Mirta Benšić

Nenad Šuvak

Uvod u vjerojatnost i statistiku

Osijek, 2014.

M. Benšić, N. Šuvak – Uvod u vjerojatnost i statistiku.

Izdavač: Sveučilište J.J. Strossmayera, Odjel za matematiku

Recenzenti: Prof.dr.sc. Miljenko Huzak
Prof.dr.sc. Tibor Pogány

Lektor: Marina Tomić, prof.

Tehnička obrada: Prof.dr.sc. Mirta Benšić, Doc.dr.sc. Nenad Šuvak

CIP zapis dostupan u računalnom katalogu Gradske i sveučilišne knjižnice Osijek pod brojem 131009075.

ISBN 978-953-6931-63-7

Udžbenik se objavljuje uz suglasnost Senata Sveučilišta J. J. Strossmayera u Osijeku pod brojem 20/13.

Udžbenik se tiska uz financijsku potporu Ministarstva znanosti, obrazovanja i sporta.

 Tisak: Grafika d.o.o., Osijek

Predgovor

Ova knjiga nastala je za potrebe kolegija koji se bave uvođenjem osnovnih pojmova i koncepata vjerojatnosti i statistike. Za razumijevanje sadržaja potrebno je znanje iz standardnih matematičkih kolegija prve studijske godine tehničkih fakulteta u Republici Hrvatskoj, tj. osnove diferencijalnog i integralnog računa realnih funkcija više varijabli te geometrije ravnine i prostora.

Knjiga je podijeljena u pet poglavlja. Prva tri poglavlja posvećena su teoriji vjerojatnosti s obzirom da je ona temelj za razumijevanje modela potrebnih za statističke analize. Četvrto poglavlje odnosi se na statistiku i napisano je s ciljem razumijevanja koncepata koji se koriste u statističkim analizama. Pretpostavka je autora da će ovladavanje sadržajima iz tog poglavlja dati studentima osnovna znanja potrebna za razumijevanje i primjenu statističkih modela i statističkih procedura prezentiranih u literaturi koja se bavi specijalno statistikom. U petom poglavlju navedeni su pojmovi iz algebre skupova te osnovni rezultati iz kombinatorike i teorije ponovljenih redova koji se koriste u prethodnim poglavljima.

Zahvaljujemo svima koji su pomogli da se ova knjiga tiska i bude što bolja. To se posebno odnosi na recenzente koji su pažljivo pročitali rukopis te svojim primjedbama i sugestijama utjecali na poboljšanje mnogih dijelova teksta, kao i na kolege Andreu Krajina, Slobodana Jelića, Mariju Miloloža-Pandur i Ivonu Puljić jer su svojim sugestijama doprinijeli kvaliteti primjera i zadataka.

Autori će biti zahvalni svim čitateljima na primjedbama vezanima uz eventualne pogreške, nepreciznosti ili nedostatke.

U Osijeku, ožujak 2014. Mirta Benšić i Nenad Šuvak

Sadržaj

Poglavlje 1

Pojam i osnovna svojstva vjerojatnosti

Pojam vjerojatnosti nastao je u pokušaju brojčanog izražavanja stupnja vjerovanja da će se dogoditi neki zamišljeni događaj. Na primjer, često možemo čuti ili pročitati: "vjerojatnost da će sutra padati kiša je 75%", "vjerojatnost da dobijem prolaznu ocjenu na ispitu mi je oko 50%", "100% sam siguran u pobjedu Blanke Vlašić na ovom natjecanju", itd. Pitanje je kako nastaju brojevi kojima je izražen stupanj vjerovanja da se dogodi neki događaj i kako ih možemo iskoristiti. Teorija vjerojatnosti dio je matematike koji se bavi tom problematikom.

Nastanak se teorije vjerojatnosti tradicionalno stavlja u 17. stoljeće iako postoje dokazi da su se već indijski matematičari (3. st. pr. Kr.) bavili pitanjima koja pripadaju današnjoj teoriji vjerojatnosti te da je u 14. stoljeću postojala praksa pomorskog osiguranja koja je omogućila srednjovjekovnim trgovcima ocjenjivanje različitih faktora rizika koji se pojavljuju prilikom prekomorskog trgovanja. Prvi matematički rezultati koji se mogu jasno prepoznati kao temelj teorije vjerojatnosti vezani su uz igre na sreću i uz izradu tablica smrtnosti. Više detalja o nastanku teorije vjerojatnosti pogledajte npr. na internetskoj adresi http://www.economics.soton.ac.uk/staff/aldrich/Figures.htm.

Povijesno gledano, koncept vjerojatnosti događaja temelji se na ideji odnosa dijela i cjeline. Pri tome se odnos dijela i cjeline koristi na dva načina:

- klasičan pristup i
- statistički pristup.

U sljedećim poglavljima opisat ćemo detaljno oba povijesna koncepta vjerojatnosti kao i aksiomatsku definiciju vjerojatnosti kojom ćemo se koristiti u ovom kolegiju, ali prije toga trebamo upoznati temeljni objekt koji se koristi prilikom matematičkog modeliranja vjerojatnosti. Zvat ćemo ga **skup ili prostor elementarnih događaja**.

1.1 Prostor elementarnih događaja

Bacanje igraće kocke jedan je od klasičnih i jednostavnih pokusa kojim ćemo se koristiti u ovom kolegiju za ilustraciju mnogih novih pojmova. Prilikom bacanja igraće kocke kao rezultat jednog bacanja pojavljuje se točno jedan broj iz skupa $\{1, 2, 3, 4, 5, 6\}$. Ako je kockica pravilno izrađena, opravdano je pretpostaviti jednaku mogućnost pojavljivanja bilo kojeg od navedenih šest brojeva. Koristeći se idejom vjerojatnosti događaja kao odnosa dijela i cjeline, možemo zaključiti da je vjerojatnost pojavljivanja parnog broja pri bacanju igraće kocke ista kao i vjerojatnost pojavljivanja neparnog broja s obzirom da je parnih brojeva isto koliko i neparnih u cjelini koju čini navedeni skup. Također, vjerojatnost pojavljivanja broja dva mora biti ista kao i vjerojatnost pojavljivanja broja pet jer oni čine po brojnosti jednake dijelove cjeline.

Na osnovu principa odnosa dijela i cjeline uočavamo da je prvi korak prema definiranju vjerojatnosti događaja spoznavanje cjeline. Naime, da bismo imali mjeru za vjerojatnost da se dogodi ono što nas konkretno zanima, moramo prvo znati što je to "sve" što se može dogoditi za naš pokus ili promatranje, tj. što čini "cjelinu". Zato ćemo, da bismo počeli graditi matematički model, uz jedan pokus (odnosno promatranje) vezati skup koji kao elemente sadrži sve što se može realizirati kad izvedemo pokus (promatranje) i zvat ćemo ga **skup svih mogućih ishoda** ili **prostor elementarnih događaja**. Najčešće takav skup označavamo sa Ω. Jedina pretpostavka na prostor elementarnih događaja jest da je neprazan i da zaista sadrži sve što se može realizirati u pokusu (odnosno promatranju).

Ako skup svih mogućih ishoda Ω sadrži samo jedan element, onda pokus ima samo jednu moguću realizaciju pa točno znamo što će se dogoditi kada ga izvedemo, zbog čega takav pokus zovemo **deterministički pokus**.

Primjer 1.1 (DETERMINISTIČKI POKUS).

a) Rezultat zagrijavanja vode pod normalnim atmosferskim tlakom na temperaturu od $100°C$ *ima samo jedan ishod: voda isparava, tj. voda iz tekućeg prelazi u plinovito agregatno stanje.*

b) Miješanje vode i jestivog ulja pri normalnim atmosferskim uvjetima ima jedinstveni ishod - stvara se emulzija ulja i vode pri čemu ulje pliva na vodi.

c) Pritiskanje papučice "gasa" pri vožnji tehnički ispravnog automobila ima jedinstveni ishod - brzina automobila povećava se. Analogno tomu, pritiskanje kočnice u istim uvjetima ima za ishod smanjenje brzine kretanja automobila.

Prostore elementarnih događaja koji imaju više elemenata (barem dva, a može i beskonačno mnogo!) koristimo ako ne možemo sa sigurnošću znati realizaciju pokusa (promatranja). Za takav pokus kažemo da ima slučajne ishode, tj. da je pokus **slučajan**.

Primjer 1.2 (Slučajan pokus)**.**

a) **Bacanje igraće kocke** - *ako nas zanima ishod jednog bacanja igraće kocke, "sve", tj. prostor elementarnih događaja jest skup*

$$\{1, 2, 3, 4, 5, 6\}.$$

b) **Slučajan izbor znamenke** - *ako nas zanima ishod slučajnog izbora jedne znamenke u dekadskom sustavnu, prostor elementarnih događaja jest skup*

$$\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}.$$

c) **Loto 6 od 45** - *ako nas zanima rezultat izvlačenja u igri "Loto 6 od 45", prostor elementarnih događaja jest skup*

$$\{\{i_1, \ldots, i_6\} : i_1, \ldots, i_6 \in \{1, \ldots, 45\},\ 1 \leq i_1 < \ldots < i_6 \leq 45\}.$$

d) Ako prognoziramo sutrašnju temperaturu zraka u Osijeku (u hladu), prostor elementarnih događaja možemo postaviti na različite načine. Jedna je mogućnost, npr. $\Omega = [-50, 50]$, *ili* $\Omega = [-100, 100]$. *Zapravo, možemo pretpostaviti i da je* $\Omega = [-273.15, \infty\rangle$, *gdje je* $-273.15°C$ *temperatura poznata pod nazivom apsolutna nula. Važno je da* Ω *bude neprazan skup i da uistinu sadrži sve moguće realizacije prognoziranja. Naravno da je ponekad praktično ugraditi već u* Ω *što više informacija o pokusu koji modeliramo, tj. suziti* Ω, *ali to nije nužno za modeliranje.*

Jednom kad smo odredili prostor elementarnih događaja slučajnog pokusa, tj. Ω, treba izraziti stupanj vjerovanja da se neki događaj dogodi. Međutim, kako opisujemo događaj kao matematički objekt? Do sada znamo da prostor elementarnih događaja sadrži sve moguće ishode pokusa, no sadrži li i sve događaje koji su nama zanimljivi? Na primjer, u pokusu bacanja igraće kocke zanima nas hoće li se okrenuti paran broj. Očigledno je da će se dogoditi paran broj ako se realizira bilo koji od ishoda 2, 4 ili 6. Dakle, taj je događaj prirodno modelirati kao podskup od Ω i to kao skup $\{2, 4, 6\}$. Očigledno je da ćemo događaje vezane uz neki slučajan pokus modelirati kao podskupove od Ω. Općenito u teoriji vjerojatnosti neki podskupovi skupa Ω nisu prikladni da ih zovemo događajima, ali zasad nećemo detaljno opisivati familiju događaja o kojoj će više riječi biti u poglavlju 1.4. Bit će nam dovoljno znati da je **događaj** vezan uz dani slučajan pokus uvijek **podskup prostora elementarnih događaja** tog pokusa.

U nastavku navodimo povijesne pristupe za modeliranje vjerojatnosti događaja (klasičan i statistički), kao i definiciju vjerojatnosti koja se danas standardno koristi u matematičkoj teoriji.

1.2 Klasičan pristup

U praksi često susrećemo primjere pokusa u kojima je **prostor elementarnih događaja konačan** i svaki je pojedini ishod **jednako moguć**. Veliki broj klasičnih igara na sreću odgovara tim pretpostavkama. Na primjer:

za bacanje pravilno izrađenog novčića (idealan slučaj, tj. nema varanja!) prostor elementarnih događaja jest $\Omega = \{pismo, glava\}$ i svaki je od tih ishoda jednako moguć,

za bacanje pravilno izrađene igraće kocke (idealan slučaj, tj. nema varanja!) prostor elementarnih događaja jest $\Omega = \{1, 2, 3, 4, 5, 6\}$ i svaki je od tih ishoda jednako moguć,

za "Loto 6 od 45" prostor elementarnih događaja jest

$$\Omega = \{\{i_1, \ldots, i_6\} : i_1, \ldots, i_6 \in \{1, \ldots, 45\}, 1 \leq i_1 < \ldots < i_6 \leq 45\}$$

i realizacija je svakog šestorčlanog skupa iz Ω jednako moguća,

za rulet je prostor elementarnih događaja $\Omega = \{0, 1, 2, \ldots, 36\}$ i svaki je od tih ishoda jednako moguć.

Kod takvih je pokusa moguće prebrojati sve ishode koji su povoljni za događaj i sve ishode koji nisu povoljni za događaj. Omjer se tih dvaju brojeva često u praksi koristi za izražavanje stupnja vjerovanja u realizaciju događaja. Evo nekoliko primjera:

Primjer 1.3. *U slučajnom pokusu bacanja igraće kocke, na temelju pretpostavke o istoj mogućnosti da se okrene bilo koji broj, kaže se da je šansa* $3 : 3$ *da se okrene paran broj. Dakle, stupanj vjerovanja u pojavu događaja izražen je stavljanjem u omjer broja ishoda koji su povoljni i broja ishoda koji su nepovoljni. Međutim, isti se omjer može iskazati i na mnogo drugih načina, npr.* $1 : 1$.

Primjer 1.4. *U skupini ljudi iz koje se sasvim slučajno bira jedna osoba nalazi se* 30 *muškaraca i* 60 *žena. U tom je pokusu šansa da izaberemo ženu* 60 : 30. *Taj se omjer takoder može prikazati na više načina, tj. kao* 6 : 3 *ili* 2 : 1.

Omjer broja povoljnih i nepovoljnih događaja ilustriran u navedenim primjerima ne stavlja u odnos dio i cjelinu nego dva dijela iste cjeline - dio koji znači realizaciju događaja koji nas zanima i dio koji znači da se taj događaj nije realizirao. Klasičan pristup modeliranju vjerojatnosti ne računa vjerojatnost na taj način nego određuje vjerojatnost kao mjeru koja dio suprotstavlja cjelini, tj. dijeli broj ishoda povoljnih za događaj brojem svih mogućih ishoda.

Klasičan način računanja vjerojatnosti

Pretpostavke na pokus	konačan prostor elementarnih događaja, svi su ishodi pokusa jednako mogući
Događaj	$A \subseteq \Omega$
Vjerojatnost događaja A	kvocijent broja elemenata skupa A i broja elemenata skupa Ω, tj. $P(A) = \dfrac{k(A)}{k(\Omega)}$

Ovdje je $P(A)$ oznaka za vjerojatnost događaja A, dok je $k(A)$ oznaka za broj elemenata skupa A.

Problem se određivanja vjerojatnosti po klasičnom pristupu svodi na problem prebrojavanja elemenata skupova. Dakle, da bismo mogli računati vjerojatnost na spomenuti način, morat ćemo nešto znati o rješavanju kombinatornih problema.

Primjer 1.5. *Skup svih mogućih ishoda bacanja igraće kocke jest skup* $\Omega = \{1, 2, 3, 4, 5, 6\}$. *Budući da je* Ω *konačan skup i kocka je za igru (pa je pretpostavka da je pravilno izrađena, tj. mogućnost realizacije bilo kojeg od brojeva jedan do šest jest jednaka), ovdje možemo primijeniti klasičan pristup određivanja vjerojatnosti. Na primjer, vjerojatnost događaja "na gornjoj strani kocke realizirao se paran broj", koji predstavljamo skupom* $A = \{2, 4, 6\}$ *svih parnih elemenata skupa* Ω*, jednaka je vjerojatnosti događaja "na gornjoj strani kocke realizirao se neparan broj", koji predstavljamo skupom* $B = \{1, 3, 5\}$ *svih neparnih elemenata skupa* Ω*. Osim toga, vjerojatnosti događaja A i B jednake su vjerojatnosti bilo kojeg događaja koji možemo predstaviti nekim tročlanim podskupom skupa* Ω.

Primjer 1.6. *Pretpostavimo da se u šeširu nalazi sedam kuglica jednakih karakteristika (od istog su materijala, jednake mase i promjera). Poznato je da su kuglice numerirane brojevima od jedan do sedam te da su dvije kuglice crvene, a pet kuglica zelene boje. Promotrimo pokus nasumičnog izvlačenja jedne kuglice iz šešira. Na temelju načina provođenja postavljenog pokusa i sadržaja šešira možemo promatrati sljedeće slučajeve.*

Zanimaju nas ishodi slučajnog pokusa temeljeni na broju kojim je kuglica numerirana - u tom je slučaju prostor elementarnih događaja skup

$$\Omega = \{1, 2, 3, 4, 5, 6, 7\}.$$

Primjenom klasičnog pristupa računanja vjerojatnosti svi događaji koji se mogu predstaviti jednakobrojnim podskupovima skupa Ω *jednako su mogući, no kako u šeširu ima više kuglica numeriranih neparnim brojevima, vjerojatnost događaja "izvučena je kuglica numerirana neparnim brojem" koji predstavljamo skupom* $A = \{1, 3, 5, 7\}$ *je veća od vjerojatnosti događaja "izvučena kuglica numerirana je parnim brojem" koji predstavljamo skupom* $B = \{2, 4, 6\}$*, tj.*

$$P(A) = \frac{4}{7}, \quad P(B) = \frac{3}{7}, \quad P(A) > P(B).$$

Zanimaju nas ishodi slučajnog pokusa temeljeni na boji kuglice - u tom je slučaju prostor elementarnih događaja skup

$$\Omega = \{C, Z\},$$

pri čemu C *označava događaj "izvučena je kuglica crvene boje", a* Z *događaj "izvučena je kuglica zelene boje". Budući u šeširu ima dvije crvene i pet zelenih kuglica, vidimo da elementarni događaji* C *i* Z *nisu jednako mogući, tj.*

$$P(C) = \frac{2}{7}, \quad P(Z) = \frac{5}{7}, \quad P(C) < P(Z).$$

Klasičan koncept pripisivanja vjerojatnosti pojedinim događajima odigrao je veliku ulogu u razvoju teorije vjerojatnosti, ali nije uvijek primjenjiv. Čak i ako je prostor elementarnih događaja konačan, ne moraju biti svi ishodi jednako mogući. Zamislimo samo da igramo 1000 igara s kockom te da se u 90% bacanja okrenuo broj 6. Tko je od nas sklon vjerovati da su u tom pokusu svi ishodi jednako mogući?

Primjer 1.7. *Slučajni pokusi u kojima nije ispunjen uvjet jednake vjerojatnosti svih elementarnih događaja su npr.*

bacanje nepravilno izrađenog (tzv. nestandardnog) novčića (tj. novčića pri čijem je bacanju favorizirana realizacija ili pisma ili glave),

bacanje nepravilno izrađene (tzv. nestandardne) igraće kocke (tj. kocke pri čijem je bacanju favorizirana realizacija jednog ili više brojeva iz skupa $\{1, 2, 3, 4, 5, 6\}$*).*

1.3 Statistički pristup

Ideja određivanja stupnja vjerovanja u pojavu događaja kao dijela cjeline ili kao omjer dijela povoljnog za događaj i dijela nepovoljnog za događaj može se iskoristiti i u drugom kontekstu. Primjer za to jest prognoziranje pobjednika nekog sportskog susreta na temelju rezultata prethodnih natjecanja istih suparnika. Npr., ako je uzajamnim natjecanjima u tenisu igrač A pobijedio igrača B u 9 od 11 susreta, njegova se šansa za pobjedu u sljedećem susretu najčešće prognozira kao 9 : 2. Princip primijenjen u tom primjeru temelji se na mogućnosti nezavisnog ponavljanja uvijek istog pokusa, a stupanj vjerovanja u pojavu događaja izražava se na temelju broja pojavljivanja i nepojavljivanja događaja prilikom ponavljanja. Izražavanje stupnja vjerovanja u pojavu događaja moglo bi se numerički izraziti također stavljanjem u omjer dijela i cjeline, tj. kao $\frac{9}{11}$.

Za računanje vjerojatnosti koja slijedi tu logiku iskoristit ćemo pojmove frekvencije i relativne frekvencije događaja.

Definicija 1.1. *Pokus je ponovljen n puta. Ako se pritom događaj A dogodio n_A puta, broj n_A zovemo* **frekvencija događaja** *A. Broj*

$$f_A(n) = \frac{n_A}{n}$$

zovemo **relativna frekvencija događaja** *A.*

Iz definicije frekvencije slijedi da je frekvencija n_A cijeli broj za koji vrijedi

$$0 \leq n_A \leq n,$$

a relativna frekvencija $f_A(n)$ racionalan broj takav da je

$$0 \leq \frac{n_A}{n} \leq 1.$$

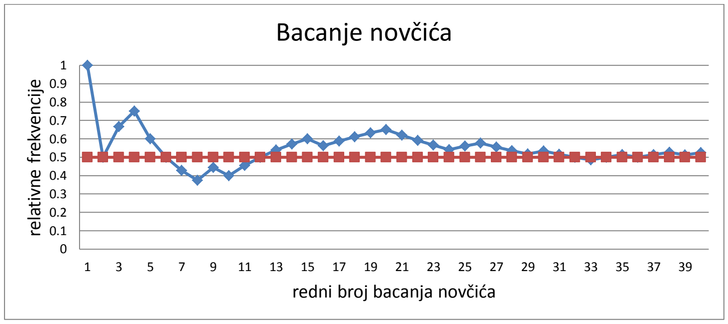
Primjer 1.8. *Promotrimo slučajan pokus bacanja pravilno izrađenog novčića sa skupom elementarnih događaja $\Omega = \{p, g\}$, gdje je p oznaka za događaj "palo je pismo", a g oznaka za događaj "pala je glava". U pokušaju definiranja vjerojatnosti da će pri jednom bacanju pasti pismo možemo koristiti klasičan pristup. Na taj način određujemo vjerojatnost da padne pismo kao*

$$P(\{p\}) = \frac{1}{2}.$$

Međutim, kako znamo da je novčić pravilno izrađen? Iskustveno znamo da to možemo provjeriti na sljedeći način: ako isti pokus ponovimo n puta, gdje je $n \in \mathbb{N}$ velik broj, kod pravilno se izrađenog novčića pismo i glava realiziraju približno jednak broj puta. U tom će slučaju biti $n_p \approx n/2$ i $n_g \approx n/2$. Prema tome, relativna frekvencija realizacije pisma iznosit će

$$\frac{n_p}{n} \approx \frac{1}{2}$$

(slika 1.1). U takvom je slučaju razumno uzeti $1/2$ kao vjerojatnost pojavljivanja pisma pri jednom bacanju novčića, tj. prihvatiti pretpostavku da je novčić pravilno izrađen.



Slika 1.1: Relativna frekvencija pojave pisma u ovisnosti o broju bacanja "n" za jedan niz bacanja.

Iskustvo nas uči da se kod mnogo nezavisnih ponavljanja istog pokusa relativna frekvencija nekog događaja stabilizira u okolini nekog broja. To svojstvo zovemo **statistička stabilnost relativnih frekvencija**.

Statistički pristup određivanja vjerojatnosti baziran je upravo na tom pricipu. Naime, ako slučajan pokus ima svojstvo statističke stabilnosti relativnih frekvencija, tada za vjerojatnost proizvoljnog događaja A vezanog uz taj pokus uzimamo realan broj $P(A) = p_A$ oko kojega se grupiraju relativne frekvencije $f_A(n)$ tog događaja.

1.4 Definicija vjerojatnosti

Oba navedena povijesna pristupa u određivanju vjerojatnosti imaju veliku ulogu u primjeni rezultata teorije vjerojatnosti u praksi, ali niti jedan od njih ne daje općenitu definiciju vjerojatnosti. Iz dosadašnjih razmatranja jasno je da vjerojatnost treba definirati za podskupove prostora elementarnih događaja. Zapravo, vjerojatnost ćemo definirati kao jednu mjeru skupova (podskupova od Ω) slično kao što je duljina jedna mjera skupova na pravcu ili površina jedna mjera skupova u ravnini. Iz teorijskih razloga (zainteresirani čitatelj dodatne informacije može pronaći u svakoj knjizi iz teorije mjere, npr. [3], [14]) nećemo vjerojatnost definirati uvijek za svaki podskup prostora elementarnih događaja, nego ćemo definirati dovoljno bogatu familiju podskupova od Ω koja će činiti temelj za definiciju vjerojatnosti. Takvu familiju podskupova od Ω zvat ćemo **familijom događaja**, a zahtjev koji mora ispunjavati dan je definicijom 1.2.

Definicija 1.2. *Neka je dan neprazan skup Ω. Familija $\mathcal{F}$ podskupova skupa Ω jest* **σ-algebra skupova na** Ω *ako vrijedi:*

i) $\emptyset \in \mathcal{F}$,

ii) *ako je $A \in \mathcal{F}$, onda je i $A^c \in \mathcal{F}$,*

iii) *ako je dana prebrojiva familija skupova $(A_n, n \in \mathbb{N}) \subseteq \mathcal{F}$, onda $\mathcal{F}$ sadrži i njihovu uniju, tj.*

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}.$$

Ako je Ω prostor elementarnih događaja nekog pokusa, bilo koja σ-algebra na njemu može igrati ulogu familije događaja tog pokusa.

Svojstvo ii) naziva se zatvorenost σ-algebre na komplementiranje (tj. ako sadrži neki skup, sadrži i njegov komplement), a svojstvo iii) zatvorenost σ-algebre na prebrojivu uniju (tj. ako imamo prebrojivo mnogo skupova iz dane σ-algebre, onda će i njihova unija biti element te σ-algebre).

Iako je σ-algebra definirana samo s tri zahtjeva, ona može biti vrlo bogata. Naime, sadrži sve skupove koji će nam biti zanimljivi u primjenama, tj. skupove koji nastaju primjenom konačno ili prebrojivo mnogo standardnih skupovnih operacija nad skupovima koji su njezini elementi. Prije svega, uočimo da je cijeli Ω sigurno element σ-algebre, s obzirom da je komplement praznoga skupa. Nadalje, σ-algebra sigurno sadrži i sve konačne unije svojih elemenata jer se konačna unija uvijek može nadopuniti do prebrojive korištenjem prebrojivo mnogo praznih skupova. Primjerom 1.9 ilustriran je način na koji možemo provjeriti zatvorenost σ-algebre i za neke druge slučajeve.

Primjer 1.9.

- *Ako su $A, B \in \mathcal{F}$, prema svojstvu zatvorenosti σ-algebre na konačnu uniju jest skup $A \cup B$ također iz $\mathcal{F}$. Prema De Morganovom zakonu za komplementiranje konačne unije skupova i svojstvu* ii) *iz definicije σ-algebre vrijedi*

$$A^c \cap B^c = (A \cup B)^c \in \mathcal{F}.$$

Dakle, σ-algebra sadrži i presjek komplemenata svojih dvaju elemenata. S obzirom da za svaki skup vrijedi $(A^c)^c = A$, slijedi da σ-algebra sadrži i presjeke svakih svojih dvaju elemenata.

- *Pokažimo da je σ-algebra zatvorena i na prebrojive presjeke svojih elemenata. Naime, ako je $(A_n, n \in \mathbb{N}) \subseteq \mathcal{F}$ prebrojiva familija skupova, tada je prema svojstvu* ii) *iz definicije σ-algebre i $(A_n^c, n \in \mathbb{N})$ također prebrojiva familija događaja iz $\mathcal{F}$. Nadalje, prema svojstvu* iii) *iz definicije σ-algebre slijedi da je*

$$\bigcup_{n=1}^{\infty} A_n^c \in \mathcal{F}.$$

Naposljetku, prema De Morganovom zakonu za komplementiranje prebrojive unije skupova i svojstvu ii) *iz definicije σ-algebre slijedi da je*

$$\left(\bigcup_{n=1}^{\infty} A_n^c\right)^c = \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}.$$

Time je dokazana zatvorenost σ-algebre na prebrojivo mnogo presjeka svojih elemenata.

- *Ako su $A, B \in \mathcal{F}$, prema svojstvu* ii) *iz definicije σ-algebre slijedi da su A^c i B^c elementi σ-algebre $\mathcal{F}$. Zbog zatvorenosti σ-algebre na konačne presjeke slijedi da je*

$$A \setminus B = A \cap B^c \in \mathcal{F}.$$

Dakle, σ-algebra sadrži i razliku svojih dvaju elemenata.

S obzirom na vezu koja postoji između povezivanja rečenica jezika i skupovnih operacija, σ-algebra zadovoljava potrebe koje nastaju u primjeni. Naime, čim sadrži neke skupove, sadrži i one skupove koji nastaju primjenom uobičajenih rečenica za njihovo kombiniranje. U sljedećem je primjeru ilustrirana ta tvrdnja na najčešće korištenim rečenicama.

Primjer 1.10. *Od svih nastavnika zaposlenih u nekoj srednjoj školi biramo jednu osobu te promatramo sljedeće događaje:*

A = *izabrana osoba predaje matematiku,*

B = *izabrana je osoba ženskog spola.*

Standardne skupovne operacije ilustrirane na primjeru tih dvaju događaja opisujemo sljedećim rečenicama:

$A \cup B$ = *izabrana osoba predaje matematiku ili je ženskog spola,*

$A \cap B$ = *izbrana osoba predaje matematiku i ženskog je spola*
(izabrana je osoba profesorica matematike),

$A \setminus B$ = *izabrana osoba predaje matematiku i nije ženskog spola*
(izabrana je osoba profesor matematike),

$B \setminus A$ = *izabrana je osoba ženskog spola i ne predaje matematiku*
(izabrana profesorica ne predaje matematiku),

A^c = *izabrana osoba ne predaje matematiku,*

B^c = *izabrana osoba nije ženskog spola, tj. izabrana je osoba muškog spola.*

Najmanja je σ-algebra na nekom skupu Ω tzv. **trivijalna σ-algebra** $\mathcal{F}_0 = \{\emptyset, \Omega\}$, dok je najveća ona koja sadrži sve podskupove od Ω, tj. **partitivni skup skupa** Ω, kojega označavamo s $\mathcal{P}(\Omega)$.

Nakon što smo definirali prostor elementarnih događaja slučajnog pokusa, odabiremo neku σ-algebru na njemu. Zvat ćemo je **familija događaja** toga pokusa, a pojedine elemente te familije događaja zovemo jednostavno **događajima**. Na odabranoj familiji događaja definirat ćemo vjerojatnost aksiomatskom definicijom 1.3.

Definicija 1.3. *Neka je Ω neprazan prostor elementarnih događaja i $\mathcal{F}$ σ-algebra skupova na njemu. Funkciju*

$$P \colon \mathcal{F} \to \mathbb{R}$$

zovemo **vjerojatnost** *na Ω ako zadovoljava sljedeće zahtjeve:*

A1. **nenegativnost vjerojatnosti***: $P(A) \geq 0$, za sve $A \in \mathcal{F}$,*

A2. **normiranost vjerojatnosti***: $P(\Omega) = 1$,*

A3. ***σ-aditivnost vjerojatnosti****: ako je dana prebrojiva familija međusobno disjunktnih skupova $(A_i, i \in I) \subseteq \mathcal{F}$, $I \subseteq \mathbb{N}$, tj. $A_i \cap A_j = \emptyset$ čim je $i \neq j$, tada vrijedi*

$$P\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} P(A_i).$$

Zahtjeve A1. - A3. *nazivamo* **aksiomima vjerojatnosti**.

Na temelju rezultata poglavlja 1.2 vidimo da vjerojatnost koja je definirana na konačnom prostoru elementarnih događaja korištenjem klasičnog pristupa udovoljava svim zahtjevima iz definicije 1.3[1].

Aksiomatski pristup daje puno veće mogućnosti u načinu zadavanja vjerojatnosti. Vidimo da je moguće zadati vjerojatnost ne samo na konačnom skupu nego i na bilo kojem nepraznom skupu s pripadnom σ-algebrom, dok na konačnom skupu možemo jasno zadati vjerojatnost i u slučaju da nemamo jednako moguće ishode. Na primjer, ukoliko smo dugo bilježili rezultate bacanja jedne igraće kocke te utvrdili da se niti jednom nije okrenuo broj 3, dok se broj 2 pojavljuje približno dvostruko češće nego brojevi 1, 4, 5 i 6, koristeći statistički pristup logično bi bilo pretpostaviti da

[1]Prvu aksiomatsku definiciju vjerojatnosti dao je Kolmogorov 1933. godine.

ishodi nisu jednako mogući te definirati vjerojatnost kao funkciju P koja zadovoljava zahtjeve iz definicije vjerojatnosti, a ujedno ima i sljedeća svojstva: $P(\{3\}) = 0$, $P(\{2\}) = 2P(\{1\})$, $P(\{1\}) = P(\{4\}) = P(\{5\}) = P(\{6\})$.

Definicija 1.4. *Neka je Ω neprazan skup, $\mathcal{F}$ σ-algebra događaja na njemu, a P vjerojatnost na Ω. Uređenu trojku $(\Omega, \mathcal{F}, P)$ zovemo* **vjerojatnosni prostor**.

1.5 Osnovna svojstva vjerojatnosti

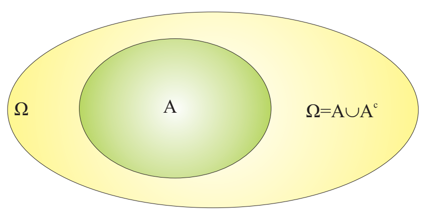
Na temelju zahtjeva navedenih u aksiomatskoj definiciji vjerojatnosti mogu se dokazati mnoga druga svojstva koja će automatski biti zadovoljena čim znamo da je zadanom funkcijom definirana vjerojatnost. U ovom poglavlju navedena su i dokazana neka od njih koja se najčešće koriste u praksi.

S1. Vjerojatnost suprotnog događaja

Neka je $(\Omega, \mathcal{F}, P)$ dani vjerojatnosni prostor i $A \in \mathcal{F}$ događaj. Suprotni događaj događaju A je njegov komplement, tj. događaj A^c. Vrijedi:

$$P(A^c) = 1 - P(A).$$

Dokaz. Skup Ω možemo prikazati kao uniju disjunktnih skupova A i A^c (slika 1.2).



Slika 1.2: Skup Ω kao unija disjunktnih skupova A i A^c.

Primjenom aksioma A3. i A2. iz definicije vjerojatnosti slijedi:

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c) \;\Rightarrow\; P(A^c) = 1 - P(A).$$

Primjer 1.11. *U šeširu se nalazi dvadeset crvenih i dvije zelene kuglice. Pretpostavimo da svaka kuglica, bez obzira na boju, može biti izvučena s jednakom vjerojatnošću, tj. ako označimo kuglice $k_1, \ldots k_{22}$ tada pretpostavljamo da je*

$$P(\{k_i\}) = \frac{1}{22}, \quad \forall i \in \{1, \ldots, 22\}.$$

Pretpostavimo da n puta, $n \in \mathbb{N}$, izvlačimo točno jednu kuglicu iz šešira, ali tako da se nakon svakog izvlačenja kuglica vraća u šešir i pomiješa s ostalim kuglicama. Dakle, jedan ishod slučajnog pokusa koji se sastoji od n izvlačenja jedne kuglice shvaćamo kao jednu varijaciju s ponavljanjem n-tog razreda skupa od 22 različita elementa, a pripadni je prostor elementarnih događaja skup svih takvih varijacija s ponavljanjem. Znamo da takvih varijacija ima ukupno 22^n te da su, zbog pretpostavke o jednakoj vjerojatnosti izvlačenja bilo koje od 22 kuglice, svi elementi od Ω jednako vjerojatni. Dakle, za računanje je vjerojatnosti zanimljivih podskupova od Ω opravdano koristiti klasičan pristup.

Na primjer, zanima nas kolika je vjerojatnost da u tih n izvlačenja kuglice iz šešira niti jednom nije izvučena zelena kuglica. U tu svrhu praktično je koristiti svojstvo vjerojatnosti suprotnog događaja. Naime, definirajmo događaj A na sljedeći način:

A - u n je ponavljanja slučajnog pokusa barem je jednom izvučena zelena kuglica.

Njemu suprotan događaj jest

A^c - u n ponavljanja slučajnog pokusa niti jednom nije izvučena zelena kuglica

= u n je ponavljanja slučajnog pokusa svaki put izvučena crvena kuglica.

Događaj A^c sastoji se od onih elemenata iz Ω čiji su svi elementi crvene kuglice, a takvih ima 20^n. Slijedi da je

$$P(A^c) = \left(\frac{20}{22}\right)^n = \left(\frac{10}{11}\right)^n.$$

Primjenom svojstva vjerojatnosti suprotnog događaja vidimo da je

$$P(A) = 1 - \left(\frac{10}{11}\right)^n.$$

S2. Vjerojatnost praznog skupa

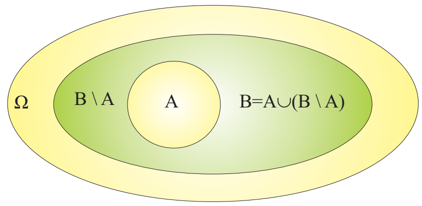
Neka je $(\Omega, \mathcal{F}, P)$ dani vjerojatnosni prostor. Tada vrijedi $P(\emptyset) = 0$.

Dokaz. S obzirom da je $\emptyset = \Omega^c$, primjena aksioma A2. iz definicije vjerojatnosti i svojstva S1. dokazuje tu tvrdnju.

S3. Monotonost vjerojatnosti

Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor te $A, B \in \mathcal{F}$ takvi da je $A \subseteq B$. Tada je $P(A) \leq P(B)$.

Dokaz. Ako je $A \subseteq B$, tada se B može prikazati kao unija disjunktnih skupova A i $(B \setminus A)$ (slika 1.3).



Slika 1.3: Skup B kao unija disjunktnih skupova A i $(B \setminus A)$.

S obzirom da je vjerojatnost nenegativna funkcija, slijedi da je

$$P(B) = P(A) + P(B \setminus A) \geq P(A).$$

Iz prethodnog izraza možemo zaključiti da ako je $A \subseteq B$, tada vjerojatnost razlike $B \setminus A$ računamo kao razliku vjerojatosti događaja B i A, tj. ako je $A \subseteq B$, tada je

$$P(B \setminus A) = P(B) - P(A).$$

Primjer 1.12. *Odredimo vjerojatnost da iz skupa svih pozitivnih dvoznamenkastih brojeva na slučajan način odaberemo broj djeljiv s tri koji istovremeno nije djeljiv s devet. Pretpostavimo da svaki dvoznamenkasti broj može biti odabran s jednakom vjerojatnošću. U tom je slučaju za računanje spomenute vjerojatnosti opravdano koristiti klasičan pristup.*

Definirajmo sljedeće događaje:

skup svih pozitivnih dvoznamenkastih brojeva (prostor elementarnih događaja):

$$\Omega = \{n \in \mathbb{N} : 10 \leq n \leq 99\}, \quad k(\Omega) = 90;$$

skup svih pozitivnih dvoznamenkastih brojeva djeljivih brojem tri:

$$T = \{n \in \Omega : \exists t \in \mathbb{N} \text{ takav da } n = 3t\}, \quad k(T) = 30;$$

skup svih pozitivnih dvoznamenkastih brojeva djeljivih brojem devet:

$$D = \{n \in \Omega : \exists d \in \mathbb{N} \text{ takav da } n = 9d\}, \quad k(D) = 10.$$

Budući da je svaki broj koji je djeljiv brojem devet ujedno djeljiv i brojem tri, zaključujemo da je $D \subset T$. Primjenom klasičnog pristupa određivanju vjerojatnosti slijedi da je

$$P(D) = \frac{k(D)}{k(\Omega)} = \frac{1}{9}, \quad P(T) = \frac{k(T)}{k(\Omega)} = \frac{1}{3}.$$

Sada možemo izračunati vjerojatnost događaja od interesa, tj. događaja $T \setminus D$. Naime, skup $T \setminus D$ sadrži sve pozitivne dvoznamenkaste brojeve djeljive brojem tri koji nisu djeljivi brojem devet. Budući da je $D \subset T$, slijedi da je

$$P(T \setminus D) = P(T) - P(D) = \frac{1}{3} - \frac{1}{9} = \frac{2}{9}.$$

S4. Vjerojatnost unije događaja

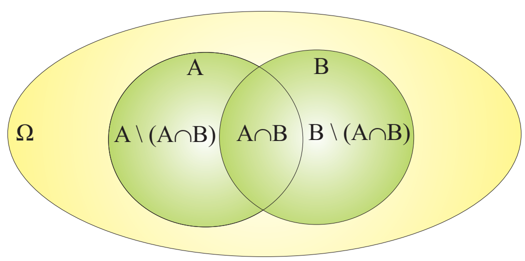
O vjerojatnosti unije događaja govori treći zahtjev iz definicije vjerojatnosti, tj. σ-aditivnost vjerojatnosti. Međutim, tim zahtjevom obuhvaćene su samo familije disjunktnih skupova. To svojstvo daje formulu za izračun vjerojatnosti unije dvaju skupova i u slučaju kada skupovi nisu disjunktni.

Neka je $(\Omega, \mathcal{F}, P)$ dani vjerojatnosni prostor te $A, B \in \mathcal{F}$. Tada vrijedi

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Dokaz. Unija skupova A i B može se prikazati kao unija disjunktnih skupova na sljedeći način (slika 1.4):

$$A \cup B = (A \setminus (A \cap B)) \cup (B \setminus (A \cap B)) \cup (A \cap B)\,.$$



Slika 1.4: Skup $(A \cup B)$ kao unija disjunktnih skupova.

Iz navedenoga proizlazi:

$$P(A \cup B) = P\left(A \setminus (A \cap B)\right) + P\left(B \setminus (A \cap B)\right) + P(A \cap B).$$

S obzirom da je $(A \cap B) \subseteq A$ i $(A \cap B) \subseteq B$, vrijedi:

$$\begin{aligned} P(A \cup B) &= P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

Primjer 1.13. *Želimo izračunati vjerojatnost slučajnog odabira dvoznamenkastog broja koji je djeljiv brojem tri ili brojem sedam. Pretpostavimo da svaki dvoznamenkasti broj može biti odabran s jednakom vjerojatnošću. U tom je slučaju za računanje spomenute vjerojatnosti opravdano koristiti klasičan pristup.*

Sada uz prostor elementarnih događaja

$$\Omega = \{n \in \mathbb{N} : 10 \leq n \leq 99\}\,, \quad k(\Omega) = 90,$$

i skup dvoznamenkastih brojeva djeljivih brojem tri

$$T = \{n \in \Omega : \exists t \in \mathbb{N} \text{ takav da } n = 3t\}, \quad k(T) = 30,$$

iz primjera 1.12 promatramo i skup svih dvoznamenkastih brojeva djeljivih brojem sedam, tj. skup

$$S = \{n \in \Omega : \exists s \in \mathbb{N} \text{ takav da } n = 7s\}, \quad k(S) = 13.$$

Događaj koji nas zanima jest $T \cup S$, a njegovu ćemo vjerojatnost odrediti primjenom formule za vjerojatnost unije dvaju događaja. U tu svrhu trebamo odrediti i kardinalni broj skupa svih dvoznamenkastih brojeva djeljivih i brojem tri i brojem sedam, tj. skupa

$$T \cap S = \{n \in \Omega : \exists t, s \in \mathbb{N} \text{ takvi da } n = 3t \wedge n = 7s\}.$$

Budući da je $T \cap S = \{21, 42, 63, 84\}$, tj. $k(T \cap S) = 4$, slijedi da je

$$P(T \cup S) = P(T) + P(S) - P(T \cap S) = \frac{1}{3} + \frac{13}{90} - \frac{4}{90} = \frac{39}{90} = \frac{13}{30}.$$

Formula za računanje vjerojatnosti unije više skupova može se dati i općenito, tj. za uniju konačno mnogo skupova. Ta formula poznata je pod imenom **Sylvesterova formula**. Njezin je dokaz moguće provesti koristeći formulu za vjerojatnost unije dvaju skupova i princip matematičke indukcije (vidi [26]).

Vezano uz vjerojatnost unije dvaju skupova valja uočiti da vrijedi nejednakost

$$P(A \cup B) \leq P(A) + P(B),$$

gdje su $A, B \in \mathcal{F}$, a $(\Omega, \mathcal{F}, P)$ dani vjerojatnosni prostor. Takva se nejednakost također može generalizirati i to ne samo na uniju konačno mnogo događaja nego i na uniju prebrojivo mnogo događaja. To generalizirano svojstvo zove se σ-subaditivnost vjerojatnosti.

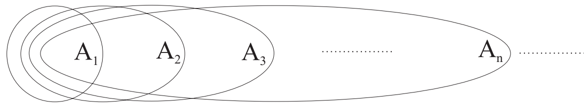
S5. σ-subaditivnost vjerojatnosti

Neka je $(\Omega, \mathcal{F}, P)$ dani vjerojatnosni prostor i neka je $(A_i, i \in I) \subseteq \mathcal{F}$, $I \subseteq \mathbb{N}$, prebrojiva familija događaja toga prostora. Tada je

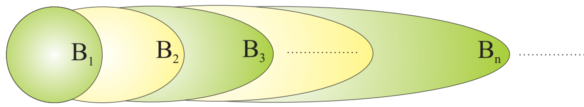
$$P\left(\bigcup_{i \in I} A_i\right) \leq \sum_{i \in I} P(A_i).$$

Dokaz. Iz zadane familije događaja $(A_i, i \in I)$, $I \subseteq \mathbb{N}$, formirat ćemo novu familiju međusobno disjunktnih događaja na sljedeći način (slika 1.5, slika 1.6):

$$B_1 = A_1, \quad B_2 = A_2 \setminus A_1, \quad \ldots, \quad B_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i, \ldots.$$



Slika 1.5: Familija događaja $(A_i, i \in I)$, $I \subseteq \mathbb{N}$.



Slika 1.6: Familija disjunktnih događaja $(B_i, i \in I)$, $I \subseteq \mathbb{N}$.

Tako definirani skupovi jesu međusobno disjunktni, a vrijedi i da je

$$B_i \subseteq A_i \quad \text{i} \quad \bigcup_{i \in I} B_i = \bigcup_{i \in I} A_i.$$

Primjenom aksioma A3. na familiju $(B_i, i \in I)$ slijedi da je

$$P\left(\bigcup_{i \in I} A_i\right) = P\left(\bigcup_{i \in I} B_i\right) = \sum_{i \in I} P(B_i) \leq \sum_{i \in I} P(A_i).$$

Primjer 1.14. *Automat za igre na sreću programiran je tako da je vjerojatnost pojavljivanja prirodnog broja jednaka 2^{-n}, tj.*

$$P(\{n\}) = \frac{1}{2^n}, \quad \forall n \in \mathbb{N}.$$

Uočimo da takvim definiranjem vjerojatnosti pojavljivanja različitih prirodnih brojeva nisu jednako mogući događaji.

Odredimo vjerojatnost pojavljivanja bilo kojeg prirodnog broja manjeg ili jednakog od $(n+1)$, tj. vjerojatnost događaja

$$S = \{1, 2, \ldots, n+1\}, \quad n \in \mathbb{N}.$$

Skup S može se na različite načine prikazati kao konačna unija familije disjunktnih skupova ili skupova koji nisu disjunktni. Na primjer, ako definiramo skupove

$$\begin{aligned} A_1 &= \{1, 2\}, \\ A_2 &= \{2, 3\}, \\ A_3 &= \{3, 4\}, \\ &\vdots \\ A_n &= \{n, n+1\}, \quad n \in \mathbb{N}, \end{aligned}$$

vidimo da za konačnu familiju skupova $(A_i, i \leq n)$, $n \in \mathbb{N}$, vrijedi:

$$S = \bigcup_{i=1}^{n} A_i, \quad P(A_i) = \frac{1}{2^i} + \frac{1}{2^{i+1}} = \frac{3}{2^{i+1}}, \quad \forall i \leq n.$$

Međutim, $P(S)$ ne možemo dobiti sumiranjem svih $P(A_i)$ jer $(A_i, i \leq n)$ nije familija disjunktnih skupova, no ako S prikažemo kao konačnu uniju jednočlanih skupova (uočimo da su oni ujedno disjunktni!), tada lako možemo izračunati $P(S)$:

$$S = \bigcup_{i=1}^{n} \{i\}, \;\; P(S) = P\left(\bigcup_{i=1}^{n} \{i\}\right) = \sum_{i=1}^{n} \frac{1}{2^i} = \frac{2^n - 1}{2^n}.$$

Svojstvo σ-subaditivnosti vjerojatnosti primijenjeno na uniju skupova A_i, $i \leq n$, u navedenom primjeru možemo potvrditi na osnovu nejednakosti

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P(S) = \frac{2^n - 1}{2^n} < \frac{3}{2}\,\frac{2^n - 1}{2^n} = \sum_{i=1}^{n} P(A_i), \quad n \in \mathbb{N}.$$

S6. Neprekidnost vjerojatnosti u odnosu na rastuću familiju događaja

Neprekidnost vjerojatnosti u odnosu na rastuću familiju događaja omogućava računanje vjerojatnosti unije rastuće familije događaja kao limesa niza vjerojatnosti pojedinačnih događaja u familiji. Preciznije, neka je $(\Omega, \mathcal{F}, P)$ dani vjerojatnosni prostor i neka je $(A_n, n \in \mathbb{N}) \subseteq \mathcal{F}$ rastuća familija događaja, tj. $A_n \in \mathcal{F}$ za sve $n \in \mathbb{N}$ i

$$A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subseteq A_{n+1} \ldots;$$

tada vrijedi da je

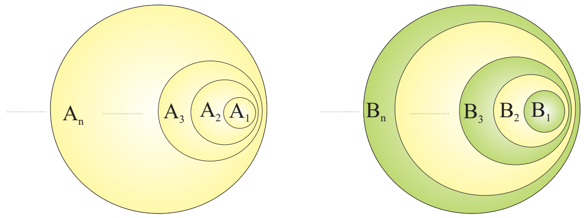
$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \to \infty} P(A_n).$$

Dokaz. Prvo uočimo da $\lim\limits_{n\to\infty} P(A_n)$ postoji. Naime, zbog monotonosti vjerojatnosti niz je brojeva $(P(A_n), n \in \mathbb{N})$ monotono rastući. S obzirom da se radi o nizu vjerojatnosti, taj je niz brojeva sadržan u segmentu $[0, 1]$, tj. ograničen je. Budući da je monoton i ograničen niz realnih brojeva konvergentan, navedena granična vrijednost zaista postoji. Od zadane rastuće familije događaja $(A_n, n \in \mathbb{N})$ formirat ćemo novu familiju $(B_n, n \in \mathbb{N})$ međusobno disjunktnih događaja na sljedeći način (slika 1.7):

$$B_1 = A_1, \quad B_2 = A_2 \setminus A_1, \quad \ldots B_n = A_n \setminus A_{n-1}, \ldots.$$

Osim što je $(B_n, n \in \mathbb{N})$ familija disjunktnih događaja, vrijedi i sljedeće:

$$A_n = \bigcup_{i=1}^{n} B_i, \quad \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i.$$



Slika 1.7: Rastuća familija $(A_n, n \in \mathbb{N})$ i disjunktna familija $(B_n, n \in \mathbb{N})$.

Dakle, vrijedi:

$$P(A_n) = \sum_{i=1}^{n} P(B_i),$$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \lim_{n\to\infty} P(A_n).$$

Primjer 1.15. *Pretpostavimo da se radi o istom automatu za igre na sreću kao u primjeru 1.14, ali zanima nas vjerojatnost pojavljivanja bilo kojeg neparnog broja, tj. vjerojatnost skupa*

$$N = \{2n-1 : n \in \mathbb{N}\}.$$

Vjerojatnost skupa N možemo izračunati na nekoliko načina. Ovdje ilustriramo pristup koji koristi neprekidnost vjerojatnosti u odnosu na rastuću familiju događaja.

Prikažimo N pomoću familije skupova $(A_n, n \in \mathbb{N})$ gdje su skupovi A_n definirani na sljedeći način:

$$\begin{aligned}
A_1 &= \{1\},\\
A_2 &= \{1,3\},\\
A_3 &= \{1,3,5\},\\
&\vdots\\
A_n &= \{1,3,5,\ldots,2n-1\}, \quad n \in \mathbb{N},\\
&\vdots
\end{aligned}$$

Vidimo da je $A_1 \subset A_2 \subset A_3 \subset \ldots \subset A_n \subset \ldots$, tj. $(A_n, n \in \mathbb{N})$ je rastuća familija skupova i vrijedi da je

$$\begin{aligned}
N &= \bigcup_{i=1}^{\infty} A_i,\\
P(A_n) &= \sum_{i=1}^{n} \frac{1}{2^{2i-1}} = 2\sum_{i=1}^{n} \frac{1}{4^i} = \frac{2}{3}\left(1 - \frac{1}{4^n}\right), \quad \forall n \in \mathbb{N}.
\end{aligned}$$

Zbog neprekidnosti vjerojatnosti u odnosu na rastuću familiju događaja slijedi da je

$$P(N) = \lim_{n\to\infty} P(A_n) = \frac{2}{3}\lim_{n\to\infty}\left(1 - \frac{1}{4^n}\right) = \frac{2}{3}.$$

S7. Neprekidnost vjerojatnosti u odnosu na padajuću familiju događaja

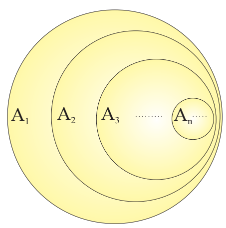
Neprekidnost vjerojatnosti u odnosu na padajuću familiju događaja omogućava računanje vjerojatnosti presjeka padajuće familije skupova kao limesa niza vjerojatnosti pojedinačnih skupova u familiji. Preciznije, neka je $(\Omega, \mathcal{F}, P)$ dani vjerojatnosni prostor i neka je $(A_n, n \in \mathbb{N}) \subseteq \mathcal{F}$ padajuća familija događaja, tj. $A_n \in \mathcal{F}$ za sve n i

$$A_1 \supseteq A_2 \supseteq \ldots A_n \supseteq A_{n+1} \ldots;$$

tada vrijedi da je

$$P\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} P(A_n).$$

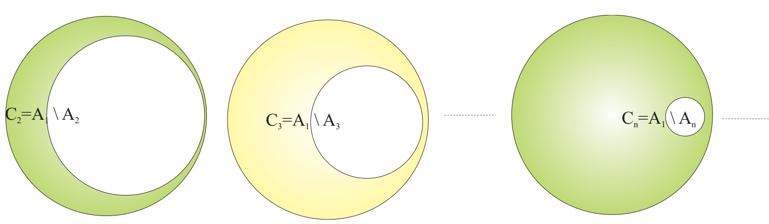
Dokaz. Slika 1.8 prikazuje padajuću familiju događaja $(A_n, n \in \mathbb{N}) \subseteq \mathcal{F}$.



Slika 1.8: Padajuća familija događaja $(A_n, n \in \mathbb{N})$.

Od zadane padajuće familije događaja $(A_n, n \in \mathbb{N})$ formirat ćemo novu familiju $(C_n, n \in \mathbb{N})$ koja će biti monotono rastuća te ćemo u dokazu iskoristiti svojstvo neprekidnosti vjerojatnosti u odnosu na monotono rastuće familije događaja (svojstvo S6). Familiju $(C_n, n \in \mathbb{N})$ definiramo na sljedeći način (slika 1.9):

$$C_1 = \emptyset, C_2 = A_1 \setminus A_2, C_3 = A_1 \setminus A_3, \ldots, C_n = A_1 \setminus A_n, \ldots$$



Slika 1.9: Rastuća familija događaja $(C_n, n \in \mathbb{N})$.

Za tako definiranu rastuću familiju događaja vrijedi sljedeće:

$$\bigcup_{i=1}^{\infty} C_i = A_1 \setminus \left(\bigcap_{i=1}^{\infty} A_i \right),$$

$$P\left(A_1 \setminus \left(\bigcap_{i=1}^{\infty} A_i\right)\right) = P\left(\bigcup_{i=1}^{\infty} C_i\right) = \lim_{i\to\infty} P(C_i) = \lim_{i\to\infty} \left(P(A_1) - P(A_i)\right).$$

Dakle, vrijedi:

$$P\left(A_1 \setminus \left(\bigcap_{i=1}^{\infty} A_i\right)\right) = P(A_1) - P\left(\bigcap_{i=1}^{\infty} A_i\right) = P(A_1) - \lim_{i\to\infty} P(A_i),$$

odakle slijedi tvrdnja.

Primjer 1.16. *Pretpostavimo da se radi o istom automatu za igre na sreću kao u primjeru 1.14. Odredimo s kojom se vjerojatnošću pojavljuje bilo koji paran broj veći ili jednak od unaprijed zadanog prirodnog broja. Skupove A_n, za $n \in \mathbb{N}$, definiramo na sljedeći način:*

$$\begin{array}{rcl} A_1 & = & \{2,4,6,8,10,\ldots\}, \\ A_2 & = & \{4,6,8,10,\ldots\}, \\ A_3 & = & \{6,8,10,\ldots\}, \\ & \vdots & \\ A_n & = & \{2n, 2(n+1), 2(n+2), \ldots\}, \quad n \in \mathbb{N}, \\ & \vdots & \end{array}$$

To znači da treba odrediti $P(A_n)$. Prvo odredimo vjerojatnost presjeka prebrojive familije $(A_n, n \in \mathbb{N})$. Očito je $A_1 \supset A_2 \supset A_3 \supset \ldots \supset A_n \supset \ldots$, tj. $(A_n, n \in \mathbb{N})$ je padajuća familija skupova za koju je

$$P(A_n) = \sum_{k=n}^{\infty} \frac{1}{2^{2k}} = \frac{1}{3 \cdot 4^{n-1}}.$$

Korištenjem neprekidnosti vjerojatnosti u odnosu na padajuću familiju događaja vidimo da je

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} \frac{1}{3 \cdot 4^{n-1}} = 0,$$

što je i logično s obzirom da je presjek svih skupova A_i zapravo prazan (ne postoji paran broj koji je veći od svakog prirodnog broja.)

1.6 Vjerojatnost na diskretnom Ω

Neka je Ω konačan ili prebrojiv prostor elementarnih događaja. Takve skupove označavat ćemo na sljedeći način:

$$\Omega = \{\omega_i : i \in I_\Omega\}, \quad I_\Omega \subseteq \mathbb{N},$$

gdje je I_Ω skup indeksa. Na primjer, ako je $\Omega = \{\omega_1, \ldots, \omega_n\}$, onda je $I_\Omega = \{1, \ldots, n\}$, a ako je $\Omega = \{\omega_i : i \in \mathbb{N}\}$, onda je $I_\Omega = \mathbb{N}$. Iz definicije vjerojatnosti znamo da je vjerojatnost funkcija kojoj je domena σ-algebra događaja. Dakle, da bismo zadali konkretnu vjerojatnost na Ω, potrebno je zadati vrijednosti te funkcije na svakom skupu A sadržanom u pridruženoj σ-algebri. **Kod konačnih i prebrojivih skupova elementarnih događaja pretpostavit ćemo da je pridružena σ-algebra točno jednaka partitivnom skupu od Ω, tj. $\mathcal{F} = \mathcal{P}(\Omega)$.** U nastavku će se pokazati da je to za naše potrebe prirodno i moguće ispoštovati. Vjerojatnosni prostor kod kojega je Ω konačan ili prebrojiv skup, a pridružena je σ-algebra $\mathcal{P}(\Omega)$, zvat ćemo **diskretan vjerojatnosni prostor**.

U ovom poglavlju pokazat ćemo kako se funkcija na $\mathcal{P}(\Omega)$, kojom je zadana vjerojatnost, može odrediti zadavanjem vrijednosti samo na jednočlanim podskupovima od Ω, tj. elementima od Ω. S obzirom da $\mathcal{P}(\Omega)$ ima mnogo više elemenata nego Ω (ako konačan Ω ima $k(\Omega)$ elemenata, onda je $2^{k(\Omega)}$ broj elemenata od $\mathcal{P}(\Omega)$), na taj smo način u velikoj mjeri pojednostavili zadavanje vjerojatnosti.

Preciznije, ako imamo zadan diskretan vjerojatnosni prostor $(\Omega, \mathcal{P}(\Omega), P)$, tada za svaki pojedini $\omega_i \in \Omega$ znamo izračunati vjerojatnost

$$p_i = P\left(\{\omega_i\}\right).$$

Koristeći svojstvo σ-aditivnosti vjerojatnosti, pomoću tako dobivenog niza brojeva $(p_i, i \in I_\Omega)$, $I_\Omega \subseteq \mathbb{N}$, možemo izračunati vjerojatnost bilo kojeg događaja $A = \{\omega_i : i \in I_A\}$, gdje je $I_A \subseteq I_\Omega$ skup indeksa elemenata od A:

$$P(A) = P\left(\bigcup_{i \in I_A} \{\omega_i\}\right) = \sum_{i \in I_A} P\left(\{\omega_i\}\right) = \sum_{i \in I_A} p_i.$$

Dakle, poznavanje vrijednosti vjerojatnosti na jednočlanim podskupovima od Ω automatski određuje vjerojatnost bilo kojeg događaja tog vjerojatnosnog prostora. Time niz brojeva koji predstavljaju vjerojatnosti jednočlanih podskupova od Ω preuzima ključnu ulogu u opisivanju i zadavanju vjerojatnosti na diskretnom vjerojatnosnom prostoru.

Valja uočiti da navedeni niz brojeva $(p_i, i \in I_\Omega)$ ima sljedeća dva svojstva:

1. $p_i \geq 0$ za sve $i \in I_\Omega$,
2. $\sum_{i \in I_\Omega} p_i = 1$.

Taj rezultat možemo iskoristiti prilikom zadavanja vjerojatnosti na diskretnom vjerojatnosnom prostoru, kao što je ilustrirano sljedećim primjerima:

Primjer 1.17 (Konačno mnogo jednako mogućih ishoda). *Pretpostavimo da slučajan pokus ima konačno mnogo jednako mogućih ishoda te da je* $\Omega = \{\omega_1, \ldots, \omega_n\}$*, tj.* $k(\Omega) = n$*. Definirajmo funkciju* $P\colon \mathcal{P}(\Omega) \to [0,1]$ *na sljedeći način:*

$$P(\{\omega_i\}) = \frac{1}{n}, \quad i \in \{1, \ldots, n\},$$

$$A = \{\omega_j : j \in I_A,\, I_A \subseteq \{1, \ldots, n\}\}, \quad P(A) = \sum_{i \in I_A} P(\{\omega_i\}).$$

Na taj je način zadana funkcija na $\mathcal{P}(\Omega)$ *koja zadovoljava sve zahtjeve definicije vjerojatnosti (provjerite!). Osim toga, tako definirana vjerojatnost podudara se s klasičnim pristupom određivanja vjerojatnosti.*

Primjer 1.18. *Neka je prostor elementarnih događaja slučajnog pokusa* $\Omega = \mathbb{N}$*. Definirajmo funkciju na* σ*-algebri* $\mathcal{P}(\mathbb{N})$ *na sljedeći način:*

$$P(\{i\}) = \frac{1}{2^i}, \quad i \in \mathbb{N},$$

$$A = \{i_1, i_2, \ldots, i_n, \ldots\}, \quad P(A) = \sum_{i_j \in A} P(\{i_j\}).$$

Dokažimo da je na taj način definirana vjerojatnost na $\mathbb{N}$*.*

1. *Očigledno je* $P(A) \geq 0$ *za sve* $A \in \mathcal{P}(\mathbb{N})$*.*
2. *Provjerimo je li* $P(\mathbb{N}) = 1$*. Koristimo formulu za sumu geometrijskog reda s kvocijentom* $\frac{1}{2}$ *i prvim članom* $\frac{1}{2}$*:*

$$P(\mathbb{N}) = \sum_{i \in \mathbb{N}} \frac{1}{2^i} = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1.$$

3. *Da bismo provjerili* σ*-aditivnost uzmimo familiju* $\{A_1, A_2, \ldots, A_n, \ldots\}$ *disjunktnih podskupova skupa* $\mathbb{N}$*. Neka je*

$$\begin{array}{rcl} A_1 & = & \{a_{11}, a_{12}, \ldots, a_{1j}, \ldots\} \subseteq \mathbb{N}, \\ A_2 & = & \{a_{21}, a_{22}, \ldots, a_{2j}, \ldots\} \subseteq \mathbb{N}, \\ & \vdots & \\ A_n & = & \{a_{n1}, a_{n2}, \ldots, a_{nj}, \ldots\} \subseteq \mathbb{N}, \quad n \in \mathbb{N}, \\ & \vdots & \end{array}$$

Pri tome je

$$P(A_n) = \sum_{j=1}^{\infty} \frac{1}{2^{a_{nj}}} \leq 1.$$

Uočimo također da, zbog disjunktnosti skupova A_i*, među brojevima* $a_{nj} \in \mathbb{N}$ *nema istih pa se njihova unija može prikazati kao*

$$\bigcup_{n=1}^{\infty} A_n = \{a_{11}, a_{12}, \ldots, a_{1j}, \ldots, a_{21}, a_{22}, \ldots, a_{2j}, \ldots, a_{n1}, a_{n2}, \ldots, a_{nj}, \ldots\}.$$

Vjerojatnost unije dobije se sumiranjem brojeva oblika (2^{-k}) *po svim* $k \in \bigcup_{n=1}^{\infty} A_n$*. S obzirom da je* $\bigcup_{n=1}^{\infty} A_n \subseteq \mathbb{N}$*, niz* $\left(2^{-k}, k \in \bigcup_{n=1}^{\infty} A_n\right)$ *podniz je geometrijskog niza s kvocijentom*

1/2. Time je i red koji se dobije sumiranjem članova tog niza konvergentan i vrijedi (vidi poglavlje 5.3 u Dodatku):

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty}\sum_{j=1}^{\infty} \frac{1}{2^{a_{nj}}} = \sum_{n=1}^{\infty} P(A_n).$$

Takav način zadavanja vjerojatnosti na diskretnom vjerojatnosnom prostoru može se primijeniti uvijek. Ako je s $\Omega = \{\omega_i, i \in I_\Omega \subseteq \mathbb{N}\}$ dan prostor elementarnih događaja, potreban je samo niz brojeva $(p_i, i \in I_\Omega)$ sa svojstvima

1. $p_i \geq 0$ za sve $i \in I_\Omega$,
2. $\displaystyle\sum_{i\in I_\Omega} p_i = 1$

da bismo definirali vjerojatnost na tom diskretnom Ω kao

$$P(A) = \sum_{i\in I_A} P(\{\omega_i\}) = \sum_{i\in I_A} p_i.$$

Naime, vrijedi sljedeća tvrdnja:

Neka je s $(p_i, i \in I)$, $I \subseteq \mathbb{N}$ zadan niz realnih brojeva sa svojstvima:

1. **$p_i \geq 0$ za sve $i \in I$,**
2. **$\sum_{i\in I} p_i = 1$,**

te neka je Ω bilo koji skup koji ima $k(I)$ elemenata. Tada je izrazom

$$P(A) = \sum_{i\in I_A} p_i, \quad A \subseteq \Omega, \tag{1.1}$$

gdje je I_A skup indeksa elemenata iz Ω koji pripadaju skupu A, dobro definirana vjerojatnost na $(\Omega, \mathcal{P}(\Omega))$.

Dokaz. Neka je $\Omega = \{\omega_i, i \in I\}$ dan neprazan skup. Dokažimo da je funkcijom (1.1) dobro definirana vjerojatnost.

1. Očigledno je $P(A) \geq 0$ za sve $A \in \mathcal{P}(\Omega)$.
2. $P(\Omega) = \sum_{i\in I} p_i = 1$.
3. Neka su $A_1, A_2, \ldots, A_n, \ldots$ disjunktni podskupovi od Ω i $A = \bigcup_{i=1}^{\infty} A_i$. Tada familija skupova $\{A_1, A_2, \ldots, A_n, \ldots\}$ čini jednu particiju skupa A. Koristeći rezultate teorije ponovljenih redova (poglavlje 5.3 u Dodatku) možemo

zaključiti da vrijedi:

$$P(A) = \sum_{j\in I_A} p_j = \sum_{i\in\mathbb{N}} \sum_{j\in I_{A_i}} p_j = \sum_{i\in\mathbb{N}} P(A_i),$$

pa vrijedi σ-aditivnost.

Primjer 1.19. *Dan je niz triju brojeva: $\frac{1}{6}$, $\frac{1}{5}$ i $\frac{1}{3}$. Možemo li taj niz nadopuniti četvrtim brojem tako da ta četiri broja dobro definiraju vjerojatnost na diskretnom vjerojatnosnom prostoru u smislu ovoga poglavlja? Odgovor je potvrdan. Naime, navedena tri broja pozitivna su i manja od 1, a suma im iznosi $\frac{21}{30}$, što je još uvijek manje od 1. Dakle, ako za četvrti broj izaberemo razliku $1 - \frac{21}{30} = \frac{9}{30}$, prema rezultatima ovoga poglavlja slijedi da postoji vjerojatnosni prostor na kojemu pomoću tih brojeva možemo zadati vjerojatnost. Npr. uzmemo $\Omega = \{1, 2, 3, 4\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ i*

$$P(\{1\}) = \frac{1}{6}, \quad P(\{2\}) = \frac{1}{5}, \quad P(\{3\}) = \frac{1}{3}, \quad P(\{4\}) = \frac{9}{30}.$$

1.7 Primjeri vjerojatnosti na $\mathbb{R}$

Postoji mnoštvo slučajnih pokusa za koje je prirodno pretpostaviti da prostor elementarnih događaja sadrži neki interval.

Primjer 1.20. *Vrijeme (mjereno u sekundama) koje postiže atletičar u utrci na 100 metara može se modelirati kao rezultat slučajnog pokusa s vrijednostima iz intervala $[0, 13]$. S obzirom da se vrijeme na utrkama danas mjeri vrlo precizno, nije mudro modelirati rezultate takvog pokusa kao diskretan skup. Zbog toga uzimamo npr. $\Omega = [0, 13]$ ili neki veći skup koji sadrži interval $[0, 13]$.*

Vjerojatnost u takvim vjerojatnosnim prostorima ne može se odrediti koristeći niz vrijednosti koje ona postiže na pojedinačim ishodima već i zbog činjenice da prostor elementarnih događaja nije prebrojiv. Jedan od načina na koji je moguće zadati vjerojatnost na intervalima, a koji je prikladan za računanje, jest korištenje nenegativne realne funkcije definirane na skupu $\mathbb{R}$ koja s osi apscisa zatvara jediničnu površinu. Naime, odaberemo takvu nenegativnu realnu funkciju i zadamo vjerojatnost nekog skupa A kao površinu koju zatvara graf te funkcije s osi apscisa nad skupom A.

Primjer 1.21. *Pretpostavimo da računalo izvodi neku numeričku proceduru s točnošću na 6 decimanih mjesta i daje rezultat 1.234567, no da je stvaran rezultat te procedure zapravo neki realan broj iz intervala $[1.234567, 1.234568\rangle$. U tom kontekstu kao prostor elementarnih događaja promatramo interval $\Omega = [1.234567, 1.234568\rangle$. Kako nemamo razloga preferirati određene brojeve iz Ω kao stvarne rezultate te numeričke procedure, pretpostavljamo da su svi podintervali od Ω koji imaju jednaku duljinu jednako mogući. U skladu s tom pretpostavkom i shvaćanjem vjerojatnosti kao dijela cjeline, vjerojatnost da je stvaran rezultat iz intervala $A \subseteq \Omega$ u ovom se primjeru može zadati kao kvocijent duljine intervala A i duljine cijelog Ω:*

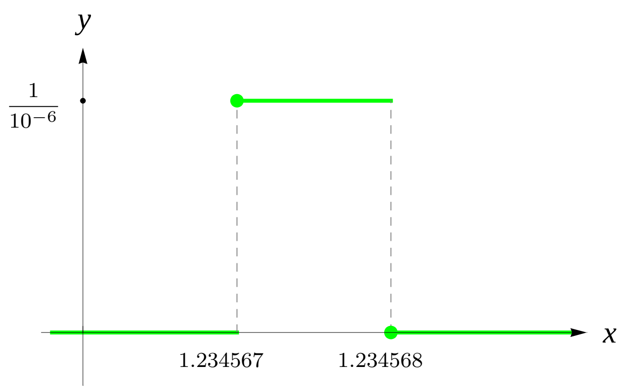
$$P(A) = \frac{d(A)}{d(\Omega)}.$$

Tako definirana funkcija na σ-algebri koja sadrži sve podintervale od Ω zadovoljavat će zahtjeve postavljene u definiciji vjerojatnosti[2] *te je time dobro definirana vjerojatnost na* $\Omega = [1.234567, 1.234568\rangle$. *Zovemo ju* **geometrijska vjerojatnost** *na* $\Omega = [1.234567, 1.234568\rangle$.

Primjer 1.22. *Vjerojatnost iz prethodnog primjera definiranu na* $\Omega = [1.234567, 1.234568\rangle$ *možemo proširiti na* $\Omega' = \mathbb{R}$ *korištenjem nenegativne realne funkcije*

$$f(x) = \begin{cases} \frac{1}{10^{-6}} & , x \in [1.234567, 1.234568\rangle \\ 0 & , x \notin [1.234567, 1.234568\rangle \end{cases} .$$

Graf funkcije f prikazan je na slici 2.15.



Slika 1.10: Graf funkcije $f(x) = \frac{1}{10^{-6}} \cdot \mathrm{I}_{[1.234567, 1.234568\rangle}(x)$.

Vjerojatnost događaja $A \subseteq \Omega'$ (označimo je P') definiramo kao površinu koju zatvara graf te funkcije s osi apscisa nad intervalom A. Uočimo da za intervale $A \subseteq \Omega'$ koji su ujedno i podintervali od Ω iz primjera 1.21 vrijedi

$$P'(A) = P(A),$$

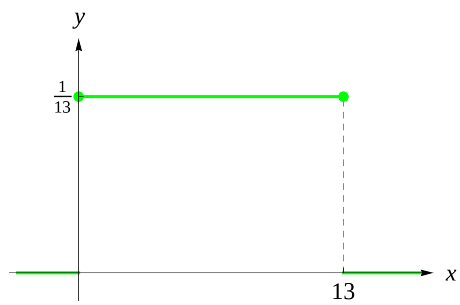
gdje je P geometrijska vjerojatnost definirana u primjeru 1.21.

Prednost je zadavanja vjerojatnosti pomoću nenegativne funkcije u odnosu na vjerojatnost kao kvocijent duljina intervala u tome što se može generalizirati i na probleme u kojima nema razloga pretpostavljati da su svi ishodi jednako mogući.

Primjer 1.23. *U primjeru 1.20 nije razumno očekivati istu vjerojatnost da atletičar postigne vrijeme u intervalu* $[0, 5]$ *kao u intervalu* $[8, 13]$ *iako su duljine tih intervala jednake. Naime,*

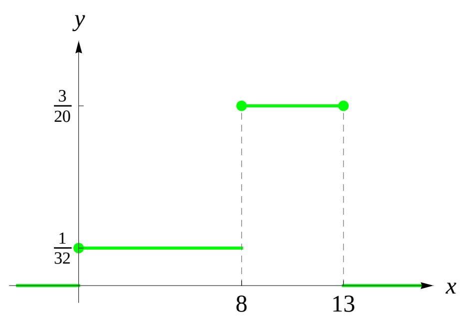
[2] Dovoljno je provjeriti da vrijede prva dva zahtjeva iz definicije vjerojatnosti te treći zahtjev za konačno mnogo disjunktnih intervala.

poznato je da srednjoškolci uobičajeno postižu vrijeme od oko 12 *sekundi pri trčanju na* 100 *metara, a rezultati se atletičara uglavnom kreću između* 9 *i* 10 *sekundi. Vjerojatnost koja je definirana na intervalu* $[0, 13]$ *korištenjem principa jednako mogućih ishoda imala bi pripadni graf funkcije kao što je prikazano na slici 1.11.*



Slika 1.11: Graf funkcije koja uniformno definira vjerojatnost na [0, 13]

Preferiranje intervala $[8, 13]$ *nad ostatkom možemo npr. predočiti grafom funkcije na slici 1.12.*



Slika 1.12: Graf funkcije koja definira vjerojatnost na [0, 13] uz preferiranje intervala [8, 13].

Da bi realnom funkcijom mogli modelirati vjerojatnost na opisani način bitno je da površina koju zatvara njezin graf s osi apscisa iznad cijelog Ω *bude jednaka* 1*, a graf funkcije može rasti i padati na* Ω *tako da odražava stupanj vjerovanja da se pojedini intervali realiziraju u stvarnom pokusu. Ako želimo neki interval preferirati nad ostalim područjem, iznad njega i vrijednosti funkcije kojom opisujemo vjerojatnost trebaju biti veće.*

U ovom trenutku nećemo se baviti odgovorom na pitanje je li vjerojatnost određena funkcijom čiji je graf dan na slici 1.12 u skladu s problemom opisanim u primjeru ili treba raditi dodatne modifikacije. O usklađenosti modela sa stvarnim problemima bit će riječi u drugom dijelu knjige koji se bavi statistikom.

Općenito, neka je dana realna funkcija f realne varijable, tj. $f\colon \mathbb{R} \to \mathbb{R}$ takva da je

- $f(x) \geq 0$, za sve $x \in \mathbb{R}$

- $\int\limits_{-\infty}^{\infty} f(x)\,dx = 1.$

Neka je $\mathcal{B}$ najmanja σ-algebra podskupova od $\mathbb{R}$ koja sadrži sve intervale. Izrazom

$$P(A) = \int\limits_A f(x)\,dx, \quad A \in \mathcal{B}$$

definirana je vjerojatnost na $\mathcal{B}$. Matematička teorija koja omogućava dokaz te činjenice može se naći npr. u [26].

Ukoliko je prostor elementarnih događaja samo jedan interval Ω koji je pravi podskup od $\mathbb{R}$, potrebno je da $\int\limits_\Omega f(x)\,dx$ bude jednak 1. To možemo lako zadovoljiti tako da na Ω^c stavimo vrijednosti funkcije $f(x)$ na nulu.

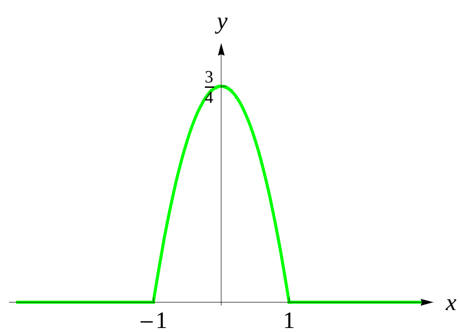
Primjer 1.24. *Pokažimo da pomoću funkcije $f : \mathbb{R} \to \mathbb{R}$ (slika 1.13) definirane formulom*

$$f(x) = \begin{cases} \frac{3}{4}\,(1 - x^2) & , \; x \in [-1, 1] \\ 0 & , \; x \notin [-1, 1] \end{cases},$$

možemo definirati vjerojatnost na skupu $\mathbb{R}$, tj. da je funkcija f nenegativna i normirana:

- *nenegativnost - iz analitičke definicije funkcije f slijedi da je $f(x) \geq 0$, $\forall x \in \mathbb{R}$,*
- *normiranost - površina je ispod grafa funkcije f (pogledati sliku) nad skupom $\mathbb{R}$*

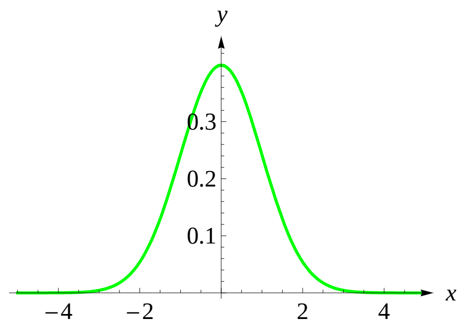
$$\int\limits_{\mathbb{R}} f(x)\,dx = \int\limits_{-1}^{1} \frac{3}{4}\,(1 - x^2)\,dx = \frac{3}{4} \cdot \frac{4}{3} = 1.$$



Slika 1.13: Graf funkcije $f(x) = \frac{3}{4}\,(1 - x^2) \cdot \mathrm{I}_{[-1,1]}(x)$.

Primjer 1.25. *Važan je primjer nenegativne funkcije $f : \mathbb{R} \to \mathbb{R}$ pomoću koje možemo definirati vjerojatnost na $\mathbb{R}$ funkcija definirana formulom (slika 1.14)*

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$



Slika 1.14: Gaussova krivulja - graf funkcije $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

1.8 Primjeri vjerojatnosti na $\mathbb{R}^2$ i $\mathbb{R}^3$

Ukoliko je prostor elemetarnih događaja Ω pravokutnik ili neki drugi geometrijski lik, za modeliranje vjerojatnosti na Ω može se prenijeti logika nastala pri modeliranju vjerojatnosti na intervalima. Slično se može primijeniti i na skupove Ω koji su zadani kao geometrijska tijela.

Primjer 1.26. *U pokusu slučajnog izbora točke iz kruga radijusa r vjerojatnost možemo zadati pomoću kvocijenta dijela i cjeline. Tako je vjerojatnost da bude izabrana točka iz kružnog odsječka A površine $s(A)$ dana izrazom*

$$P(A) = \frac{s(A)}{r^2\pi}.$$

Naime, $s(\Omega) = r^2\pi$ predstavlja površinu cijelog kruga, a površina dijela koji nas zanima označena je sa $s(A)$.

Primjer 1.27. *Pri slučajnom izboru točke iz kugle radijusa r vjerojatnost također možemo zadati pomoću kvocijenta dijela i cjeline. Tako je vjerojatnost da bude izabrana točka iz kuglinog isječka A volumena $v(A)$ dana izrazom*

$$P(A) = \frac{v(A)}{\frac{4}{3}r^3\pi}$$

s obzirom da je $v(\Omega) = \dfrac{4}{3}r^3\pi$ volumen cijele kugle.

Za modeliranje se vjerojatnosti koje preferiraju neke dijelove prostora elementarnih događaja u odnosu na druge dijelove istih dimenzija (površine odnosno volumena) također može prenijeti logika zadavanja vjerojatnosti pomoću nenegativne realne funkcije. U slučaju kad je $\Omega = \mathbb{R}^2$ (odnosno, $\Omega = \mathbb{R}^3$) radi se o nenegativnoj realnoj funkciji $f\colon \mathbb{R}^2 \to \mathbb{R}$ (odnosno, $f\colon \mathbb{R}^3 \to \mathbb{R}$) koja mora zadovoljati sljedeće svojstvo:

$$\int\limits_{\Omega} f(\mathbf{x})\, d\mathbf{x} = 1,$$

gdje je $\mathbf{x} \in \mathbb{R}^2$ (odnosno, $\mathbf{x} \in \mathbb{R}^3$).
Uočimo da je za dvodimenzionalne Ω to dvostruki integral, pa postavljeni zahtjev zapravo znači da volumen tijela koje je omeđeno grafom funkcije f i ravninom (x_1, x_2) na dijelu za koji vrijedi $(x_1, x_2) \in \Omega$ mora iznositi 1. Za trodimenzionalne Ω to je trostruki integral.
Vjerojatnost je skupa[3] $A \subseteq \Omega$ tada dana izrazom

$$P(A) = \int\limits_A f(\mathbf{x})\, d\mathbf{x}.$$

Primjer 1.28. *Pokažimo da pomoću funkcije $f\colon \mathbb{R}^2 \to \mathbb{R}$ definirane formulom*

$$f(x,y) = \begin{cases} \frac{3}{5}\,(x^2+y) & ,\ (x,y) \in [-1,1]\times[0,1] \\ \quad 0 & ,\ \text{za ostale } (x,y) \in \mathbb{R}^2 \end{cases}$$

možemo definirati vjerojatnost na $\mathbb{R}^2$, tj. da je funkcija f nenegativna i normirana:

- *nenegativnost - iz analitičke definicije funkcije f slijedi da je $f(x,y) \geq 0$, $\forall (x,y) \in \mathbb{R}^2$,*
- *normiranost - volumen kojega graf funkcije f zatvara sa skupom $\mathbb{R}^2$, tj. volumen ispod grafa funkcije f nad pravokutnikom $[-1,1]\times[0,1]$ iznosi*

$$\int\limits_{\mathbb{R}^2} f(x,y)\,dx\,dy = \frac{3}{5}\int\limits_0^1\int\limits_{-1}^1 (x^2+y)\,dx\,dy = 1.$$

Prema tome, vjerojatnost skupa $A = [0,1]\times[0,0.5] \subset \mathbb{R}^2$ računamo na sljedeći način:

$$P(A) = \int\limits_A f(x,y)\,dx\,dy = \frac{3}{5}\int\limits_0^{0.5}\int\limits_0^1 (x^2+y)\,dx\,dy = \frac{7}{40}.$$

Primjer 1.29. *Pokažimo da pomoću funkcije $f\colon \mathbb{R}^3 \to \mathbb{R}$ definirane formulom*

$$f(x,y,z) = \begin{cases} \frac{1}{8}\,z^2(x+y) & ,\ (x,y,z) \in [0,2]\times[0,1]\times[-2,0] \\ \quad 0 & ,\ \quad \text{za ostale } (x,y,z) \in \mathbb{R}^3 \end{cases}$$

možemo definirati vjerojatnost na $\mathbb{R}^3$, tj. da je funkcija f nenegativna i normirana:

- *nenegativnost - iz analitičke definicije funkcije f slijedi da je $f(x,y,z) \geq 0$, $\forall (x,y,z) \in \mathbb{R}^3$,*
- *normiranost - volumen kojega graf funkcije f zatvara sa skupom $\mathbb{R}^3$ iznosi*

$$\int\limits_{\mathbb{R}^2} f(x,y,z)\,dx\,dy\,dz = \frac{1}{8}\int\limits_0^2\int\limits_0^1\int\limits_{-2}^0 z^2(x+y)\,dx\,dy\,dz = 1.$$

Prema tome, vjerojatnost skupa $B = [0,1]\times[0,1]\times[-1,0] \subset \mathbb{R}^2$ računamo na sljedeći način:

$$P(B) = \int\limits_B f(x,y,z)\,dx\,dy\,dz = \frac{1}{8}\int\limits_0^1\int\limits_0^1\int\limits_{-1}^0 z^2(x+y)\,dx\,dy\,dz = \frac{1}{24}.$$

[3] O obliku σ-algebre na kojoj se može definirati vjerojatnost pogledati [26]. Specijalno, u dvodimenzionalnom slučaju vjerojatnost se na taj način može definirati na najmanjoj σ-algebri koja sadrži sve skupove oblika $(a,b)\times(c,d)$, gdje su $a,b,c,d \in \mathbb{R}$, $a < b$, $c < d$.

1.9 Uvjetna vjerojatnost. Nezavisnost

Primjer 1.30. *Nakon izleta napravili smo 5 kopija istog CD-a s fotografijama izleta, ali su samo tri kopije uspjele. Na CD-ove nismo stavili nikakve oznake i ne znamo koji su ispravni, a koji ne. Pretpostavimo da smo u žurbi i da imamo vremena za provjeru samo jednog CD-a. Izabrali smo jedan, provjerili i utvrdili da je neispravan. Međutim, u žurbi nam se taj provjereni CD pomiješao s ostalima i sada moramo uzeti jedan bez provjere pa kako bude. Mislite li da bi vjerojatnost odabira ispravnog CD-a u drugom biranju bila veća da se provjereni CD nije pomiješao s ostalima? Analizirajmo te slučajeve odvojeno.*

1. *Pomiješani slučaj. Definirajmo događaje:*

   L_i *- u i-tom izvlačenu izvučen je neispravan (loš) CD,*

   D_i *- u i-tom izvlačenju izvučen je ispravan (dobar) CD,*

   gdje je $i \in \{1, 2\}$. *S obzirom da je nakon prvog izvlačenja sve ponovo pomiješano, pri drugom izvlačenju CD-a ponovo smo na početku, tj. rezultat prvog izvlačenja uopće nema utjecaja na rezultat drugog izvlačenja. Dakle, vjerojatnost događaja* D_2 *ne ovisi o realizaciji događaja* L_1. *Slijedi da je vjerojatnost odabira ispravnog CD-a u drugom izvlačenju jednaka vjerojatnosti odabira ispravnog CD-a u prvom izvlačenju i iznosi* $\frac{3}{5}$.

2. *Nepomiješani slučaj. Ovaj puta je drugi pokus bitno promijenjen u odnosu na prvi. Sada znamo da u drugom izvlačenju biramo od četiri CD-a, od kojih su točno tri ispravna, pa je vjerojatnost događaja* D_2 *uz uvjet da se dogodio* L_1 *jednaka* $\frac{3}{4}$.

U oba slučaja prethodnog primjera pojavljuju se dva izvlačenja CD-a. U pomiješanom slučaju vjerojatnost u drugom izvlačenju ne ovisi o rezultatima prvog izvlačenja i zapravo je ista kao u prvom izvlačenju. Kažemo da drugo izvlačenje **ne ovisi** o prvom izvlačenju i možemo ih promatrati odvojeno. U nepomiješanom slučaju rezultat prvog izvlačenja utječe na vjerojatnost u drugom izvlačenju i nije mudro ta dva izvlačenja promatrati kao sasvim odvojene cjeline. Da bismo mogli proučavati takve probleme definirat ćemo tzv. **uvjetne vjerojatnosti** koje će opisivati rezultate drugog izvlačenja i to: uvjetnu vjerojatnost uz uvjet da je u prvom pokušaju izvučen dobar CD i uvjetnu vjerojatnost uz uvjet da je u prvom pokušaju izvučen loš CD.

Definicija 1.5. *Neka je dan vjerojatnosni prostor* $(\Omega, \mathcal{F}, P)$ *i događaj* $A \in \mathcal{F}$ *koji ima pozitivnu vjerojatnost, tj.* $P(A) > 0$. *Funkcija* $P(\cdot \mid A)$ *definirana na* $\mathcal{F}$ *izrazom*

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}, \quad B \in \mathcal{F}, \tag{1.2}$$

je **uvjetna vjerojatnost uz uvjet da se dogodio događaj** A.

Tako definirana funkcija zadovoljava sve zahtjeve postavljene u definiciji vjerojatnosti, tj. na taj je način definirana jedna vjerojatnost na Ω (provjerite zahtjeve iz aksiomatske definicije vjerojatnosti!). Navedena činjenica ima za posljedicu to da se na $P(B \mid A)$ mogu primijeniti sva svojstva vjerojatnosti koja su dokazana u potpoglavlju 1.5. Tako npr. vrijedi:

- $P(\emptyset \mid A) = 0$,
- $P(B^c \mid A) = 1 - P(B \mid A)$,
- za $B_1 \subseteq B_2$ je $P(B_1 \mid A) \leq P(B_2 \mid A)$, itd.

U pojedinim slučajnim pokusima uvjetne vjerojatnosti modeliraju se po istom principu kao i sve vjerojatnosti.

Primjer 1.31. *Vratimo se primjeru 1.30. Odredimo uvjetne vjerojatnosti ishoda drugog izvlačenja u nepomiješanom slučaju uvjetovane na moguće ishode prvog izvlačenja.*
Prvo uočimo da je u nepomiješanom slučaju iz primjera 1.30 vjerojatnost da je u drugom izvlačenju izvučen ispravan CD ako je u prvom izvlačenu izvučen neispravan CD zapravo vjerojatnost $P(D_2|L_1)$. Dakle, imamo:

$$P(D_2 \mid L_1) = \frac{3}{4}, \quad P(L_2 \mid L_1) = \frac{1}{4},$$
$$P(D_2 \mid D_1) = \frac{2}{4} = \frac{1}{2}, \quad P(L_2 \mid D_1) = \frac{2}{4} = \frac{1}{2}.$$

Vjerojatnost presjeka.

Izraz se iz definicije uvjetne vjerojatnosti često koristi za izračun vjerojatnosti presjeka. To omogućuje jednakost (1.2) napisanu na drugi način (pri čemu je $P(A) > 0$):

$$P(A \cap B) = P(B \mid A)P(A).$$

Primjer 1.32. *Ako u primjeru 1.30, u nepomiješanom slučaju, želimo odrediti vjerojatnost da su u oba izvlačenja izvučeni loši CD-ovi možemo primijeniti ovaj izraz. Dobijemo:*

$$P(L_2 \cap L_1) = P(L_2 \mid L_1)P(L_1) = \frac{1}{4} \cdot \frac{2}{5} = \frac{1}{10}.$$

Pojam nezavisnosti događaja.

Intuitivno je jasno da bi pojam nezavisnosti dvaju događaja A i B, pri čemu zahtijevamo da je $P(A) > 0$, morao odražavati činjenicu da realizacija jednog od njih ne utječe na realizaciju drugog. To bi se na uvjetnu vjerojatnost trebalo odraziti tako da je

$$P(B \mid A) = P(B),$$

tj. da za vjerojatnost presjeka vrijedi

$$P(A \cap B) = P(B \mid A)P(A) = P(B)P(A).$$

Primjer 1.33. *Za slučaj u kojemu smo u primjeru 1.30 pomiješali provjereni CD s ostalima također možemo odrediti uvjetnu vjerojatnost uz uvjet da je u prvom izvlačenju izvučen loš CD, kao i uvjetnu vjerojatnost uz uvjet da je u prvom izvlačenju izvučen dobar CD. Međutim, te će dvije uvjetne vjerojatnosti biti jednake, neovisno uvjetujemo li na događaj D_1 ili L_1. Uvjetne vjerojatnosti izvlačenja ispravnog, odnosno neispravnog, CD-a u drugom izvlačenju, uz uvjet da je u prvom izvlačenju izvučen neispravan CD, iznose:*

$$P(D_2 \mid L_1) = P(D_2) = \frac{3}{5}, \quad P(L_2 \mid L_1) = P(L_2) = \frac{2}{5}.$$

Nadalje, uvjetne vjerojatnosti izvlačenja ispravnog, odnosno neispravnog CD-a u drugom izvlačenju, uz uvjet da je u prvom izvlačenju izvučen ispravan CD, iznose:

$$P(D_2 \mid D_1) = P(D_2) = \frac{3}{5}, \quad P(L_2 \mid D_1) = P(L_2) = \frac{2}{5}.$$

Slijedi da je vjerojatnost da su u oba izvlačenja izvučeni loši CD-ovi

$$P(L_1 \cap L_2) = \frac{2}{5} \cdot \frac{2}{5} = \frac{4}{25}.$$

Definicija 1.6. *Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor. Kažemo da su događaji $A, B \in \mathcal{F}$* **nezavisni** *ako vrijedi:*

$$P(A \cap B) = P(A)\,P(B).$$

Sljedeća definicija pojam nezavisnosti događaja generalizira na proizvoljnu familiju događaja.

Definicija 1.7. *Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor. Kažemo da je proizvoljna familija događaja $(A_x, x \in I) \subseteq \mathcal{F}$ nezavisna ako za svaki konačan skup različitih indeksa $\{i_1, \ldots, i_n\} \subseteq I$ vrijedi*

$$P\left(\bigcap_{j=1}^{n} A_{i_j}\right) = \prod_{j=1}^{n} P(A_{i_j}).$$

Važno je napomenuti da komplementiranje događaja neće promijeniti njegovo stanje nezavisnosti od nekog drugog događaja. Naime, vrijedi sljedeća tvrdnja:

ako su A i B nezavisni događaji, onda su

$$A^c \textbf{ i } B, \qquad A \textbf{ i } B^c, \qquad A^c \textbf{ i } B^c$$

također nezavisni događaji.

Dokaz. Dokažimo samo prvu tvrdnju, ostale se dokazuju analogno. Zbog nezavisnosti je događaja A i B $P(A \cap B) = P(A)P(B)$. Upotrebom toga svojstva slijedi:

$$\begin{aligned} P(A^c \cap B) &= P(B \setminus (A \cap B)) = P(B) - P(A \cap B) = \\ &= P(B) - P(A)\,P(B) = (1 - P(A))\,P(B) = \\ &= P(A^c)\,P(B), \end{aligned}$$

tj. događaji A^c i B nezavisni su.

Primjer 1.34. *Pojam nezavisnosti događaja olakšava računanje vjerojatnosti kod nezavisnog ponavljanja istog pokusa. Promotrimo pokus izvlačenja kuglice iz šešira koji sadrži 20 crvenih i 2 zelene kuglice. Odredimo vjerojatnost izvlačenja točno dvije zelene kuglice u* 30 *ponavljanja toga pokusa. Prilikom ponavljanja prethodno izvučena kuglica vraća se u šešir i sve se dobro promiješa.*

Da bi riješili taj problem, prvo ćemo uočiti da je vjerojatnost izvlačenja zelene kuglice u svakom pojedinom izvlačenju uvijek ista i iznosi $1/11$*, dok je vjerojatnost izvlačenja crvene kuglice* $10/11$*. Prostor elementarnih događaja* Ω *tog slučajnog pokusa sastoji se od uređenih* 30*-torki slova "c" i "z" koja označavaju boju kuglice izvučene u pojedinom izvlačenju (tj.* Ω *je skup svih varijacija s ponavljanjem* 30*-og razreda dvočlanog skupa* $\{c, z\}$*).*

Vjerojatnost da su u prvim dvama pokušajima izvučene zelene kuglice, a u svim ostalima crvene, možemo izračunati kao vjerojatnost presjeka nezavisnih događaja, pa ona iznosi

$$\left(\frac{1}{11}\right)^2 \cdot \left(\frac{10}{11}\right)^{28}.$$

Međutim, isto toliko iznosi vjerojatnost pojave svake uređene 30*-torke toga pokusa u kojoj su točno dva slova "z". S obzirom da u* Ω *ima* $\binom{30}{2}$ *elemenata koji imaju "z" na točno dvjema pozicijama, a svaki od njih ima vjerojatnost* $\left(\frac{1}{11}\right)^2 \cdot \left(\frac{10}{11}\right)^{28}$*, onda vjerojatnost izvlačenja točno dviju zelenih kuglica u tih* 30 *ponavljanja pokusa iznosi*

$$\binom{30}{2} \cdot \left(\frac{1}{11}\right)^2 \cdot \left(\frac{10}{11}\right)^{28}.$$

Formula potpune vjerojatnosti

Korištenje uvjetnih vjerojatnosti može olakšati računanje vjerojatnosti u složenijim slučajevima.

Primjer 1.35. *Student je pristupio pismenom ispitu, no ne može doći pogledati rezultate, stoga ne zna je li prošao ili pao. Zamolio je svog prijatelja da to učini umjesto njega i da mu pošalje poruku po bratu. Međutim, svjestan je da obojica govore istinu samo s vjerojatnošću* $2/3$*. Kolika je vjerojatnost da će student dobiti točnu informaciju?*

Označimo A *događaj koji znači da je student dobio točnu informaciju,* L_p *činjenicu da je slagao prijatelj, a* L_b *činjenicu da je slagao brat. Događaje da su brat i prijatelj govorili istinu označit ćemo sa* I_p *i* I_b*.*

Taj primjer možemo riješiti tako da prebrojimo sve mogućnosti i izračunamo vjerojatnosti presjeka iz skupa

$$\{L_p \cap L_b, I_p \cap L_b, L_p \cap I_b, I_p \cap I_b\}.$$

Student može čuti točnu informaciju u dvama od navedenih slučajeva: $L_p \cap L_b$ *i* $I_p \cap I_b$*. Ako mu brat i prijatelj lažu neovisno jedan od drugoga, onda je*

$$P(L_p \cap L_b) = P(L_p)P(L_b) = \frac{1}{9},$$

$$P(I_p \cap I_b) = P(I_p)P(I_b) = \frac{4}{9},$$

pa je vjerojatnost da čuje točnu informaciju

$$P(A) = P(L_p \cap L_b) + P(I_p \cap I_b) = \frac{5}{9}.$$

Problem je jednostavan. Međutim, sljedeći način rješavanja istog primjera sugerira princip koji se može generalno primijeniti te olakšati računanje u mnogo složenijim situacijama.

Uočimo da prijatelj može samo lagati ili reći istinu. Dakle, događaji se L_p *i* I_p *međusobno isljučuju (disjunktni su) i opisuju sve što se može dogoditi u odnosu na informaciju koju prenosi prijatelj. Također je*

$$\begin{aligned} P(A) &= P(A \cap I_p) + P(A \cap L_p) = \\ &= P(A \mid I_p)P(I_p) + P(A \mid L_p)P(L_p) = \\ &= \frac{2}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{3} = \frac{5}{9}. \end{aligned}$$

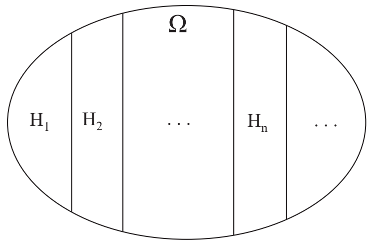
Definicija 1.8. *Konačna ili prebrojiva familija događaja* $\{H_i : i \in I\}$, $I \subseteq \mathbb{N}$, *u vjerojatnostnom prostoru* $(\Omega, \mathcal{F}, P)$ *je* **potpun sustav događaja** *ako vrijedi:*

1. $P(H_i) > 0$ *za sve* $i \in I$,
2. $H_i \cap H_j = \emptyset$ *za sve* $i \neq j$, $i, j \in I$,
3. $\bigcup_{i \in I} H_i = \Omega$.

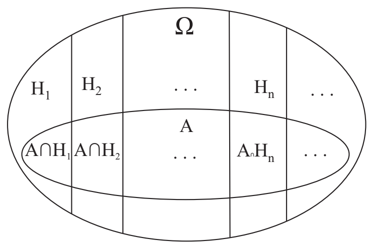
Napomena 1.1. *Potpun sustav događaja* $\{H_i : i \in I\}$, $I \subseteq \mathbb{N}$, *čini jednu particiju prostora elementarnih događaja* Ω *(slika 1.15). Prema tome, bilo koji događaj* $A \subset \Omega$ *moguće je prikazati kao uniju presjeka događaja* A *s događajima* H_i, $I \subseteq \mathbb{N}$, *potpunog sustava događaja, tj.*

$$A = \bigcup_{i \in I} (A \cap H_i),$$

što prikazujemo slikom 1.16.



Slika 1.15: Potpun sustav događaja.



Slika 1.16: Prikaz događaja $A \subset \Omega$ kao unije događaja $A \cap H_i$, $i \in I \subseteq \mathbb{N}$.

Teorem 1.1 (Formula potpune vjerojatnosti). *Neka je* $(\Omega, \mathcal{F}, P)$ *vjerojatnosni prostor i* $\{H_i : i \in I\}$, $I \subseteq \mathbb{N}$, *potpun sustav događaja na njemu. Tada za proizvoljan događaj* $A \in \mathcal{F}$ *vrijedi*

$$P(A) = \sum_{i \in I} P(A \mid H_i) P(H_i). \tag{1.3}$$

Dokaz. *Uočimo da događaj* $A \in \mathcal{F}$ *možemo predstaviti na sljedeći način:*

$$A = \bigcup_{i \in I} A \cap H_i.$$

Također uočimo da je $\{A \cap H_i : i \in I\}$, $I \subseteq \mathbb{N}$, *familija disjunktnih skupova. Prema tome slijedi da je*

$$P(A) = \sum_{i \in I} P(A \cap H_i) = \sum_{i \in I} P(A \mid H_i) P(H_i).$$

Primjer 1.36. *Ulični kockar ima dva novčića od jedne kune: jedan je pravilno izrađen (tzv. standardni novčić), a drugi s obiju strana ima slavuja (tj. nepravilno je izrađen, tzv. nestandardni novčić). Vama je ponudio da na slučajan način odaberete jedan novčić koji će on baciti dva puta. Kolika je vjerojatnost da je u oba bacanja odabrani novčić pao na slavuja?*

Očito tražena vjerojatnost ovisi o tome koji smo od dvaju ponuđenih novčića odabrali. Dakle, kao potpun sustav događaja promatramo familiju $\{H_1, H_2\}$ *gdje su događaji* H_1 *i* H_2 *definirani na* *sljedeći način:*

$$H_1 \quad - \quad \textit{odabran je standardan novčić,}$$
$$H_2 \quad - \quad \textit{odabran je nestandardan novčić.}$$

Budući da novčić izvlačimo na slučajan način, vjerojatnosti događaja H_1 i H_2 iznose

$$P(H_1) = P(H_2) = \frac{1}{2}.$$

Događaj čiju vjerojatnost želimo izračunati jest

$$A - \textit{u dvama je bacanjima odabrani novčić oba puta pao na slavuja.}$$

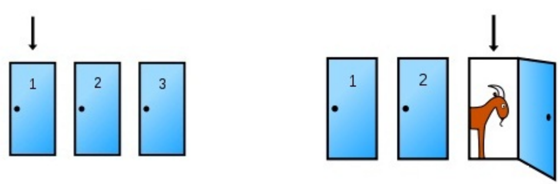
Iz definicije promatranog problema slijedi:

$$P(A \mid H_1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}, \quad P(A \mid H_2) = 1 \cdot 1 = 1.$$

Prema formuli potpune vjerojatnosti slijedi tražena vjerojatnost:

$$P(A) = P(H_1)P(A \mid H_1) + P(H_2)P(A \mid H_2) = \frac{1}{2}\left(\frac{1}{4} + 1\right) = \frac{5}{8}.$$

Primjer 1.37 (**Monty Hall problem** [4]**).** *Pretpostavite da igrač sudjeluje u igri u kojoj bira jedna od triju vrata. Iza jednih je vrata automobil, a iza preostalih dvaju vrata nalaze se koze. Npr. ako igrač izabere prva vrata, voditelj igre (koji zna iza kojih vrata se nalazi automobil, a iza kojih koze) otvorit će ona od preostalih dvaju vrata za koja zna da kriju kozu (slika 1.17).*



(a) Početni izbor (b) Potez voditelja

Slika 1.17: Monty Hall problem

Nakon toga voditelj pita igrača: "Želite li promjeniti vaš izbor vrata?". Mi se pitamo je li mudro da igrač promijeni izbor?

Promotrimo situaciju u kojoj se automobil nalazi iza drugih vrata, a igrač je odabrao prva vrata - to znači da će voditelj otvoriti treća vrata jer zna da se iza njih nalazi koza. Uvedimo sljedeće oznake za promatrane događaje:

A_i - automobil se nalazi iza i-tih vrata,

I_i - igrač je izabrao i-ta vrata,

[4] Monty Hall problem prvi put postavljen je 1975. u pismu Stevea Selvina uredniku časopisa *American Statistician*, a nazvan je prema voditelju američkog kviza *Let's make a deal*.

V_i - voditelj je otvorio i-ta vrata,

$V_i \mid I_j$ - ako je igrač odabrao j-ta vrata, voditelj je otvorio i-ta vrata, $i \neq j$,

gdje su $i, j \in \{1, 2, 3\}$. Budući da igrač ne zna iza kojih se vrata nalazi automobil, događaji A_i i I_j nezavisni su za svaki izbor i i j. Odgovor na pitanje je li mudro za igrača promijeniti izbor vrata možemo dobiti tako da izračunamo vjerojatnost da se automobil nalazi iza drugih vrata ako je igrač odabrao prva vrata i voditelj potom otvorio treća vrata. Tu vjerojatnost računamo primjenom definicije uvjetne vjerojatnosti:

$$P(A_2|I_1 \cap V_3) = \frac{P\left(V_3 \cap (A_2 \cap I_1)\right)}{P(I_1 \cap V_3)} = \frac{P(V_3 \mid A_2 \cap I_1)P(A_2)}{P(V_3 \mid I_1)}.$$

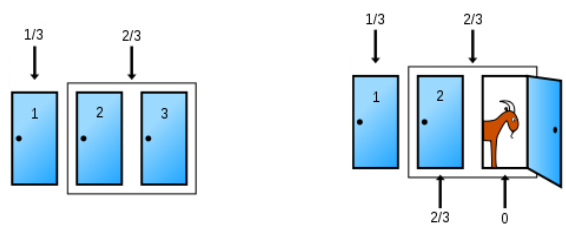
Kako će u promatranoj situaciji voditelj otvoriti treća vrata jer zna da se iza njih nalazi koza, slijedi da je $P(V_3 \mid A_2 \cap I_1) = 1$. Događaji A_1, A_2 i A_3 čine potpun sustav događaja i $P(A_1) = P(A_2) = P(A_3) = 1/3$. Preostaje izračunati $P(V_3 \mid I_1)$. Prema formuli potpune vjerojatnosti slijedi da je

$$P(V_3 \mid I_1) = P_{I_1}(V_3) = \sum_{i=1}^{3} P_{I_1}(A_i)P_{I_1}(V_3 \mid A_i) = \sum_{i=1}^{3} P(A_i \mid I_1)P\left(V_3 \mid A_i \cap I_1\right).$$

Kako su događaji A_i i I_i, $i \in \{1, 2, 3\}$ nezavisni, slijedi da je

$$P(A_i \mid I_1) = P(A_i) = \frac{1}{3}, \quad i \in \{1, 2, 3\}.$$

Nadalje, zaključujemo da je $P(V_3 \mid A_1 \cap I_1) = 1/2$, $P(V_3 \mid A_2 \cap I_1) = 1$ i $P(V_3 \mid A_3 \cap I_1) = 0$, iz čega slijedi da je $P_{I_1}(V_3) = P(V_3 \mid I_1) = 1/2$. Sada lako slijedi da je $P(A_2|I_1 \cap V_3) = 2/3$ te zaključujemo da je u opisanoj situaciji promjena izbora vrata mudar potez za igrača (slika 1.18).



(a) Vjerojatnosti bez sudjelovanja voditelja

(b) Vjerojatnosti sa sudjelovanjem voditelja

Slika 1.18: Vjerojatnosti ishoda u Monty Hall problemu

Analizirajte probleme vezane uz druge odabire vrata i smještaje automobila.

U primjenama često elemente potpunog sustava događaja $\{H_i : i \in I\} \subseteq \mathcal{F}$, $I \subseteq \mathbb{N}$, tj. skupove H_i zovemo **hipotezama**. Ponekad je korisno razmišljati o računanja vjerojatnosti da je postavljena hipoteza istinita ako utvrdimo da se realizirao neki događaj. Npr. ako u primjeru 1.36 utvrdimo da se u oba pokušaja okrenuo slavuj, možemo se pitati kolika je vjerojatnost da je izvučeni novčić standardan, a kolika da nije standardan. Za izračunavanje vjerojatnosi tog oblika može se koristiti rezultat poznat pod nazivom **Bayesova formula**.

Teorem 1.2. *Neka je $\{H_i : i \in I\}$, $I \subseteq \mathbb{N}$, potpun sustav događaja na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, P)$ i neka je $A \in \mathcal{F}$ događaj s pozitivnom vjerojatnosti, tj. $P(A) > 0$. Tada za svaki $i \in I$ vrijedi:*

$$P(H_i \mid A) = \frac{P(H_i)P(A \mid H_i)}{P(A)}. \tag{1.4}$$

Dokaz. Primjenom definicije uvjetne vjerojatnosti slijedi:

$$P(H_i \mid A) = \frac{P(H_i \cap A)}{P(A)} = \frac{P(H_i)P(A \mid H_i)}{P(A)}.$$

Primjer 1.38. *Za problem opisan u primjeru 1.36 pomoću Bayesove formule izračunajmo vjerojatnosti $P(H_1 \mid A)$ i $P(H_2 \mid A)$, pri čemu događaje $H_1 \mid A$ i $H_2 \mid A$ interpretiramo na sljedeći način:*

$H_1 \mid A$ – *ako je u oba bacanja novčić pao na slavuja, odabran je standardan novčić.*

$H_2 \mid A$ – *ako je u oba bacanja novčić pao na slavuja, odabran je nestandardan novčić.*

Primjenom Bayesove formule slijedi:

$$P(H_1 \mid A) = \frac{\frac{1}{2} \cdot \frac{1}{4}}{\frac{5}{8}} = \frac{1}{5}, \quad P(H_2 \mid A) = \frac{\frac{1}{2} \cdot 1}{\frac{5}{8}} = \frac{4}{5}.$$

Uočimo da je

$$P(H_1 \mid A) + P(H_2 \mid A) = \frac{1}{5} + \frac{4}{5} = 1.$$

1.10 Zadaci

Zadatak 1.1. Odredite skupove elementarnih događaja za sljedeće slučajne pokuse:

a) uzastopno bacanje pravilno izrađenog novčića dva puta,

b) bacanje jedne pravilno izrađene igraće kockice,

c) istovremeno bacanje dviju pravilno izrađenih igraćih kockica,

d) uzastopno bacanje pravilnom izrađene igraće kockice n puta, $n \in \mathbb{N}$,

e) slučajan izbor delegacije od dvaju članova iz skupa osoba $\{A, B, C, D, E, F\}$.

Rješenje:
a) $\Omega = \{(P,P), (P,G), (G,P), (G,G)\}$,
b) $\Omega = \{1, 2, 3, 4, 5, 6\}$,
c) $\Omega = \{(i,j) : i, j \in \{1, \ldots, 6\}\}$,
d) $\Omega = \{(i_1, \ldots, i_n) : i_1, \ldots, i_n \in \{1, \ldots, 6\}\}$,
e) Ω je familija svih dvočlanih podskupova skupa $\{A, \ldots, F\}$.

Zadatak 1.2. Odredite prostor elementarnih događaja slučajnog pokusa koji se sastoji od bacanja pravilno izrađenog novčića i pravilno izrađene igraće kockice redom, pri čemu se kao ishod registriraju pojava pisma ili glave na novčiću i broj na gornjoj strani kockice.

Rješenje: $\Omega = \{(i, j) : i \in \{P, G\}, j \in \{1, \ldots, 6\}\}$.

Zadatak 1.3. Odredite prostor elementarnih događaja slučajnog pokusa koji se sastoji od bacanja pravilno izrađenog novčića i odabira jednog elementa skupa $\{A, B, C, D, E\}$ redom.

Rješenje: označimo 1 realizaciju pisma i 0 realizaciju glave pri bacanju novčića; $\Omega = \{0, 1\} \times \{A, B, C, D, E\}$.

Zadatak 1.4. Odredite prostor elementranih događaja slučajnog pokusa koji se sastoji redom od bacanja pravilno izrađene igraće kockice i pravilno izrađenog tetraedra čije su strane numerirane slovima a, b, c i d.

Rješenje: $\Omega = \{(i, j) : i \in \{1, \ldots, 6\}, j \in \{a, b, c, d\}\}$.

Zadatak 1.5. Slučajan pokus sastoji se od uzastopnog bacanja pravilno izrađenog novčića sve dok ne padne pismo. Odredite prostor elementarnih događaja tog pokusa.

Rješenje: označimo 1 realizaciju pisma i 0 realizaciju glave; $\Omega = \{1, (0, 1), (0, 0, 1), (0, 0, 0, 1), \ldots\}$

Zadatak 1.6. U kutiji se nalaze četiri papirića numerirana brojevima 1, 2, 3 i 4. Iz kutije se na slučajan način izvlači jedan po jedan papirić i to:

a) bez vraćanja,

b) s vraćanjem,

sve dok se ne izvuče papirić na kojemu je neparan broj. Ako se kao ishod tog slučajnog pokusa registriraju izvučeni brojevi, odredite prostor elementarnih događaja.

Rješenje:
a) $\Omega = \{1, 3, (2, 1), (4, 1), (2, 3), (4, 3), (2, 4, 1), (2, 4, 3), (4, 2, 1), (4, 2, 3)\}$,
b) $\Omega = \{1, 3, (2, 1), (2, 2, 1), (2, 2, 2, 1), \ldots\}$.

Zadatak 1.7. Strijelac gađa metu 4 puta, pri čemu se registriraju pogoci (označeni nulom) i promašaji (označeni jedinicom). Odredite prostor elementarnih događaja i sljedeće događaje:

a) A - gađanje je započelo promašajem,

b) B - rezultat je svih gađanja isti,

c) C - cilj je pogođen dva puta,

d) D - cilj je pogođen barem dva puta.

Rješenje:
$\Omega = \{(0,0,0,0),(1,0,0,0),(0,1,0,0),(0,0,1,0),(0,0,0,1),(1,1,0,0),$
$(0,1,1,0),(0,0,1,1),(1,0,1,0),(1,0,0,1),(0,1,0,1),(1,1,1,0),$
$(0,1,1,1),(1,0,1,1),(1,1,0,1),(1,1,1,1)\}.$

Zadatak 1.8. Vitez Okruglog stola kralja Arthura, Sir Lancelot, odlučio je strijelom gađati jabuku kroz zlatni prsten u čast kraljice Guinevere. Sir Lancelot ima na raspolaganju samo luk i 4 strijele i može gađati jabuku sve dok je ne pogodi 2 puta ili ne potroši sve strijele (pogoci i promašaji ne moraju se realizirati zaredom). Odredite prostor elementarnih događaja Ω te sljedeće događaje:

- A - jabuka je promašena točno dva puta,
- B - jabuka je promašena u posljednjem pokušaju.

Jesu li događaji A i B disjunktni?

Rješenje:
$\Omega = \{(1,1),(0,1,1),(1,0,1),(0,0,1,1),(0,1,0,1),(1,0,0,1),(1,0,0,0),$
$(0,1,0,0),(0,0,1,0),(0,0,0,1),(0,0,0,0)\},$
$A = \{(1,0,0,1),(0,0,1,1),(0,1,0,1)\},$
$B = \{(0,0,0,0),(1,0,0,0),(0,1,0,0),(0,0,1,0)\}, \quad A \cap B = \emptyset.$

Zadatak 1.9. U natjecanju u streličarstvu lukom i strijelom Robin Hood gađa jednostavnu metu (jedine moguće realizacije gađanja su ili pogodak ili promašaj) sve dok ju ili ne pogodi dva puta ili ne promaši tri puta (pogoci i promašaji ne moraju se realizirati zaredom). Odredite prostor elementarnih događaja Ω te sljedeće događaje:

- A - druga odapeta strelica promašila je metu,
- B - prva i treća odapeta strelica pogodile su metu.

Jesu li događaji A i B disjunktni?

Rješenje:
$\Omega = \{(1,1),(1,0,1),(0,1,1),(0,0,0),(0,0,1,1),(0,0,1,0),$
$(1,0,0,0),(0,1,0,1),(1,0,0,1),(0,1,0,0)\}$, $A = \{(1,0,1),(0,0,0),(0,0,1,1),(0,0,1,0),(1,0,0,0),(1,0,0,1)\},$

$B = \{(1,0,1)\}, \quad B \subset A.$

Zadatak 1.10. Strijelac gađa cilj oblika kružne mete polumjera R, pri čemu se mjeri udaljenost od mjesta pogotka do središta mete. Odredite prostor elementarnih događaja.

Rješenje: označimo O promašaj mete pri gađanju; $\Omega = [0,R] \cup \{O\}$.

Zadatak 1.11. Neka je Ω prostor elementarnih događaja nekog slučajnog pokusa, te neka su A, B i C događaji ($A, B, C \subseteq \Omega$). Pomoću događaja A, B i C izrazite sljedeće događaje:

a) realizirao se samo događaj A,

b) realizirali su se događaji A i B,

c) realizirala su se sva tri događaja,

d) realizirao se barem jedan od događaja A, B i C,

e) realizirao se točno jedan od događaja A, B i C,

f) realizirali su se barem dva od događaja A, B i C,

g) realizirali su se točno dva od događaja A, B i C,

h) realizirali su se najviše dva od događaja A, B i C,

i) nije se realizirao niti jedan od događaja A, B i C.

Rješenje:
a) $A \cap B^c \cap C^c$,
b) $A \cap B$,
c) $A \cap B \cap C$,
d) $A \cup B \cup C$,
e) $(A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C)$,
f) $(A \cap B) \cup (A \cap C) \cup (B \cap C)$,
g) $(A \cap B \cap C^c) \cup (A \cap B^c \cap C) \cup (A^c \cap B \cap C)$,
h) $A^c \cup B^c \cup C^c$,
i) $(A \cup B \cup C)^c$.

Zadatak 1.12. Među studentima se okupljenima na predavanju slučajno bira jedan student. Promatramo sljedeće događaje:

- A - student je muškog spola,
- B - student je vegetarijanac,
- C - student živi u studentskom domu.

a) Opišite događaj $A \cap B \cap C$.

b) Kada će vrijediti $A \cap B \cap C = A$?

c) Kada će vrijediti $C^c \subseteq B$?

d) Kada će vrijediti $A^c = B$? Vrijedi li nužno ta jednakost ako su svi studenti muškog spola vegetarijanci?

Rješenje:

a) Student je muškog spola, vegetarijanac je i živi u studentskom domu.

b) Vrijedi kad su svi studenti muškog spola vegetarijanci i žive u studentskom domu.

c) Vrijedi kad su svi studenti koji ne žive u studentskom domu vegetarijanci.

d) $A^c = B$ *vrijedi ako su svi studenti vegetarijanci ženskog spola (*$B \subseteq A^c$*) i ako su sve studentice vegetarijanci (*$A^c \subseteq B$*). Ako su svi studenti muškog spola vegetarijanci (tj. ako je* $A \subseteq B$*) zaključujemo sljedeće:*

- $A^c \subseteq B$ *samo ako su svi studenti (bez obzira na spol) vegetarijanci,*
- $B \subseteq A^c$ *u opisanoj situaciji ne vrijedi,*

što znači da ako je $A \subseteq B$, skupovi A^c i B nisu jednaki.

Zadatak 1.13. Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor. Ako su $A, B \in \mathcal{F}$ takvi da je $P(A \triangle B) = 0$, dokažite da je $P(A) = P(B)$.

Uputa: $A \triangle B = (A \setminus B) \cup (B \setminus A)$.

Zadatak 1.14. Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor te neka za $A, B \in \mathcal{F}$ vrijedi da je $P(A \triangle B) = 0$ i $P(A) = p$. Izračunajte:

a) $P(A \cap B)$ i $P(A \cup B)$,

b) $P(A \setminus B)$ i $P(B \setminus A)$.

Zadatak 1.15. Dokažite da nezavisnost događaja A i B, $A, B \in \mathcal{F}$, povlači nezavisnost

a) događaja A i B^c,

b) događaja A^c i B^c.

Zadatak 1.16. Na raspolaganju nam je kutija u kojoj se nalazi 100 papirića numeriranih brojevima $1, 2, \ldots, 100$. Slučajan pokus sastoji se od izvlačenja jednog papirića iz kutije. Upotrebom klasične definicije vjerojatnosti odredite vjerojatnosti sljedećih događaja:

a) A - izvučeni je broj jednoznamenkast,

b) B - izvučeni je broj dvoznamenkast,

c) C - izvučeni je broj manji ili jednak broju k, $k \in \{1, 2, \ldots, 100\}$,

d) D - izvučeni je broj strogo veći od k, $k \in \{1, 2, \ldots, 100\}$,

e) E - suma je znamenaka izvučenog broja 3,

f) F - umnožak je znamenaka izvučenog broja 6.

Rješenje:

a) $P(A) = 0.09$,	b) $P(B) = 0.9$,	c) $P(C) = k/100$,
d) $P(D) = (100 - k)/100$,	e) $P(E) = 0.04$,	f) $P(F) = 0.04$.

Zadatak 1.17. Pravilno izrađena igraća kockica baca se dva puta. Upotrebom klasične definicije vjerojatnosti odredite vjerojatnosti sljedećih događaja:

a) A - pali su jednaki brojevi,

b) B - suma je brojeva koji su pali 8,

c) C - produkt je brojeva koji su pali 8,

d) D - suma brojeva koji su pali veća je od produkta brojeva koji su pali,

e) E - produkt brojeva koji su pali veći je od sume brojeva koji su pali.

Rješenje:
a) $P(A) = 1/6$, b) $P(B) = 5/36$, c) $P(C) = 1/18$,
d) $P(D) = 11/36$, e) $P(E) = 2/3$.

Zadatak 1.18. Pretpostavimo da u pošiljci od ukupno 500 jabuka ima 2% prezrelih jabuka. Kolika je vjerojatnost da slučajan uzorak od 20 jabuka uzet iz te pošiljke sadrži točno dvije prezrele jabuke?

Rješenje: $\dfrac{\binom{10}{2}\binom{490}{18}}{\binom{500}{20}}$.

Zadatak 1.19 (Problem rođendana). Kolika je vjerojatnost da između n osoba, $n \leq 365$, barem dvije osobe imaju rođendan istog datuma? Pretpostavljamo da svaka od n osoba može biti rođena s jednakom vjerojatnošću bilo kojeg dana u godini i zanemarujemo postojanje prijestupnih godina.

Rješenje: $1 - \dfrac{364 \cdots (366-n)}{365^{n-1}}$.

Zadatak 1.20. U kutiji se nalazi a crvenih i b zelenih kuglica ($a \geq 2$, $b \geq 2$). Iz kutije na slučajan način istovremeno izvadimo dvije kuglice. Definirajmo sljedeće događaje:

A – izvučene su kuglice iste boje,
B – izvučene su kuglice različitih boja.

Koji je od događaja A i B vjerojatniji?

Rješenje: $P(A) = \dfrac{\binom{a}{2} + \binom{b}{2}}{\binom{a+b}{2}}$, $P(B) = \dfrac{ab}{\binom{a+b}{2}}$.

Zadatak 1.21. Pravilno izrađen novčić bacamo 10 puta zaredom. Kolika je vjerojatnost da se pismo realizira točno tri puta?

Rješenje: $\dfrac{120}{2^{10}}$.

Zadatak 1.22. Pravilno izrađenu igraću kockicu bacamo 10 puta. Kolika je vjerojatnost da se kao rezultat bacanja brojevi $1, 2, 3, 4, 5, 6$ pojave redom $2, 3, 1, 1, 1, 2$ puta?

Rješenje: $\dfrac{10!}{4 \cdot 6^{11}}$.

Zadatak 1.23. U autobusu se nalazi 15 ljudi. Do kraja putovanja ostale su 4 stanice. Kolika je vjerojatnost da svi putnici izađu na istoj stanici?

Rješenje: 4^{-14}.

Zadatak 1.24. Pretpostavimo da n ljudi na slučajan način sjeda za okrugli stol. Izračunajte vjerojatnost da će dva unaprijed odabrana čovjeka sjediti zajedno.

Rješenje: $\frac{2}{n-1}$.

Zadatak 1.25. m muškaraca i n žena raspoređuju se na slučajan način u kazalištu na $(m+n)$ sjedala složenih u redu. Kolika je vjerojatnost da sve žene sjede zajedno (tj. jedna do druge)?

Rješenje: $\frac{n!(m+1)!}{(m+n)!} = \frac{m+1}{\binom{m+n}{n}}$.

Zadatak 1.26. Svežanj od 52 karte podijeli se na dva jednakobrojna dijela. Odredite vjerojatnost sljedećih događaja:

a) A - u svakom dijelu nalaze se po dva kralja,

b) B - u jednom dijelu ne nalazi se niti jedan kralj,

c) C - u jednom dijelu nalazi se jedan, a u drugom dijelu tri kralja.

Rješenje: $P(A) = \frac{\binom{4}{2}\binom{48}{24}}{\binom{52}{26}}$, $P(B) = \frac{2\binom{48}{26}}{\binom{52}{26}}$, $P(C) = \frac{8\binom{48}{25}}{\binom{52}{26}}$.

Zadatak 1.27. Iz špila od 52 karte na slučajan način biramo 8 karata. Izračunajte vjerojatnost da su izvučena

a) tri asa,

b) tri kralja,

c) tri asa ili tri kralja.

Rješenje: a) $\frac{\binom{4}{3}\binom{48}{5}}{\binom{52}{8}}$, c) $\frac{2\binom{4}{3}\binom{48}{5} - \binom{4}{3}^2\binom{44}{2}}{\binom{52}{8}}$.

Zadatak 1.28. Student je došao na ispit znajući odgovore na 90 od 100 pitanja. Izvlači se pet pitanja.

a) Kolika je vjerojatnost da će student znati odgovore na svih pet pitanja?

b) Kolika je vjerojatnost da će student znati odgovore na barem tri od pet izvučenih pitanja?

Rješenje: a) $\frac{\binom{90}{5}}{\binom{100}{5}} \approx 0.584$, b) $\frac{\binom{90}{5} + 10\binom{90}{4} + \binom{10}{2}\binom{90}{3}}{\binom{100}{5}} \approx 0.99$.

Zadatak 1.29. Iz šešira u kojem se nalazi n kuglica na slučajan način izaberemo nekoliko kuglica. Kolika je vjerojatnost da smo izvukli paran broj kuglica?

Rješenje: $\dfrac{2^{n-1}-1}{2^n-1}$.

Zadatak 1.30 (De Mereov paradoks). Bacamo tri različite pravilno izrađene igraće kockice. Zanima nas vjerojatnost sljedećih događaja:

(a) A - zbroj brojeva na svim trima kockicama jednak je 11,

(b) B - zbroj brojeva na svim trima kockicama jednak je 12.

Rješenje: $P(A) = \dfrac{27}{6^3}, \quad P(B) = \dfrac{25}{6^3}$.

Zadatak 1.31. Standardni svežanj od 52 karte sastoji se od 26 crvenih karata (13 hercova i 13 karo karata) i 26 crnih karata (13 pikova i 13 trefova). Na slučajan način iz špila biramo 11 karata. Odredite prostor elementarnih događaja Ω i izračunajte vjerojatnosti sljedećih događaja:

a) A - izvučena je barem jedna crvena karta,

b) B - izvučeno je točno pet crnih karata ili točno sedam hercova.

Rješenje: $P(A) = \sum\limits_{k=1}^{26} \dfrac{\binom{26}{k}\binom{26}{11-k}}{\binom{52}{11}}, \quad P(B) = \dfrac{\binom{26}{5}\binom{26}{6}+\binom{13}{7}\binom{39}{4}}{\binom{52}{11}}$.

Zadatak 1.32. U šeširu se nalazi 100 listića s oznakama od 00 do 99. Ako je xy (ne radi se o umnošku $x \cdot y$) oznaka na slučajno odabranom listiću, odredite vjerojatnosti događaja $A = \{xy : x \cdot y < 5\}$, $B = \{xy : x + y = 11\}$ te događaja $C = \{xy : xy > 60\}$ pod uvjetom da se realizirao događaj $D = \{xy : y > 7\}$.

Rješenje: $P(A) = 27/100$, $P(B) = 2/25$, $P(C|D) = 2/9$.

Zadatak 1.33. Pravilno izrađen novčić bacamo n puta. Odredite vjerojatnost da se pismo realizira u točno m bacanja, $m \leq n$.

Rješenje: $\dfrac{\binom{n}{m}}{2^n}$.

Zadatak 1.34. Odredite vjerojatnost da rođendani 12 nasumično odabranih ljudi padnu u različite mjesece u godini?

Rješenje: $\dfrac{11!}{12^{11}}$.

Zadatak 1.35. U kutiji se nalazi $(n+m)$ srećki od kojih n dobiva. Na slučajan način odabere se k srećki. Kolika je vjerojatnost da među njima bude j dobitnih?

Rješenje: $\dfrac{\binom{n}{j}\binom{m}{k-j}}{\binom{n+m}{k}}$.

Zadatak 1.36. Odredite vjerojatnost da je slučajno odabrana točka iz segmenta $[0,1]$ racionalna?

Rješenje: vjerojatnost odabira racionalne točke iz segmenta $[0,1]$ jest nula, tj. slučajno će odabrana točka iz segmenta $[0,1]$ gotovo sigurno biti iracionalna.

Zadatak 1.37. Za fiksiranu točku $c \in [a,b]$ i slučajno odabranu točku $x \in [a,b]$ odredite vjerojatnosti sljedećih događaja:

a) $x \leq c$,

b) $x < c$,

c) $x = c$,

d) x je bliže točki a nego točki b.

Rješenje: a) $\dfrac{c-a}{b-a}$, b) $\dfrac{c-a}{b-a}$, c) 0, d) 0.5.

Zadatak 1.38. Neka su x i y dva slučajno odabrana broja iz segmenta $[0,1]$. Odredite vjerojatnost da za njih vrijedi:

a) $x > y$,

b) $x + y < 3/2$,

c) $x = y$,

d) $xy \leq 2/9$ i $x + y < 1$.

Rješenje: a) $1/2$, b) $7/8$, c) 0, d) $\dfrac{1}{3} + \dfrac{2}{9}\ln 2$.

Zadatak 1.39. Neka su x i y dva slučajno odabrana broja iz segmenta $[-2,2]$. Odredite vjerojatnost da za njih vrijedi $y \leq x^2$ i $x < \frac{1}{2}y + 1$.

Zadatak 1.40. Dva broda A i B trebaju stići u istu luku. Njihova su vremena dolaska u luku slučajna, međusobno nezavisna i jednako vjerojatna tijekom 24 sata. Kad brod A stigne u luku, on ostaje u njoj 1 sat, a brod B 2 sata. Budući da luka ne može odjednom primiti dva broda, odredite vjerojatnost da jedan od brodova čeka na oslobađanje luke dok drugi brod ne ode.

Rješenje: 0.121.

Zadatak 1.41. Na kružnici polumjera $r = 2\,cm$ na slučajan način biramo dvije točke. Kolika je vjerojatnost da će udaljenost među tim točkama mjerena dužinski (ne po kružnici) biti najviše $2\,cm$?

Rješenje: $1/3$.

Zadatak 1.42. Na kružnici polumjera r na slučajan način odabrane su tri točke: A, B i C. Kolika je vjerojatnost da je kružnici upisan trokut ABC šiljastokutan?

Rješenje: $1/4$. Pretpostavite da je položaj točke A na kružnici fiksan, promatrajte duljine kružnih lukova AB i BC te tako određen prostor elementarnih događaja skicirajte u Kartezijevom koordinatnom sustavu.

Zadatak 1.43. Izračunajte vjerojatnost da korijeni kvadratne jednadžbe $x^2 + px + q = 0$ budu realni, ako se koeficijenti p i q izabiru na slučajan način iz segmenta $[-1, 1]$.

Rješenje: $13/24$.

Zadatak 1.44. U pravokutniku kojemu se stranice odnose kao $1 : 2$ na slučajan način biramo jednu točku. Kolika je vjerojatnost da je točka pala:

a) na dijagonalu pravokutnika,

b) bliže duljoj stranici pravokutnika?

Rješenje: a) 0, b) $3/4$.

Zadatak 1.45. Na dužini $\overline{AB}$ duljine a na slučajan način izabrane su točke M i N. Odredite vjerojatnost da je točka M bliže točki A nego točki N.

Rješenje: $1/4$.

Zadatak 1.46. Kolika je vjerojatnost da slučajno odabrana točka u kvadratu bude bliža jednoj od dijagonala nego stranicama kvadrata?

Rješenje: $\dfrac{\sqrt{2}}{1+\sqrt{2}}$.

Zadatak 1.47. Vjerojatnost da se knjiga nalazi u knjižnici iznosi p. Ako je knjiga u knjižnici, nalazi se na jednoj od n polica s jednakom vjerojatnošću (za svaku od tih n polica). Pregledano je m polica, $m < n$, i knjiga nije nađena. Kolika je vjerojatnost da je knjiga ipak u knjižnici?

Rješenje: $\dfrac{(n-m)p}{n-mp}$.

Zadatak 1.48. Pretpostavimo da na slučajan način odabiremo realne brojeve $x \in [0, 1]$ i $y \in [0, 2]$. Kolika je vjerojatnost da je njihov zbroj veći od dva, a produkt manji od jedan?

Rješenje: $\dfrac{1}{2} \ln 2 - \dfrac{1}{4}$.

Zadatak 1.49. Točka A bira se na slučajan način unutar jednakostraničnog trokuta sa stranicom duljine 3 cm. Kolika je vjerojatnost da udaljenost točke A od bilo kojeg vrha trokuta bude veća od 1 cm?

Rješenje: $\dfrac{2\pi}{9\sqrt{3}}$.

Zadatak 1.50. Zadano nam je pet dužina čije su duljine 2 cm, 4 cm, 5 cm, 7 cm i 9 cm. Kolika je vjerojatnost da se od tri slučajno odabrane dužine može konstruirati trokut?

Rješenje: $2/5$.

Zadatak 1.51. Točka A bira se na slučajan način unutar raznostraničnog pravokutnog trokuta. Kolika je vjerojatnost da je odabrana točka A bliža hipotenuzi nego katetama?

Rješenje: duljinu hipotenuze označimo c, a duljine kateta a i b; tada je tražena vjerojatnost $\dfrac{c}{a+b+c}$.

Zadatak 1.52. Trenutak u kojem će neki signal stići do prijemnika slučajno je odabrani trenutak iz intervala $[0, T]$. Prijemnik neće registrirati drugi signal ako je razlika između dva uzastopna signala manja od τ, $\tau < T$. Odredite vjerojatnost da prijemnik neće registrirati drugi signal.

Rješenje: $\dfrac{\tau(2T-\tau)}{T^2}$.

Zadatak 1.53 (Buffonov problem). Ravnina je podijeljena paralelnim pravcima koji su jedan od drugog udaljeni za $2a$. Na tu se ravninu na slučajan način baca igla duljine $2l$, $l < a$. Odredite vjerojatnost da igla siječe neki od pravaca.

Rješenje: $\dfrac{2l}{a\pi}$.

Zadatak 1.54. Slučajan pokus sastoji se od bacanja pravilno izrađenog novčića tri puta za redom. Želimo naći vjerojatnost događaja A uz dani događaj B kada su A i B sljedeći događaji:

- A - glava je pala više puta nego pismo,
- B - prvo je palo pismo.

Rješenje: $P(A|B) = 0.25$.

Zadatak 1.55. Iz kutije u kojoj se nalazi m crvenih i n zelenih kuglica na slučajan način izvučeno je k kuglica. Uz pretpostavku da su sve izvučene kuglice iste boje, kolika je vjerojatnost da su sve zelene?

Rješenje: $\dfrac{\binom{n}{k}}{\binom{m}{k}+\binom{n}{k}}$.

Zadatak 1.56. Slučajan pokus sastoji se od slučajnog odabira obitelji koja ima dvoje djece (uz pretpostavku da su vjerojatnosti rođenja kćerke i sina jednake (što približno odgovara stvarnoj situaciji) te da znamo redoslijed rađanja djece u obitelji). Treba provjeriti jesu li sljedeći događaji nezavisni:

- A - u obitelji je barem jedna kćerka,
- B - u obitelji su jedna kćerka i jedan sin.

Rješenje: A i B nisu nezavisni događaji.

Zadatak 1.57. Cilj se gađa iz tri topa. Topovi pogađaju cilj nezavisno jedan od drugoga s vjerojatnošću 0.4. Ako jedan top pogodi cilj uništava ga s vjerojatnošću 0.3, ako ga pogode dva topa uništavaju ga s vjerojatnošću 0.7, a ako ga pogode tri topa uništavaju ga s vjerojatnošću 0.9. Nađite vjerojatnost uništenja cilja.

Rješenje: 0.3888.

Zadatak 1.58. Imamo dvije kutije. U prvoj se nalazi a bijelih i b crnih kuglica, a u drugoj c bijelih i d crnih kuglica. Iz prve kutije na slučajan način biramo k kuglica, a iz druge m kuglica. Tih $(k+m)$ kuglica izmiješamo i stavimo u treću kutiju.

a) Iz treće kutije na slučajan način izvučemo jednu kuglicu. Kolika je vjerojatnost da ona bude bijele boje?

b) Iz treće kutije na slučajan način izvučemo tri kuglice. Kolika je vjerojatnost da je barem jedna kuglica bijele boje?

Rješenje:
a) $\dfrac{1}{k+m}\left(\dfrac{ak}{a+b}+\dfrac{cm}{c+d}\right)$

b) Pomoću formule potpune vjerojatnosti odrediti vjerojatnost suprotnog događaja.

Zadatak 1.59. Na usmenom ispitu iz Uvoda u vjerojatnost i statistiku ponuđeno je 10 pitanja od kojih je student naučio njih n, $0 \leq n \leq 10$. Student će položiti ispit ako bude točno odgovorio na dva slučajno odabrana pitanja ili ako točno odgovori na jedno od njih te na treće, dodatno postavljeno pitanje. Na koliko pitanja student treba znati točno odgovoriti da bi s vjerojatnošću većom od 0.8 položio ispit?

Rješenje: student treba znati točno odgovoriti na bar sedam pitanja.

Zadatak 1.60. Ptica slijeće u slučajno izabrano gnijezdo od ukupno tri gnijezda koja su joj na raspolaganju. Svako gnijezdo sadrži dva jaja i to: u prvom gnijezdu su oba jaja zdrava, u drugom je jedno zdravo i jedan mućak, a u trećem su oba jaja mućka. Nađite vjerojatnost da ptica sjedi na mućku. Ako je sjela na mućak, kolika je vjerojatnost da sjedi u drugom gnijezdu?

Rješenje: ako je sjela na mućak, vjerojatnost da sjedi na drugom jajetu jest $1/3$.

Zadatak 1.61. Neki izvor emitira poruke koje se sastoje od znakova 0 i 1. Vjerojatnost emitiranja znaka 1 jest 0.6, a vjerojatnost emitiranja znaka 0 jest 0.4. Na izlazu se iz kanala 10% znakova pogrešno interpretira. Ako je primljena poruka 101, kolika je vjerojatnost da je ona i poslana?

Rješenje: 0.743.

Zadatak 1.62. Svi studenti iz grupe koja se sastoji od a odličnih, b vrlo dobrih i c dobrih studenata prijavili su ispit iz Uvoda u vjerojatnost i statistiku. Odličan student na ispitu može dobiti samo odličnu ocjenu, vrlo dobar student s jednakom vjerojatnošću može dobiti odličnu ili vrlo dobru ocjenu, a dobar student s jednakom vjerojatnošću može dobiti vrlo dobru, dobru ili dovoljnu ocjenu. Na ispitu se na slučajan način proziva jedan student iz grupe. Kolika je vjerojatnost da on dobije odličnu ili vrlo dobru ocjenu?

Rješenje: $\dfrac{3a+3b+c}{3(a+b+c)}$.

Zadatak 1.63. Iz šešira u kojem se nalazi n kuglica na slučajan način izvlačimo jednu kuglicu.

a) Kolika je vjerojatnost da će ta kuglica biti bijela ako su sve pretpostavke o prethodnom broju bijelih kuglica jednako vjerojatne?

b) Ako je izvučena bijela kuglica, kolika je vjerojatnost da se sve kuglice u šeširu bijele?

Rješenje: a) $1/2$; b) $2/(n+1)$.

Zadatak 1.64. U mračnoj prostoriji nalaze se četiri svjetiljke i sedam žarulja. Od četiri svjetiljke ispravne su dvije, a od sedam žarulja ispravne su četiri. Ako na slučajan način odaberemo četiri žarulje i stavimo ih u svjetiljke, kolika je vjerojatnost da će prostoriju obasjati svjetlo?

Rješenje: $6/7$.

Zadatak 1.65. Dane su nam dvije kutije šibica. U prvoj kutiji nalazi se 12 šibica od kojih je jedna slomljena, a u drugoj kutiji 10 šibica od kojih je jedna slomljena. Uzimamo jednu šibicu iz prve kutije i prebacimo ju u drugu kutiju. Nakon toga iz druge kutije na slučajan način biramo jednu šibicu. Kolika je vjerojatnost da smo izabrali slomljenu šibicu?

Rješenje: $13/132$.

Zadatak 1.66. Na raspolaganju imamo dvije kutije: u prvoj se nalaze 2 bijele i 4 plave, a u drugoj 3 bijele i 2 plave kuglice. Iz prve kutije na slučajan način, nezavisno jednu od druge, odaberemo dvije kuglice i prebacimo ih u drugu kutiju. Kolika je vjerojatnost da potom odabrana kuglica iz druge kutije bude bijele boje?

Rješenje: $11/21$.

Zadatak 1.67. Iz šešira u kojem se nalazi n kuglica izvlačimo nasumično jednu kuglicu. Kolika je vjerojatnost da će ta kuglica biti bijela, ako su sve pretpostavke o prethodnom broju kuglica jednako vjerojatne? Nakon što je izvučena bijela kuglica, kolika je vjerojatnost da su sve kuglice u šeširu bijele?

Rješenje: $\dfrac{2}{n+1}$.

Zadatak 1.68. U tri kutije nalaze se žute i zelene kuglice. U prvoj kutiji nalazi se 5 žutih i 3 zelene kuglice, u drugoj kutiji nalaze se 4 žute i 4 zelene kuglice, a u trećoj 3 žute i 5 zelenih kuglica. Na slučajan način iz svake kutije izaberemo po četiri kuglice i stavimo u četvrtu kutiju. Zatim iz četvrte kutije na slučajan način uzmemo jednu kuglicu. Kolika je vjerojatnost da je ta kuglica zelene boje?

Zadatak 1.69. U šeširu se nalaze 4 bijele kuglice. Tri se puta na slučajan način izvlači jedna kuglica i zamjenjuje crnom kuglicom. Potom je na slučajan način izvučena jedna kuglica i vidjelo se da je bijela. Koliko iznosi vjerojatnost da su u šeširu točno dvije crne kuglice nakon opisanih zamjena?

Zadatak 1.70. Slučajan pokus zamišljen je na sljedeći način:

- iz skupa $S = \{1, 6, 7, 8, 9\}$ na slučajan način odaberemo jedan broj,
- od preostalih elemenata skupa S slučajno odaberemo još jedan broj.

Kolika je vjerojatnost da je drugi izabrani broj neparan?

Zadatak 1.71. Četiri strijelca gađaju metu (pretpostavljamo da je meta jednostavna, tj. da su moguće realizacije gađanja ili pogodak ili promašaj). Vjerojatnosti pogotka za pojedine strijelce iznose redom 0.6, 0.7, 0.8, 0.9.

a) Izračunajte vjerojatnost da barem tri strijelca pogode metu.

b) Ako je meta pogođena s tri hica, izračunajte vjerojatnost da su ju pogodili prvi i drugi strijelac.

Zadatak 1.72. Grupa od 10 studenata došla je na ispit na kojemu svaki student odgovara na tri slučajno odabrana pitanja od 20 ponuđenih. Tri su studenta ispit pripremila odlično, četvorica dobro, dvojica dovoljno i jedan loše. Odlično pripremljen student zna odgovore na svih 20 pitanja, dobro pripremljen student na 16, dovoljno pripremljen student na 10 i loše pripremljen student na 5 pitanja. Kolika je vjerojatnost da je slučajno odabrani student točno odgovorio na sva tri postavljena pitanja?

Zadatak 1.73. Pokažite da sljedećim realnim funkcijama realne varijable možemo definirati vjerojatnost na $\mathbb{R}$:

a) $f(x) = \begin{cases} \frac{3}{4}(1-x^2) & , x \in [-1,1] \\ 0 & , x \neq [-1,1] \end{cases}$.

b) $f(x) = \begin{cases} \lambda e^{-\lambda x} & , x \in (0,\infty) \\ 0 & , x \in (-\infty, 0] \end{cases}$.

c) Koristeći funkcije iz zadataka a) i b) odredite vjerojatnost događaja

$$A = \{x \in \mathbb{R} : -1/2 < x \leq 1/2\} = (-1/2, 1/2].$$

Rješenje: obje su funkcije su nenegativne i normirane, pa pomoću njih možemo definirati vjerojatnost na $\mathbb{R}$. Vjerojatnost događaja A je: a) $P(A) = 11/16$, b) $P(A) = 1 - e^{-\lambda/2}$.

Zadatak 1.74. Pokažite da funkcijom $f\colon \mathbb{R}^2 \to \mathbb{R}$ definiranom formulom

$$f(x,y) = \begin{cases} \frac{3}{5}(x^2+y) & , (x,y) \in [-1,1] \times [0,1] \\ 0 & , (x,y) \neq [-1,1] \times [0,1] \end{cases}$$

možemo definirati vjerojatnost na $\mathbb{R}^2$. Izračunajte vjerojatnost da slučajno odabrana točka iz $\mathbb{R}^2$ pripada pravokutniku

$$A = [0,1] \times [0,1/2].$$

Rješenje: funkcija f nenegativna je i normirana, stoga pomoću nje možemo definirati vjerojatnost na $\mathbb{R}^2$. Vjerojatnost događaja A jest $P(A) = 7/40$.

Poglavlje 2

Slučajna varijabla

Proučavanje cijelog prostora elementarnih događaja nekog slučajnog pokusa i vjerojatnosti zadane na njemu može biti vrlo složen zadatak. Skupovi elementarnih događaja mogu sadržavati razne objekte, npr. brojeve, uređene parove, uređene n-torke, nizove brojeva, slova, nizove slova, boje, itd. Međutim, vrlo se često naš interes za slučajan pokus svodi samo na proučavanje jedne ili nekoliko karakteristika koje možemo opisati (ili kodirati) numerički.

Primjer 2.1. *Ispit se sastoji od* 10 *pitanja višestrukog izbora. Ispit je za kandidata slučajan pokus. Prostor elementarnih događaja takvog pokusa jest skup uređenih desetorki s oznakama koje se koriste za "točan odgovor", odnosno "netočan odgovor". Međutim, prilikom polaganja ispita kandidata prvenstveno zanima koliko će ukupno imati točnih odgovora, a ne i koji će po redu odgovor biti točan. Dakle, njega zanima modeliranje na skupu mogućih ishoda* $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$*. Osim toga, zanima ga i koliku će ocjenu postići, što znači da ga zanima i skup mogućih ishoda* $\{1, 2, 3, 4, 5\}$.

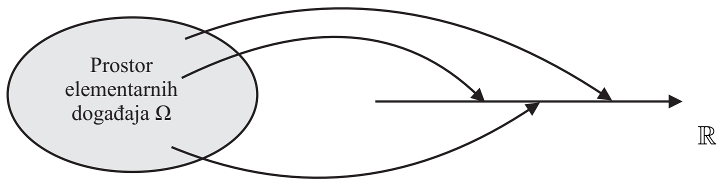
Da bismo za dani slučajni pokus analizirali i modelirali pojedinačne slučajne karakteristike ili više njih, u ovom poglavlju definirat ćemo pojam slučajne varijable, a u sljedećem pojam slučajnog vektora.

2.1 Diskretna slučajna varijabla

Kao što je već rečeno, diskretan vjerojatnosni prostor karakteriziran je činjenicom da Ω ima konačno ili prebrojivo mnogo elemenata. U poglavlju 1.6 označili smo ga $\Omega = \{\omega_i : i \in I_\Omega\}$, $I_\Omega \subseteq \mathbb{N}$. Pridružena je σ-algebra tada točno jednaka partitivnom skupu od Ω, tj. $\mathcal{F} = \mathcal{P}(\Omega)$, a vjerojatnost se može odrediti zadavanjem vrijednosti samo na jednočlanim podskupovima od Ω. Ako Ω nije diskretan (primjerice $\Omega = \mathbb{R}$), nije moguće definirati vjerojatnost samo zadavanjem vrijednosti na elementima od

Ω. Zbog toga ćemo i izučavanje slučajne varijable razdvojiti na takozvane diskretne i apsolutno neprekidne slučajne varijable. Pri tome ćemo diskretne slučajne varijable prirodno vezati uz diskretan vjerojatnosni prostor (iako to nije općenito nužno).

Definicija 2.1. *Neka je dan diskretan vjerojatnosni prostor* $(\Omega, \mathcal{P}(\Omega), P)$. *Svaku funkciju* $X : \Omega \to \mathbb{R}$ *zvat ćemo* **diskretna slučajna varijabla**.



Slika 2.1: Prikaz djelovanja slučajne varijable.

Primjer 2.2. *Neka se na skladištu nalaze četiri istovrsna proizvoda. Vjerojatnost prodaje jednog od ponuđenih proizvoda u jednom danu iznosi* $1/2$. *Broj je prodanih proizvoda u jednom danu jedna slučajna varijabla. Označimo je sa* X.

Uočimo da se u tom primjeru Ω *sastoji od uređenih četvorki nula i jedinica koje označavaju je li taj dan pojedini proizvod prodan ili ne, tj.* $\Omega = \{(i, j, k, l) : i, j, k, l \in \{0, 1\}\}$. *Vjerojatnost svakog elementa iz* Ω *iznosi* $1/2^4$. *Međutim, nama je zanimljivo koliko je proizvoda prodano, ali nam nije bitno o kojim se proizvodima radi. Zato koristimo slučajnu varijablu*

$$X : \Omega \to \mathbb{R}, \quad X(i, j, k, l) = i + j + k + l.$$

Skup $\{(i, j, k, l) \in \Omega : X(i, j, k, l) = 2\}$ *predstavlja moguće ishode iz* Ω *kod kojih su prodana točno 2 proizvoda. Obzirom da nas zapravo zanima vjerojatnost realizacije takvih skupova, a njihovo je označavanje komplicirano, ovdje ćemo koristiti oznaku za skup:*

$$\{(i, j, k, l) \in \Omega : X(i, j, k, l) = 2\} = \{X = 2\},$$

a za vjerojatnost

$$P(\{(i, j, k, l) \in \Omega : X(i, j, k, l) = 2\}) = P\{X = 2\}.$$

Oznake analogne onima iz primjera 2.2 koristimo općenito kad računamo sa slučajnim varijablama. Tako ćemo za $A \subseteq \mathbb{R}$ uvesti sljedeću oznaku:

$$\{X \in A\} = \{\omega \in \Omega : X(\omega) \in A\},$$

a za vjerojatnost tog skupa

$$P\{X \in A\} = P(\{\omega \in \Omega : X(\omega) \in A\}).$$

Primjerice, za dani $x \in \mathbb{R}$ koristimo oznaku

$$\{\omega \in \Omega : X(\omega) = x\} = \{X = x\}.$$

Primjer 2.3. *Vratimo se primjeru 2.2. Skup je svih mogućih realizacija slučajne varijable X koja broji prodane proizvode skup $\{0, 1, 2, 3, 4\}$.*
Naime, vjerojatnosna svojstva te slučajne varijable određena su ako ulogu prostora elementarnih događaja preuzme $\mathcal{R}(X) = \{0, 1, 2, 3, 4\}$. Na njemu vjerojatnost zadajemo nizom brojeva:

$$p_0 = P\{X = 0\}, \qquad p_1 = P\{X = 1\}, \qquad p_2 = P\{X = 2\},$$
$$p_3 = P\{X = 3\}, \qquad p_4 = P\{X = 4\}.$$

Prebrojavanjam elementarnih događaja iz Ω (primjer 2.2) te primjenom aditivnosti vjerojatnosti slijedi:

$$p_0 = \frac{1}{16}, \quad p_1 = \frac{1}{4}, \quad p_2 = \frac{3}{8}, \quad p_3 = \frac{1}{4}, \quad p_4 = \frac{1}{16}.$$

Radi preglednosti ćemo tu diskretnu slučajnu varijablu zapisati pomoću sljedeće tablice:

$$X = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{1}{16} & \frac{1}{4} & \frac{3}{8} & \frac{1}{4} & \frac{1}{16} \end{pmatrix}.$$

Pregledan zapis slučajne varijable prezentiran u prethodnom primjeru koristit ćemo i općenito. Dakle, ako je X dana diskretna slučajna varijabla na vjerojatnosnom prostoru $(\Omega, \mathcal{P}(\Omega), P)$, označimo skup svih vrijednosti koje ona može primiti kao $\mathcal{R}(X) = \{x_i, i \in I\}$, $I \subseteq \mathbb{N}$, a pripadne vjerojatnosti nizom brojeva $(p_i, i \in I)$ za koji vrijedi:

$$p_i = P\left(\{\omega \in \Omega : X(\omega) = x_i\}\right) = P\{X = x_i\}, \quad i \in I.$$

Bez smanjenja općenitosti, u nastavku razmatranja pretpostavit ćemo da je $I = \mathbb{N}$.

Slučajnu varijablu X prikazujemo u obliku tablice:

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n & \cdots \\ p_1 & p_2 & p_3 & \cdots & p_n & \cdots \end{pmatrix} \tag{2.1}$$

i taj prikaz zovemo **zakon razdiobe**, **tablica distribucije** ili, kraće, **distribucija** slučajne varijable X.

Dakle, za svaku diskretnu slučajnu varijablu možemo istaknuti dva bitna skupa brojeva. Jedan čine sve vrijednosti koje slučajna varijabla X može primiti, tj. skup $\mathcal{R}(X) = \{x_i : i \in \mathbb{N}\}$ kojega zovemo slika slučajne varijable, a drugi je niz pripadnih vjerojatnosti, tj. $(p_i, i \in \mathbb{N})$ takav da vrijedi $p_i = P\{X = x_i\}$, $i \in \mathbb{N}$.

Uočimo da za elemente od $\mathcal{R}(X)$ vrijedi $x_i \neq x_j$ za $i \neq j$ s obzirom da je to skup, dok u nizu $(p_i, i \in \mathbb{N})$ može biti istih elemenata, ali on ima druga dva bitna svojstva:

1. $0 \leq p_i \leq 1$ za svaki $i \in \mathbb{N}$,
2. $\sum\limits_{i=1}^{\infty} p_i = 1$.

Zaista, budući da je X slučajna varijabla koja prima vrijednosti iz skupa $\mathcal{R}(X)$, to je $P\{X \in \mathcal{R}(X)\} = P(\Omega) = 1$, pa iz svojstava vjerojatnosti slijedi:

$$0 \leq p_i = P\{X = x_i\} \leq 1.$$

Osim toga je

$$\sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} P\{X = x_i\} = P\left(\bigcup_{i=1}^{\infty} \{X = x_i\}\right) = P\{X \in \mathcal{R}(X)\} = 1.$$

Koristeći rezultate poglavlja 1.6 znamo da je pomoću niza realnih brojeva $(p_i, i \in \mathbb{N})$ koji zadovoljava svojstva

1. $0 \leq p_i \leq 1$ za svaki $i \in \mathbb{N}$,
2. $\sum\limits_{i=1}^{\infty} p_i = 1$

dobro definirana vjerojatnost na diskretnom vjerojatnosnom prostoru koji za prostor elementarnih događaja ima skup $\mathcal{R}(X) = \{x_i, i \in \mathbb{N}\}$.

Ako brojevi p_i ujedno zadovoljavaju svojstvo

$$P\{X = x_i\} = p_i, \quad \forall i \in \mathbb{N},$$

vjerojatnost definiranu na $\mathcal{R}(X) = \{x_i, i \in \mathbb{N}\}$ označavat ćemo P_X, a novonastali vjerojatnosni prostor $(\mathcal{R}(X), \mathcal{P}(\mathcal{R}(X)), P_X)$ zovemo **vjerojatnosni prostor induciran slučajnom varijablom** X.

Primjer 2.4. *Zbroj brojeva koji su se okrenuli prilikom bacanja dviju pravilno izrađenih igraćih kockica modeliran je diskretnom slučajnom varijablom sa sljedećom tablicom distribucije:*

$$X = \begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \end{pmatrix}.$$

Ako nas zanima vjerojatnost događaja

A - zbroj je brojeva na kockicama je manji od 6,

za izračun ćemo iskoristiti prethodnu tablicu distribucije:

$$P(A) = P\{X < 6\} = P\{\omega \in \Omega : X(\omega) < 6\} = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} = \frac{5}{18}.$$

Općenito, za računanje vjerojatnosti skupova (događaja) oblika $\{X \in A\}$, gdje je $A \subseteq \mathbb{R}$, možemo iskoristiti tablicu distribucije slučajne varijable X (tablica 2.1) na sljedeći način:

$$P\{X \in A\} = \sum_{x_i \in A} p_i.$$

Grafički prikaz distribucije

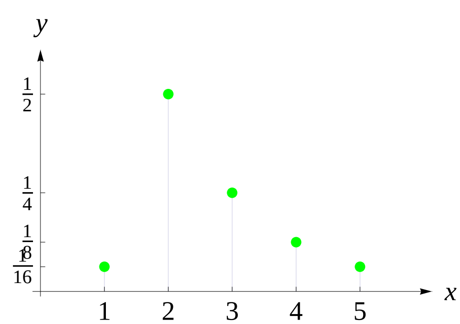
Distribuciju diskretne slučajne varijable koja ima konačan skup mogućih vrijednosti $\mathcal{R}(X)$ i zadana je tablicom distribucije

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

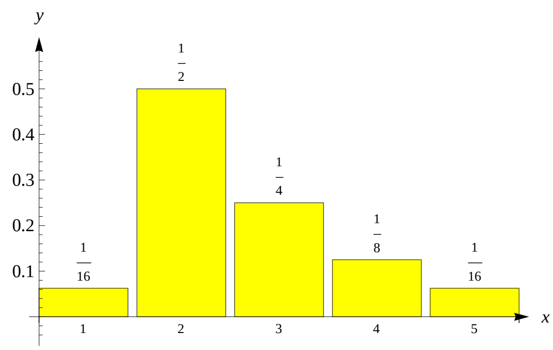
grafički prikazujemo grafom funkcije definirane na $\mathcal{R}(X) = \{x_1, \dots, x_n\}$ s vrijednostima iz skupa $\{p_1, \dots, p_n\}$ koja svakom elementu $x_i \in \mathcal{R}(X)$ pridružuje pripadnu vjerojatnost $p_i = P\{X = x_i\}$, $i \in \{1, \dots, n\}$. Očito se graf te funkcije sastoji od točaka (x_i, p_i).

Također, za grafički prikaz distribucije takve diskretne slučajne varijable često koristimo **stupčasti dijagram**. Stupčasti dijagram sastoji se od niza pravokutnika iste širine, svaki za jednu vrijednost $x_i \in \mathcal{R}(X)$ visine p_i.

Primjer 2.5. *Na temelju tablice distribucije diskretne slučajne varijable definirane u primjeru 2.2 možemo napraviti grafičke prikaze njezine distribucije. Na slici 2.2 prikazan je graf, a na slici 2.3 stupčasti dijagram distribucije te slučajne varijable.*



Slika 2.2: Graf distribucije diskretne slučajne varijable iz primjera 2.2.



Slika 2.3: Stupčasti dijagram distribucije diskretne slučajne varijable iz primjera 2.2.

2.2 Neprekidna slučajna varijabla

Ukoliko prostor elementarnih događaja nije diskretan skup, funkcije definirane na njemu ne moraju nužno imati prebrojiv skup svih mogućih vrijednosti, iako to nije isključeno. Ako je za neku funkciju X, definiranu na Ω koji nije diskretan, skup $\mathcal{R}(X)$ diskretan skup[1], moći ćemo napraviti konstrukciju diskretnog vjerojatnosnog prostora induciranog tom slučajnom varijablom, tj. proučavanje takve slučajne varijable svest ćemo na diskretan slučaj. Međutim, ako prostor elementarnih događaja sadrži neki interval, često nas zanimaju i slučajne karakteristike s neprebrojivim skupom vrijednosti.

Primjer 2.6. *Pretpostavimo da promenadom uz rijeku Dravu želimo prošetati od Tvrđe do hotela Osijek. Taj put dug je oko* $2\,km$. *Zbog različite brzine hoda i ostalih subjektivnih utjecaja na trajanje šetnje, jasno je da put koji osoba pritom prijeđe u prvih 5 minuta možemo modelirati kao slučajnu karakteristiku s neprebrojivim skupom mogućih ishoda* $[0, 2]$.

Primjer 2.7. *Praćenje vremena koje protekne od dana stavljanja stroja u upotrebu do prvog kvara jedno je promatranje sa slučajnim ishodima. Pretpostavimo da dobit ostvarena na tom stroju proporcionalno ovisi o vremenu rada stroja te da nema drugih utjecaja na visinu (iznos) dobiti. Time je dobit ostvarena do njegovog prvog stavljanja izvan pogona zbog kvara (koji će uz to uzrokovati dodatne troškove!) slučajna karakteristika za koju je prirodno pretpostaviti da njezin skup mogućih realizacija sadrži neki interval.*

U tom poglavlju definirat ćemo jedan tip slučajne varijable čija je slika $\mathcal{R}(X)$ podskup od $\mathbb{R}$ i sadrži interval.

Definicija 2.2. *Neka je dan vjerojatnosni prostor* $(\Omega, \mathcal{F}, P)$ *i funkcija* $X\colon \Omega \to \mathbb{R}$ *za koju vrijedi:*

[1] Za definiciju diskretne slučajne varijable u općenitom slučaju pogledati [26]

- $\{\omega \in \Omega : X(\omega) \leq x\} = \{X \leq x\} \in \mathcal{F}$ *za svaki* $x \in \mathbb{R}$,
- *postoji nenegativna realna funkcija realne varijable,* f, *takva da vrijedi:*

$$P\{\omega \in \Omega : X(\omega) \leq x\} = P\{X \leq x\} = \int_{-\infty}^{x} f(t)\, dt, \quad \forall x \in \mathbb{R}.$$

Funkciju X *zovemo* **apsolutno neprekidna slučajna varijabla** *na* Ω *ili, kraće,* **neprekidna slučajna varijabla**. *Funkciju* f *tada zovemo* **funkcija gustoće vjerojatnosti** *slučajne varijable* X *ili, kraće,* **funkcija gustoće** *slučajne varijable* X.

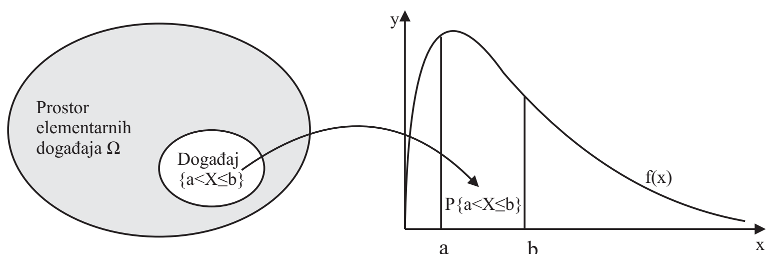
Valja napomenuti da nisu sve slučajne varijable, koje nisu diskretne, apsolutno neprekidne u tom smislu. Za opću teoriju pogledati npr. [26], [4], [6].

Bitna svojstva funkcije gustoće neprekidne slučajne varijable

1. Nenegativnost: $f(x) \geq 0$ za sve $x \in \mathbb{R}$.
2. Normiranost:
$$\int_{-\infty}^{\infty} f(x)\, dx = 1.$$
Dokaz. $\int_{-\infty}^{\infty} f(x)\, dx$ možemo prikazati kao
$$\int_{-\infty}^{\infty} f(x)\, dx = \lim_{n\to\infty} \int_{-\infty}^{n} f(x)\, dx.$$
Koristeći neprekidnost vjerojatnosti u odnosu na monotono rastuću familiju skupova imamo:
$$\lim_{n\to\infty} \int_{-\infty}^{n} f(x)\, dx = \lim_{n\to\infty} P\{X \in \langle -\infty, n]\} = P\{\Omega\} = 1.$$
3. Vjerojatnost da slučajna varijabla X, čija je funkcija gustoće f, primi vrijednost iz intervala $(a, b]$ može se izračunati korištenjem funkcije gustoće na sljedeći način (vidi sliku 2.4):
$$P\{a < X \leq b\} = P\{X \in \langle a, b]\} = \int_{a}^{b} f(x)\, dx.$$

Dokaz. Koristeći svojstva vjerojatnosti i svojstva integrala slijedi:

$$P\{a < X \leq b\} = P\{X \leq b\} - P\{X \leq a\} =$$
$$= \int_{-\infty}^{b} f(x)\, dx - \int_{-\infty}^{a} f(x)\, dx = \int_{a}^{b} f(x)\, dx.$$



Slika 2.4: Vjerojatnost kao površina od osi x do grafa funkcije gustoće nad izabranim intervalom.

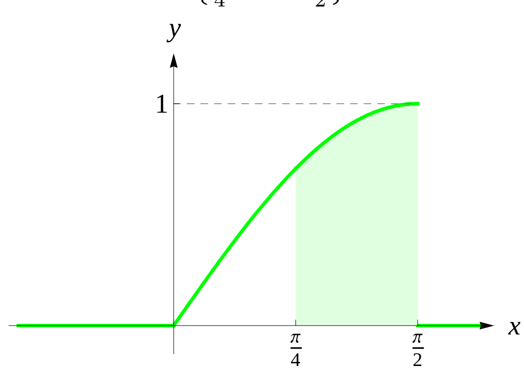
Ako je zadana nenegativna funkcija $f\colon \mathbb{R} \to \mathbb{R}$ koja zadovoljava svojstvo normiranosti, može se dokazati da postoji vjerojatnosni prostor $(\Omega, \mathcal{F}, P)$ i neprekidna slučajna varijabla X na njemu za koju je f funkcija gustoće (vidi [26]).

Primjer 2.8. *Analizirajmo može li funkcija $f\colon \mathbb{R} \to \mathbb{R}$ (vidi sliku 2.5) definirana izrazom*

$$f(x) = \begin{cases} \sin x & , \ x \in \langle 0, \frac{\pi}{2}] \\ 0 & , \ x \notin \langle 0, \frac{\pi}{2}] \end{cases}$$

poslužiti kao funkcija gustoće neke neprekidne slučajne varijable X. Ako može, odredimo

$$P\left\{\frac{\pi}{4} < X \leq \frac{\pi}{2}\right\}.$$



Slika 2.5: Funkcija gustoće slučajne varijable iz primjera 2.8.

Iz definicije funkcije f očito je da se radi o nenegativnoj funkciji:

- *na intervalu* $\langle 0, \pi/2]$ *jest* $f(x) = \sin x$, *a funkcija je sinus na tom intervalu pozitivna,*
- *na* $\langle -\infty, 0] \cup \langle \pi/2, \infty\rangle$ *je* $f(x) = 0$.

Pokažimo da je funkcija f *normirana:*

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{\frac{\pi}{2}} f(x)\,dx + \int_{\frac{\pi}{2}}^{\infty} f(x)\,dx = \int_{0}^{\frac{\pi}{2}} \sin x\,dx = 1.$$

Budući da je funkcija f *nenegativna i normirana, zaključujemo da može poslužiti kao funkcija gustoće neprekidne slučajne varijable.*

Traženu vjerojatnost računamo na sljedeći način:

$$P\left\{\frac{\pi}{4} < X \le \frac{\pi}{2}\right\} = \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \sin x\,dx = \frac{\sqrt{2}}{2}.$$

2.3 Funkcija distribucije slučajne varijable

Diskretna i neprekidna slučajna varijabla samo su dva tipa slučajnih varijabli koje se najčešće koriste u praksi. Općenito, **ako je dan bilo kakav vjerojatnosni prostor** $(\Omega, \mathcal{F}, P)$**, svaku funkciju** X **definiranu na** Ω **sa skupom vrijednosti iz** $\mathbb{R}$ **koja zadovoljava svojstvo**

$$\{X \le x\} \in \mathcal{F}, \quad \forall x \in \mathbb{R},$$

zovemo slučajna varijabla na Ω**.** Vjerojatnosna svojstva za sve slučajne varijable opisana su tzv. **funkcijom distribucije** o kojoj će biti riječi u ovom poglavlju.

Definicija 2.3. *Neka je* $(\Omega, \mathcal{F}, P)$ *vjerojatnosni prostor i neka je* X *slučajna varijabla. Funkciju* $F\colon \mathbb{R} \to [0,1]$ *koja realnom broju* x *pridružuje vjerojatnost da dana slučajna varijabla bude manja ili jednaka tom broju, tj. funkciju*

$$F(x) = P\{\omega \in \Omega : X(\omega) \le x\} = P\{X \le x\},$$

zovemo **funkcija distribucije slučajne varijable X**.

Svojstva funkcije distribucije:

1. Funkcija distribucije slučajne varijable monotono je rastuća funkcija, tj.

$$x_1 < x_2 \Rightarrow F(x_1) \le F(x_2).$$

Dokaz. *Ako je* $x_1 < x_2$, *tada je* $\{X \le x_1\} \subseteq \{X \le x_2\}$. *Primjenom svojstva monotonosti vjerojatnosti na prethodnu činjenicu slijedi tvrdnja.*

2. Neka je F funkcija distribucije neke slučajne varijable. Tada vrijedi:

$$\lim_{x \to -\infty} F(x) = F(-\infty) = 0.$$

Dokaz. *Neka je $(a_n, n \in \mathbb{N})$ monotono padajući niz koji divergira u $-\infty$. Tada vrijedi:*

$$\lim_{n \to \infty} F(a_n) = \lim_{n \to \infty} P\{X \leq a_n\} = P\left(\bigcap_{n \in \mathbb{N}} \{X \leq a_n\}\right) = P(\emptyset) = 0.$$

Odavde lako slijedi tvrdnja.

3. Neka je F funkcija distribucije neke slučajne varijable. Tada vrijedi:

$$\lim_{x \to \infty} F(x) = F(\infty) = 1.$$

Dokaz. *Slično dokazu prethodne tvrdnje.*

4. Funkcija distribucije neprekidna je zdesna, tj.

$$\lim_{x \downarrow x_0} F(x) = F(x_0).$$

Dokaz. *Neka je $x \in \mathbb{R}$ i $(h_n, n \in \mathbb{N})$ monotono padajući niz brojeva takvih da je $\lim\limits_{n \to \infty} h_n = 0$. Tada vrijedi:*

$$F(x + h_n) - F(x) = P\{x < X \leq x + h_n\},$$

$$\lim_{n \to \infty} (F(x + h_n) - F(x)) = P\left(\bigcap_{n \in \mathbb{N}} \{x < X \leq x + h_n\}\right) = P(\emptyset) = 0.$$

Napomenimo da se može dogoditi da je

$$\lim_{x \uparrow x_0} F(x) \neq F(x_0),$$

tj. funkcija distribucije ne mora nužno biti neprekidna i slijeva, no limes slijeva uvijek postoji i stoga uvodimo sljedeću oznaku:

$$\lim_{x \uparrow x_0} F(x) = F(x_0-).$$

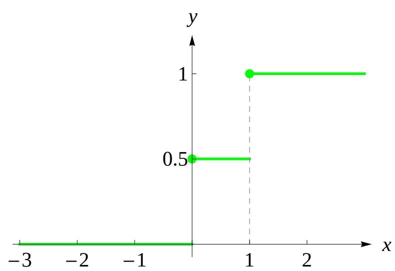
Primjer 2.9. *Neka je dana diskretna slučajna varijabla X kojom modeliramo bacanje pravilno izrađenog novčića. Ako označimo "uspjeh" (npr. palo je "pismo") brojem 1, a "neuspjeh" brojem 0, tablica distribucije ove slučajne varijable jest*

$$X = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Funkcija distribucije ima vrijednost 0 *za sve* x *koji su manji od* 0. *U broju* 0 *prima vrijednost* $\frac{1}{2}$ *i ostaje na toj vrijednosti sve do broja* 1 *kada prima vrijednost* 1, *tj.*

$$F(x) = P\{X \leq x\} = \begin{cases} 0 & , \ x \in \langle -\infty, 0\rangle \\ \frac{1}{2} & , \ x \in [0, 1\rangle \\ 1 & , \ x \in [1, \infty\rangle \end{cases} .$$

Graf funkcije distribucije te slučajne varijable prikazan je na slici 2.9. *Uočimo da navedena funkcija distribucije nije neprekidna slijeva.*



Slika 2.6: Funkcija distribucije za slučajnu varijablu kojom modeliramo bacanje pravilno izrađenog novčića.

Koristeći funkciju distribucije i svojstva vjerojatnosti možemo lako računati vjerojatnost da slučajna varijabla primi vrijednost iz nekog skupa A, gdje je skup A neki interval, prebrojiva unija ili presjek intervala, njihova razlika i slično. Primjerice, za $a, b \in \mathbb{R}$ takve da je $a < b$ vrijedi:

$$\begin{aligned} P\{a < X \leq b\} &= P\{X \leq b\} - P\{X \leq a\} = F(b) - F(a), \\ P\{X = a\} &= P\left(\bigcap_{n=1}^{\infty} \{a - \tfrac{1}{n} < X \leq a\}\right) = \\ &= \lim_{n\to\infty} (F(a) - F(a - \tfrac{1}{n})) = \\ &= F(a) - F(a-), \\ P\{a \leq X \leq b\} &= P\{X = a\} + P\{a < X \leq b\} = \\ &= F(b) - \lim_{n\to\infty} F(a - \tfrac{1}{n}) = \\ &= F(b) - F(a-) \end{aligned}$$

i slično.

Specijalno, ako je slučajna varijabla diskretna, funkcija distribucije može se izraziti koristeći tablicu distribucije slučajne varijable, a ako je neprekidna, koristeći pripadnu funkciju gustoće.

2.3.1 Funkcija distribucije diskretne slučajne varijable

Neka je dana diskretna slučajna varijabla

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n & \cdots \\ p_1 & p_2 & p_3 & \cdots & p_n & \cdots \end{pmatrix},$$

tj. $\mathcal{R}(X) = \{x_i, i \in I\}$, $I \subseteq \mathbb{N}$, je dani skup vrijednosti slučajne varijable X, a $(p_i, i \in I)$ niz pripadnih vjerojatnosti za koji vrijedi

$$p_i = P\{X = x_i\} \geq 0, \quad \forall i \in I, \qquad \sum_{i \in I} p_i = 1.$$

Za svaki $x \in \mathbb{R}$ vrijedi:

$$F(x) = P\{X \leq x\} = \sum_{x_i \leq x} p_i.$$

Uočimo da je takva funkcija distribucije stepenasta funkcija, tj. funkcija sa skokovima u točkama x_i, dok na intervalu $[x_i, x_{i+1})$ prima stalno istu vrijednost koja je jednaka $F(x_i)$ (vidi npr. sliku 2.6).

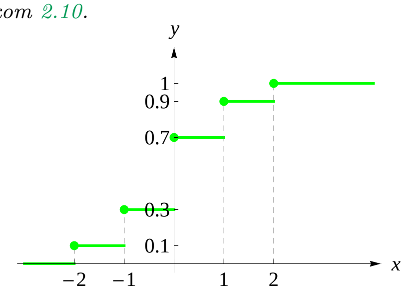
Primjer 2.10. *Neka je diskretna slučajna varijabla zadana sljedećom tablicom distribucije:*

$$X = \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 0.1 & 0.2 & 0.4 & 0.2 & 0.1 \end{pmatrix}.$$

Prema definiciji funkcije distribucije slijedi da je funkcija distribucije slučajne varijable X definirana izrazom

$$F(x) = \begin{cases} 0 & , x \in (-\infty, -2) \\ 0.1 & , x \in [-2, -1) \\ 0.3 & , x \in [-1, 0) \\ 0.7 & , x \in [0, 1) \\ 0.9 & , x \in [1, 2) \\ 1 & , x \in [2, \infty) \end{cases}.$$

Njezin graf prikazan je slikom 2.10.



Slika 2.7: Graf funkcije distribucije diskretne slučajne varijable X iz primjera 2.10.

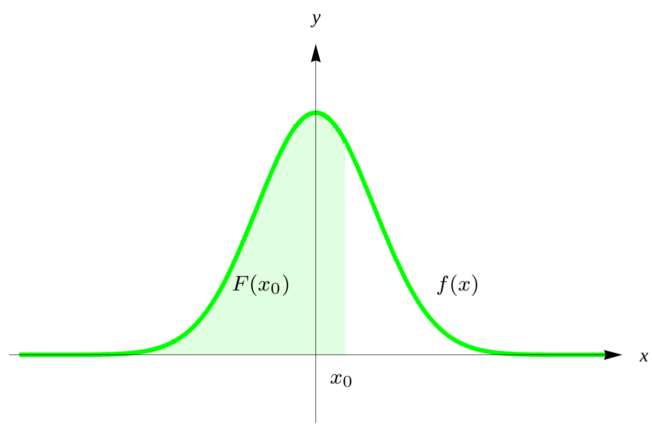
2.3.2 Funkcija distribucije neprekidne slučajne varijable

Neka je X neprekidna slučajna varijabla s funkcijom gustoće f. Dakle, $f(x) \geq 0$ za sve $x \in \mathbb{R}$, $\int\limits_{-\infty}^{\infty} f(x)\,dx = 1$ i

$$P\{X \leq x\} = \int\limits_{-\infty}^{x} f(t)\,dt, \;\; x \in \mathbb{R}.$$

Odavde direktno slijedi da vrijednost funkcije distribucije slučajne varijable X za proizvoljan realan broj x_0 određujemo na sljedeći način (slika 2.8):

$$F(x_0) = P\{X \leq x_0\} = \int\limits_{-\infty}^{x_0} f(t)\,dt, \quad \forall x_0 \in \mathbb{R}.$$



Slika 2.8: Funkcija gustoće i vrijednost funkcije distribucije u x_0 (osjenčana površina) neprekidne slučajne varijable.

Iz dobro poznatih svojstava integrala očigledno je da je, za razliku od diskretnog slučaja, funkcija distribucije neprekidne slučajne varijable neprekidna funkcija u svakoj točki $x \in \mathbb{R}$. Također, funkcija gustoće neprekidne slučajne varijable X može se dobiti kao derivacija pripadne funkcije distribucije u gotovo svakoj točki $x \in \mathbb{R}$, tj. $f(x) = F'(x)$.

Odavde slijedi i sljedeće važno svojstvo za neprekidne slučajne varijable.

Neka je X neprekidna slučajna varijabla. Tada za svaki $x_0 \in \mathbb{R}$ vrijedi:

$$P\{X = x_0\} = 0.$$

Dokaz. *Zbog neprekidnosti funkcije distribucije vrijedi:*

$$P\{X = x_0\} = P\left(\bigcap_{n=1}^{\infty}\{x_0 - \frac{1}{n} < X \leq x_0\}\right) = \\ = \lim_{n\to\infty}\left(F(x_0) - F\left(x_0 - \frac{1}{n}\right)\right) = 0.$$

Primjer 2.11. *Odredimo funkciju distribucije neprekidne slučajne varijable s funkcijom gustoće definiranom sljedećim izrazom:*

$$f(x) = \begin{cases} \sin x & , \ x \in \langle 0, \frac{\pi}{2}] \\ 0 & , \ x \notin \langle 0, \frac{\pi}{2}] \end{cases} . \tag{2.2}$$

Ako je $x \in \langle -\infty, 0]$, tada je

$$F(x) = P\{X \leq x\} = \int_{-\infty}^{x} f(x)dx = 0.$$

Za $x \in \langle 0, \pi/2]$ je

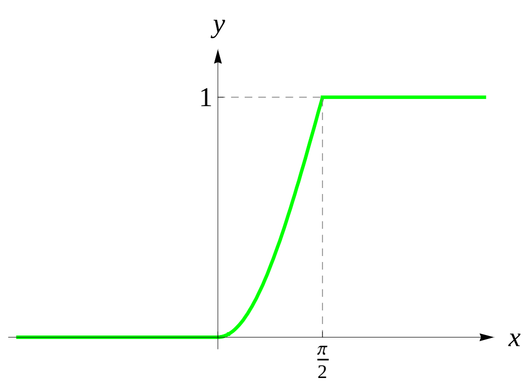
$$F(x) = P\{X \leq x\} = \int_{-\infty}^{0} f(x)dx + \int_{0}^{x} f(x)dx = \int_{0}^{x} \sin x dx = 1 - \cos x.$$

Za $x > \pi/2$ je očigledno $F(x) = F(\pi/2) = 1$.

Dakle, funkcija distribucije neprekidne slučajne varijable X definirana je izrazom

$$F(x) = \begin{cases} 0 & , \ x \in \langle -\infty, 0] \\ 1 - \cos x & , \ x \in \langle 0, \pi/2] \\ 1 & , \ x \in \langle \pi/2, \infty) \end{cases}$$

Graf te funkcije distribucije prikazan je na slici 2.11. Uočimo da je funkcija F neprekidna.



Slika 2.9: Graf funkcije distribucije neprekidne slučajne varijable X s gustoćom (2.2).

Intuitivno značenje naziva "gustoća slučajne varijable" za neprekidne slučajne varijable može se prepoznati iz činjenice da je funkcija gustoće neprekidne slučajne varijable zapravo derivacija funkcije distribucije. Naime, iz definicije derivacije slijedi:

$$f(x) = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{P\{x < X \le x + \Delta x\}}{\Delta x}.$$

Dakle, $f(x)$ je granična vrijednost po duljini intervala, kvocijenta vjerojatnosti da se X nađe u intervalu $\langle x, x+\Delta x]$ i duljine intervala, što intuitivno odgovara pojmu gustoće vjerojatnosti.

2.4 Primjeri parametarski zadanih diskretnih distribucija

Neki tipovi tablica distribucije diskretnih slučajnih varijabli koji se često koriste u praksi daju se jednostavno opisati koristeći jedan ili nekoliko realnih brojeva (parametara). Za takve distribucije kažemo da se mogu zadati parametarski i svrstavamo ih u parametarske familije distribucija s prepoznatljivim imenima i svojstvima. U ovom poglavlju definirat ćemo nekoliko takvih familija.

2.4.1 Diskretna uniformna distribucija

Za slučajnu varijablu X kažemo da ima diskretnu uniformnu distribuciju ako je njezin skup svih mogućih vrijednosti konačan podskup skupa realnih brojava, tj. $\mathcal{R}(X) = \{x_1, x_2, \ldots, x_n\} \subseteq \mathbb{R}$, a pripadni je niz vjerojatnosti definiran kao

$$p_i = P\{X = x_i\} = \frac{1}{n}, \quad i = 1, 2, \ldots, n.$$

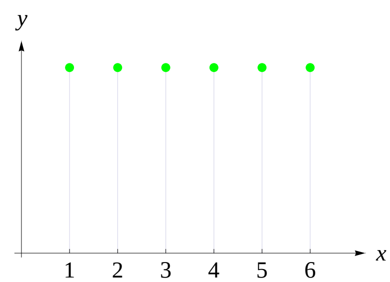
Takve distribucije koriste se ako slučajna varijabla X može primiti samo vrijednosti iz skupa $\mathcal{R}(X) = \{x_1, x_2, \ldots, x_n\}$ i to tako da je vjerojatnost realizacije svakog pojedinog ishoda ista, tj. $P\{X = x_i\} = 1/n$ za svaki $i \in \{1, \ldots, n\}$.

Uočimo da je na taj način dobro definirana tablica distribucije s obzirom da je $\sum\limits_{i=1}^{n} p_i = 1$. Također, pripadni niz vjerojatnosti ovisi samo o broju elemenata skupa $\mathcal{R}(X)$, tj. o broju n.

Primjer 2.12 (Bacanje igraće kockice.). *Neka slučajna varijabla X daje broj koji se okrenuo pri bacanju pravilno izrađene igraće kockice. Tada ona ima diskretnu uniformnu distribuciju zadanu tablicom distribucije*

$$X = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}.$$

Graf distribucije bacanja igraće kockice prikazan je slikom 2.10.



Slika 2.10: Graf distribucije bacanja igraće kockice.

Primjer 2.13. *Pretpostavimo da na ispitu na slučajan način biramo jedno od m ponuđenih pitanja. Rezultate tog slučajnog pokusa možemo opisati diskretnom uniformnom slučajnom varijablom X s tablicom distribucije*

$$X = \begin{pmatrix} 1 & 2 & 3 & \dots & m-1 & m \\ \frac{1}{m} & \frac{1}{m} & \frac{1}{m} & \dots & \frac{1}{m} & \frac{1}{m} \end{pmatrix}.$$

2.4.2 Bernoullijeva distribucija

Neka je X slučajna varijabla koja može primiti točno dvije vrijednosti, i to $\mathcal{R}(X) = \{0,1\}$. Njezina distribucija dana je tablicom distribucije

$$X = \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix},\ p \in \langle 0,1 \rangle,\ q = 1-p.$$

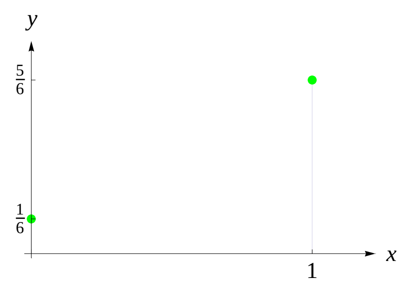
Za takvu slučajnu varijablu reći ćemo da ima Bernoullijevu distribuciju s parametrom p, pri čemu parametar p ima značenje vjerojatnosti da X primi vrijednost 1.

Bernoullijev tip distribucije koristi se pri modeliranju slučajnih karakteristika koje mogu imati točno dvije vrijednosti. Te moguće vrijednosti uglavnom zovemo "uspjeh" i "neuspjeh" te koristimo oznake 1 za "uspjeh", a 0 za "neuspjeh". Dakle, parametar p Bernoulijeve distribucije ima značenje vjerojatnosti pojavljivanja "uspjeha".

Primjer 2.14. *Igramo kockarsku igru u kojoj ostvarujemo dobitak ako se na pravilno izrađenoj igraćoj kocki okrene šestica (što interpretiramo kao uspjeh i označavamo 1). Ishod te igre modeliramo Bernoullijevom slučajnom varijablom X s tablicom distribucije*

$$X = \begin{pmatrix} 0 & 1 \\ \frac{5}{6} & \frac{1}{6} \end{pmatrix}.$$

Graf distribucije prikazan je slikom 2.11.



Slika 2.11: Graf Bernoullijeve distribucije iz primjera 2.14.

Primjer 2.15. *Izvlačimo jedan proizvod iz velike pošiljke u kojoj je* 2% *neispravnih. Ako nas zanima je li izvučen ispravan ili neispravan proizvod, rezultat izvlačenja možemo modelirati Bernoullijevom slučajnom varijablom* X. *Neka je* 1 *oznaka dobrog proizvoda. Tada je* X *zadana tablicom distribucije*

$$X = \begin{pmatrix} 0 & 1 \\ 0.02 & 0.98 \end{pmatrix}.$$

2.4.3 Binomna distribucija

Binomna distribucija vezana je uz nezavisno ponavljanje uvijek istog pokusa. Ako nas pri svakom izvođenju pokusa zanima samo je li se dogodio neki događaj (uspjeh!) ili ne (neuspjeh!), onda svako izvođenje pokusa možemo modelirati istom Bernoullijevom distribucijom

$$Y = \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix},\ p \in \langle 0, 1\rangle,\ q = 1 - p,$$

gdje p predstavlja vjerojatnost uspjeha.

Pretpostavimo da pokus ponavljamo nezavisno n puta i pri tome nas zanima broj uspjeha. Za slučajnu varijablu X koja opisuje broj uspjeha u n nezavisnih ponavljanja slučajnog pokusa modeliranog Bernoullijevog slučajnom varijablom Y kažemo da ima binomnu distribuciju s parametrima n i p.

Primjer 2.16. *Stroj proizvodi CD-ove. Vjerojatnost da bude proizveden neispravan CD je* p. *Zanima nas broj neispravnih CD-ova ako s beskonačne trake uzimamo njih* 100. *Slučajna varijabla kojom modeliramo broj neispravnih CD-ova među* 100 *odabranih jest binomna slučajna varijabla* X *parametrima* $n = 100$ *i* p, *a zadana je sljedećom tablicom distribucije:*

$$X = \begin{pmatrix} 0 & 1 & \cdots & 100 \\ p_1 & p_2 & \cdots & p_{100} \end{pmatrix},$$

gdje je:

$$p_i = P\{X = x_i\} = \binom{100}{i} p^i (1-p)^{100-i},\ i \in \{0, 1, \ldots, 100\}.$$

Definirajmo slučajnu varijablu s binomnom distribucijom.

Definicija 2.4. *Neka je* $n \in \mathbb{N}$ *i* $p \in \langle 0,1\rangle$*. Za slučajnu varijablu koja prima vrijednosti iz skupa* $\{0,1,2,...,n\}$ *s vjerojatnostima*

$$p_i = P\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i}$$

kažemo da ima **binomnu distribuciju s parametrima** n **i** p *i pišemo:* $X \sim \mathcal{B}(n,p)$.

Značenje parametara u $X \sim \mathcal{B}(n,p)$:

p je vjerojatnost uspjeha u jednom izvođenju pokusa,

n je broj nezavisnih ponavljanja pokusa.

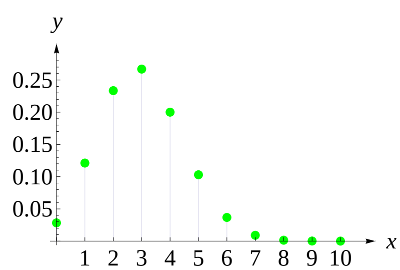
Provjerimo je li na taj način dobro definirana distribucija, tj. je li $\sum\limits_{i=0}^{n} p_i = 1$. Vrijedi:

$$\sum_{i=0}^{n} p_i = \sum_{i=0}^{n} \binom{n}{i} p^i q^{n-i} = (p+q)^n = 1.$$

Primjer 2.17. *Pretpostavimo da iz velikog skladišta slatkiša (u kojemu se nalaze čokolade, bomboni, keksi, ...) 10 puta nezavisno izvlačimo po jedan slatkiš. Ako je vjerojatnost da u jednom izvlačenju izvučemo čokoladu* $p = 0.3$*, tada slučajna varijabla koja opisuje broj izvučenih čokolada u 10 nezavisnih izvlačenja ima binomnu distribuciju s parametrima* $n = 10$ *i* $p = 0.3$*, tj.* $X \sim \mathcal{B}(10, 0.3)$*. Tablica distribucije slučajne varijable* X *jest*

$$X = \begin{pmatrix} 0 & 1 & 2 & \dots & 10 \\ 0.7^{10} & \binom{10}{1} \cdot 0.3 \cdot 0.7^9 & \binom{10}{2} \cdot 0.3^2 \cdot 0.7^8 & \dots & 0.3^{10} \end{pmatrix}.$$

Graf ove distribucije prikazan je slikom 2.12.



Slika 2.12: Graf binomne distribucija iz primjera 2.17.

Sada lako možemo odrediti sljedeće vjerojatnosti:

a) vjerojatnost da smo čokoladu izvukli točno 5 *puta:*

$$P\{X = 5\} = \binom{10}{5} 0.3^5 (1 - 0.3)^5 \approx 0.103,$$

b) vjerojatnost da smo čokoladu izvukli manje od 3 puta:

$$P\{X < 3\} = P\{X \leq 2\} = \sum_{k=0}^{2} \binom{10}{k} 0.3^k (1 - 0.3)^{10-k} \approx 0.383,$$

c) vjerojatnost da smo čokoladu izvukli barem 2 *puta:*

$$P\{X \geq 2\} = 1 - P\{X < 2\} = 1 - \sum_{k=0}^{1} \binom{10}{k} 0.3^k (1 - 0.3)^{10-k} \approx 0.851.$$

2.4.4 Poissonova distribucija

Poissonova distribucija, slično kao i binomna, može se primijeniti kao distribucija slučajne varijable koja broji uspjehe, ali ne pri nezavisnom ponavljanju pokusa konačno mnogo puta, nego u jediničnom vremenskom intervalu ili intervalu volumena, mase, itd. ako pokus zadovoljava sljedeće uvjete:

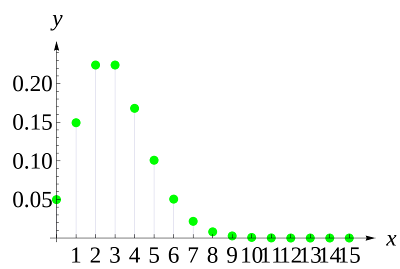
- vjerojatnost da se pojavi uspjeh ne ovisi o tome u kojem će se jediničnom intervalu dogoditi,
- broj uspjeha u jednom intervalu neovisan je o broju uspjeha u nekom drugom intervalu,
- očekivani broj uspjeha isti je za sve jedinične intervale i dan je pozitivnim realnim brojem λ.

Definicija 2.5. *Slučajna varijabla X ima* **Poissonovu distribuciju s parametrom** $\lambda > 0$ *ako prima vrijednosti iz skupa* $\{0, 1, 2, \dots\}$ *s vjerojatnostima*

$$p_i = P\{X = i\} = e^{-\lambda} \frac{\lambda^i}{i!}.$$

Tada pišemo $X \sim \mathcal{P}(\lambda)$.

Histogram distribucije Poissonove slučajne varijable s parametrom 3 za skup realizacija $\{0, 1, 2, \dots, 15\}$ prikazan je slikom 2.13.



Slika 2.13: Graf Poissonove distribucija s parametrom 3 za skup realizacija $\{0, 1, 2, \ldots, 15\}$.

Provjerimo je li na taj način dobro definirana distribucija, tj. je li $\sum_{i=0}^{\infty} p_i = 1$. Vrijedi:

$$\sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1.$$

Primjer 2.18. *Pretpostavimo da neki kafić u toku jednog sata u prosjeku posjeti* 15 *ljudi. Slučajna varijabla koja broji posjetitelje kafića tijekom jednog sata jest Poissonova slučajna varijabla s parametrom* $\lambda = 15$, *tj.* $X \sim \mathcal{P}(15)$. *Ona prima vrijednosti iz skupa* $\mathcal{R}(X) = \{0, 1, \ldots\}$ *s vjerojatnostima*

$$p_i = \frac{15^i}{i!} e^{-15}, \quad i \in \mathcal{R}(X).$$

Na temelju navedenih informacija možemo odrediti sljedeće vjerojatnosti:

a) vjerojatnost da je kafić u toku jednog sata posjetilo točno 20 *ljudi:*

$$P\{X = 20\} = \frac{15^{20}}{20!} e^{-15} \approx 0.042,$$

b) vjerojatnost da je kafić u toku jednog sata posjetilo manje od 15 *ljudi:*

$$P\{X < 15\} = P\{X \leq 14\} = \sum_{i=0}^{14} \frac{15^i}{i!} e^{-15} \approx 0.466,$$

c) vjerojatnost da je kafić u toku jednog sata posjetilo više od 10 *ljudi:*

$$P\{X > 10\} = 1 - P\{X \leq 10\} = 1 - \sum_{i=0}^{10} \frac{15^i}{i!} e^{-15} \approx 0.882.$$

Poissonova distribucija može se dobiti i kao aproksimacija binomne ako je parametar n binomne distribucije jako velik, a vjerojatnost p realizacije uspjeha mala. Ilustracija te tvrdnje dana je primjerom 2.19.

Primjer 2.19. *Telefonska centrala u vremenskom periodu u trajanju od pola sata prima* 600 *poziva uglavnom ravnomjerno raspoređenih u spomenutom vremenu (tj. u prosjeku* 20 *u minuti). Broj poziva u prvoj minuti možemo modelirati binomnom distribucijom. Naime, vjerojatnost da se poziv dogodi u prvoj od tih* 30 *minuta jest* $p = 1/30$ *(jer su pozivi uglavnom jedoliko raspoređeni u vremenu). Međutim, broj je poziva u minuti slučajna varijabla pa, iako je prosječan broj poziva* *u minuti 20, u modelu moramo pretpostaviti da se u prvoj minuti može dogoditi od* 0 *do* 600 *poziva. Dakle,* $n = 600$ *za tu binomnu slučajnu varijablu. Uočimo da prosječan broj poziva po minuti iznosi* np.

Problem je s binomnom distribucijom u navedenom primjeru prvenstveno taj da moramo pretpostaviti maksimalan broj poziva u minuti, što u modelu telefonske centrale nije razumno. Međutim, tu je $n = 600$ *velik, a* $p = 1/30$ *malen, pa vjerojatnosti zadane po binomnoj distribuciji možemo dobro aproksimirati vjerojatnostima po Poissonovoj distribuciji s parametrom koji odgovara prosječnom broju poziva po minuti, tj.* $\lambda = np = 20$.

Označimo $np = \lambda$ *i pokažimo da se, pod pretpostavkom da je* n *velik prirodan broj i da* $\lambda > 0$ *ne ovisi o* n*, vjerojatnosti po binomnoj distribuciji mogu dobro aproksimirati vjerojatnostima po Poissonovoj. Zaista, za binomnu slučajnu varijablu* $X_n \sim \mathcal{B}(n,p)$*, te za* $i \in \{0,\dots,n\}$ *jest*

$$p_i(n) = P\{X_n = i\} = \binom{n}{i} p^i(1-p)^{n-i} = \frac{n(n-1)(n-2)\cdots(n-i+1)}{i!}p^i(1-p)^{n-i}.$$

Stavimo li $\lambda = np$*, tj.* $p = \lambda/n$*, slijedi:*

$$\begin{aligned} p_i(n) &= \frac{n(n-1)(n-2)\cdots(n-i+1)}{i!}\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i} = \\ &= \frac{1}{i!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{i-1}{n}\right)\lambda^i\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-i}. \end{aligned}$$

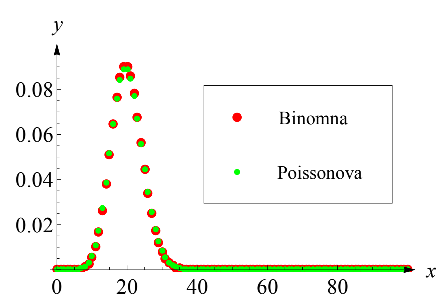
Da bismo odredili približnu vrijednost $p_i(n)$ *za velike* n*, uočimo da vrijedi:*

$$\lim_{n\to\infty} p_i(n) = \frac{1}{i!}\lambda^i e^{-\lambda}.$$

Dakle, za velike n *jest*

$$p_i(n) \approx \frac{1}{i!}\lambda^i e^{-\lambda}.$$

Slikom 2.14 prikazan je graf binomne distribucije s parametrima $p = 30$ *i* $n = 600$ *i Poissonove s parametrom* $\lambda = 20$ *koji ilustrira kvalitetu aproksimacije.*



Slika 2.14: Graf binomne $\mathcal{B}(600, \frac{1}{30})$ i Poissonove $\mathcal{P}(20)$ distribucije za $i = 0,\dots,100$.

2.4.5 Geometrijska distribucija

Geometrijska distribucija također je vezana uz nezavisno ponavljanje istog pokusa s ishodima "uspjeh" i "neuspjeh" kao i binomna. Međutim, ona se ne koristi za opisivanje broja uspjeha već za opisivanje broja ponavljanja pokusa do prvog uspjeha. Preciznije govoreći, neka je vjerojatnost pojavljivanja događaja A u svakom od nezavisnih ponavljanja istog pokusa $p \in \langle 0, 1\rangle$. Geometrijskom distribucijom opisana je slučajna varijabla koja daje broj potrebnih pokusa da bi se realizirao taj događaj. Budući da je, pod tim pretpostavkama, vjerojatnost pojavljivanja događaja A u k-tom pokusu

$$P\{X = k\} = P\left(\underbrace{A^c \bigcap A^c \bigcap \cdots \bigcap A^c}_{(k-1)\ puta} \bigcap A\right) = (1-p)^{k-1}p,$$

geometrijsku distribuciju možemo definirati na slijedeći način.

Definicija 2.6. *Slučajna varijabla X ima* **geometrijsku distribuciju s parametrom** $p \in \langle 0, 1\rangle$ *ako prima vrijednosti iz skupa $\{1, 2, 3, \dots\}$ s vjerojatnostima*

$$p_k = P\{X = k\} = p(1-p)^{k-1}.$$

Pokažimo da je na taj način dobro definirana distribucija, tj. da je $\sum\limits_{i=1}^{\infty} p_i = 1$. Koristeći formulu za sumu geometrijskog reda slijedi:

$$\sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} p(1-p)^{i-1} = p\frac{1}{1-(1-p)} = 1.$$

Primjer 2.20. *Tipkovnica na prijenosnom računalu nekog proizvođača sastoji se od 86 tipki. Pretpostavimo da zatvorenih očiju trebamo pritisnuti tipku za slovo A. Budući da je vjerojatnost pogotka bilo koje tipke jednaka 1/86 (pretpostavimo da pri svakom pokušaju pogodimo neku tipku, tj. da nikada ne promašimo tipkovnicu), zaključujemo da slučajna varijabla koja opisuje broj pritisnutih tipki do pogađanja tipke za slovo A ima geometrijsku distribuciju s parametrom $p = 1/86$. Tako definirana slučajna varijabla prima vrijednosti iz skupa $\mathcal{R}(X) = \{1, 2, \dots\}$ s vjerojatnostima*

$$p_k = P\{X = k\} = \frac{1}{86}\left(1 - \frac{1}{86}\right)^{k-1}.$$

Sada možemo izračunati sljedeće i slične vjerojatnosti:

a) vjerojatnost da je tipka za slovo A stisnuta u petnaestom pokušaju:

$$P\{X = 15\} = \frac{1}{86}\left(1 - \frac{1}{86}\right)^{14} \approx 0.009,$$

b) vjerojatnost da je tipka za slovo A stisnuta u manje od 5 pokušaja:

$$P\{X < 5\} = P\{X \leq 4\} = \sum_{k=1}^{4} \frac{1}{86}\left(1 - \frac{1}{86}\right)^{k-1} \approx 0.046,$$

c) *vjerojatnost da niti nakon* 20 *pokušaja još nismo uspjeli stisnuti tipku za slovo* A*:*

$$P\{X > 20\} = 1 - P\{X \leq 20\} = 1 - \sum_{k=1}^{20} \frac{1}{86}\left(1 - \frac{1}{86}\right)^{k-1} \approx 0.791.$$

2.4.6 Hipergeometrijska distribucija

Neka je skup iz kojeg vršimo odabir elemenata konačan i neka se sastoji od točno N elemenata od kojih je M tipa 1, a $(N - M)$ tipa 2. Pretpostavimo da smo na slučajan način iz tog skupa odabrali $n < N$ elemenata, i to bez vraćanja prethodno izvučenih elemenata u skup. Tada broj izvučenih elemenata tipa 1 modeliramo **hipergeometrijskom distribucijom** s parametrima N, M i n.

Definicija 2.7. *Diskretna slučajna varijabla* X *ima* **hipergeometrijsku distribuciju** *s parametrima* N*,* M *i* n*,* $N, M, n \in \mathbb{N}$*, ako prima vrijednosti iz skupa* $\mathcal{R}(X) = \{k \in \mathbb{N} : \max(0, n - N + M) \leq k \leq \min(n, M)\}$ *s vjerojatnostima*

$$p_k = P\{X = k\} = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}.$$

Pokažimo da je na taj način dobro definirana distribucija, tj. da je $\sum\limits_{k=0}^{n} p_k = 1$. Primjenom Vandermondeove konvolucije slijedi:

$$\sum_{k=0}^{n} p_i = \sum_{k=0}^{n} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} = \frac{1}{\binom{N}{n}} \sum_{k=0}^{n} \binom{M}{k}\binom{N-M}{n-k} = \frac{\binom{N}{n}}{\binom{N}{n}} = 1.$$

Primjer 2.21. *Na polici se nalazi* 10 *knjiga od kojih su* 4 *kriminalistički romani. Na slučajan način biramo* 5 *knjiga. Slučajna varijabla* X *koja modelira broj kriminalističkih romana od* 5 *odabranih knjiga ima hipergeometrijsku distribuciju s parametrima* $N = 10$*,* $M = 4$ *i* $n = 4$*. Slika slučajne varijable* X *jest skup* $\mathcal{R}(X) = \{0, 1, 2, 3, 4\}$*. Na temelju poznatih informacija možemo izračunati sljedeće i slične vjerojatnosti:*

a) *vjerojatnost da su odabrana točno* 3 *kriminalištička romana:*

$$P\{X = 3\} = \frac{\binom{4}{3} \cdot \binom{10-4}{5-3}}{\binom{10}{5}} = \frac{5}{21},$$

b) *vjerojatnost da su odabrana najviše* 3 *kriminalištička romana:*

$$P\{X \leq 3\} = P\{X = 0\} + P\{X = 1\} + P\{X = 2\} + P\{X = 3\} = \frac{41}{42},$$

c) vjerojatnost da su odabrana najmanje 3 kriminalistička romana:

$$P\{X \geq 3\} = 1 - P\{X < 3\} = 1 - (P\{X = 0\} + P\{X = 1\} + P\{X = 2\}) = \frac{11}{42}.$$

Neka je X hipergeometrijska slučajna varijabla s parametrima N, M i n. Ako $N \to \infty$ i $M \to \infty$ tako da $M/N \to p$, gdje je p pozitivna konstanta, tada

$$\lim_{N\to\infty} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} = \binom{n}{k} p^k (1-p)^{n-k}, \quad \forall k \in \{0, 1, \ldots, n\}.$$

Navedena tvrdnja daje aproksimaciju hipergeometrijske distribucije binomnom distribucijom u slučaju kad je n malen u odnosu na N i M, a intuitivno se može opravdati time da se odabir n elemenata iz N-članog skupa bez vraćanja prethodno izvučenih elemenata (hipergeometrijska distribucija) može za velik N "aproksimirati" odabirom n elemenata s vraćanjem izvučenih elemenata u skup (binomna distribucija).

2.5 Primjeri parametarski zadanih neprekidnih distribucija

Slično kao u diskretnom slučaju, i među neprekidnim slučajnim varijablama postoje parametarski zadane familije slučajnih varijabli koje se često koriste. Ovdje ćemo uvesti samo četiri takve familije. Još neke od njih, koje ćemo koristiti u statistici, definirat ćemo naknadno.

2.5.1 Uniformna distribucija na intervalu $\langle a, b\rangle$

Neprekidna uniformna distribucija veže se uz pokuse za koje je poznato da mogu primiti vrijednost iz ograničenog intervala $\langle a, b\rangle$, ali pritom nema razloga preferirati neko područje, tj. vjerojatnost realizacije intervala $\langle x_1, x_2\rangle$ ovisit će samo o njegovoj duljini sve dok je sadržan u $\langle a, b\rangle$.

Definicija 2.8. *Za neprekidnu slučajnu varijablu X kažemo da ima* **uniformnu distribuciju na intervalu** $\langle a, b\rangle$, *$a < b$, ako joj je funkcija gustoće vjerojatnosti dana izrazom*

$$f(x) = \begin{cases} \dfrac{1}{b-a} & , x \in \langle a, b\rangle \\ 0 & , x \notin \langle a, b\rangle \end{cases}.$$

S obzirom da za svaku neprekidnu slučajnu varijablu X vrijedi $P\{X = x_0\} = 0$, vidimo da u vjerojatnosnom smislu ne treba praviti bitnom razliku između uniformne distribucije na $\langle a, b\rangle$ i uniformne distribucije na $[a, b]$, odnosno na nekom od intervala $\langle a, b]$, $[a, b\rangle$.

Funkcija distribucije uniformne slučajne varijable na $\langle a, b\rangle$ definirana je pravilom

$$F(x) = \begin{cases} 0 & , x \in \langle -\infty, a\rangle \\ \dfrac{x-a}{b-a} & , x \in [a, b\rangle \\ 1 & , x \in [b, \infty\rangle \end{cases} .$$

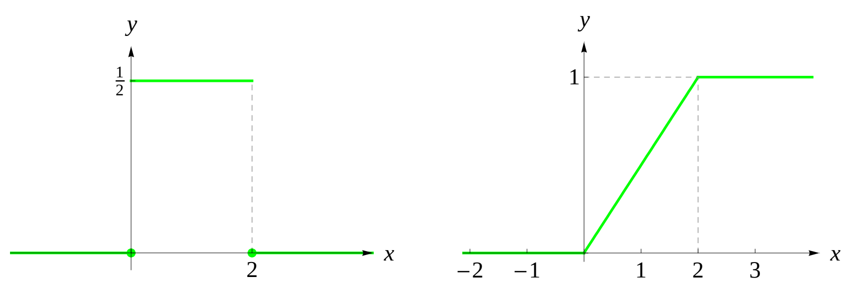
Primjer 2.22. *Pretpostavimo da imamo posudu u koju stane najviše 2 litre vode te da smo iz slavine u nju slučajno natočili nepoznatu količinu vode. Slučajna varijabla kojom opisujemo količinu vode natočenu u posudu može se modelirati kao uniformna slučajna varijabla na intervalu $\langle 0, 2\rangle$. Prema tome, njezina funkcija gustoće dana je izrazom*

$$f(x) = \begin{cases} \frac{1}{2} & , x \in \langle 0, 2\rangle \\ 0 & , x \notin \langle 0, 2\rangle \end{cases} ,$$

a funkcija distribucije izrazom

$$F(x) = \begin{cases} 0 & , x \in \langle -\infty, 0] \\ \frac{x}{2} & , x \in \langle 0, 2\rangle \\ 1 & , x \in [2, \infty\rangle \end{cases} .$$

Grafički prikazi funkcije gustoće i funkcije distribucije dani su slikom 2.15.



Slika 2.15: Graf funkcije gustoće i funkcije distribucije uniformne distribucije na intervalu $\langle 0, 2\rangle$.

Npr. vjerojatnost da smo na slučajan način iz slavine u posudu natočili više od pola litre, ali najviše jednu litru vode, računamo na sljedeći način:

$$P\{0.5 < X \leq 1\} = P\{X \leq 1\} - P\{X \leq 0.5\} = \frac{1}{2}\int_{0.5}^{1} dx = \frac{1}{4}.$$

2.5.2 Eksponencijalna distribucija

Eksponencijalna distribucija često se javlja kod slučajnih varijabli koje imaju značenje vremena čekanja do pojave nekog događaja ako se karakteristike ne mijenjaju tijekom vremena, npr. vrijeme do kvara (tj. vrijeme trajanja) jedne žarulje, vrijeme do pojave neke nesreće, itd.

Naime, analizirajmo slučajnu varijablu X koja modelira vrijeme potrebno do pojave danog događaja pod sljedećom pretpostavkom: *uvjetna vjerojatnost pojavljivanja događaja u nekom kratkom intervalu* $\langle t, t+\Delta t\rangle$, *uz uvjet da se nije dogodio prije* t, *proporcionalna je duljini tog intervala i ne ovisi o* t, *tj.*

$$P\left(\{X \le t+\Delta t\}|\{X>t\}\right) = \lambda\Delta t, \quad \lambda > 0.$$

Vjerojatnost da je vrijeme do pojave događaja veća od $t+\Delta t$ jest

$$P\{X > t+\Delta t\} = P\left(\{X > t+\Delta t\}|\{X>t\}\right) P\{X>t\}.$$

Međutim, pod tom je pretpostavkom

$$P\left(\{X > t+\Delta t\}|\{X>t\}\right) = 1 - P\left(\{X \le t+\Delta t\}|\{X>t\}\right) = 1-\lambda\Delta t,$$

pa vrijedi

$$P\{X > t+\Delta t\} = (1-\lambda\Delta t)P\{X>t\}.$$

Označimo funkciju $S(t) = P\{X>t\}$. Ta funkcija, za male Δt, treba zadovoljavati svojstvo

$$S(t+\Delta t) = (1-\lambda\Delta t)S(t),$$

tj.

$$\frac{S(t+\Delta t)-S(t)}{\Delta t} = -\lambda S(t),$$

što u limesu, za $\Delta t \to 0$, znači da mora zadovoljavati diferencijalnu jednadžbu

$$S'(t) = -\lambda S(t).$$

Rješenje te diferencijalne jednadžbe jest $S(t) = Ce^{-\lambda t}$, a konstantu možemo odrediti iz prirodnog uvjeta $S(0) = P\{X>0\} = 1$. Dakle,

$$S(t) = P\{X>t\} = e^{-\lambda t},$$

što znači da funkcija distribucije slučajne varijable X ima oblik

$$F(t) = 1 - P\{X>t\} = 1 - e^{-\lambda t}.$$

To je upravo funkcija distribucije slučajne varijable koju zovemo eksponencijalna s parametrom $\lambda > 0$.

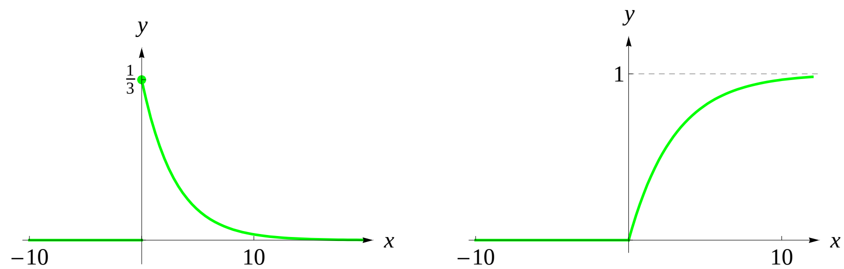
Definicija 2.9. *Neprekidna slučajna varijabla X ima* **eksponencijalnu distribuciju** *s parametrom $\lambda > 0$ ako je funkcija gustoće dana izrazom*

$$f(x) = \begin{cases} 0 & , x < 0 \\ \lambda e^{-\lambda x} & , x \geq 0 \end{cases}.$$

Funkcija distribucije te slučajne varijable dana je izrazom

$$F(x) = \begin{cases} 0 & , x < 0 \\ 1 - e^{-\lambda x} & , x \geq 0 \end{cases}.$$

Primjer 2.23. *Vrijeme u sekundama koje protekne od servisa do prvog udara teniske loptice u tlo modelirano je eksponencijalnom slučajnom varijablom s parametrom $\lambda = 1/3$. Grafovi funkcije gustoće i funkcije distribucije te slučajne varijable dani su na slici 2.16*



Slika 2.16: Graf funkcije gustoće i funkcije distribucije eksponencijalne distribucije s parametrom $\lambda = 1/3$.

Vjerojatnost da će od servisa do prvog udara teniske loptice u tlo proći više od dvije, ali najviše četiri sekunde, računamo na sljedeći način:

$$P\{2 < X \leq 4\} = P\{X \leq 4\} - P\{X \leq 2\} = \frac{1}{3}\int_2^4 e^{-x/3}\,dx \approx 0.249.$$

2.5.3 Dvostrana eksponencijalna distribucija

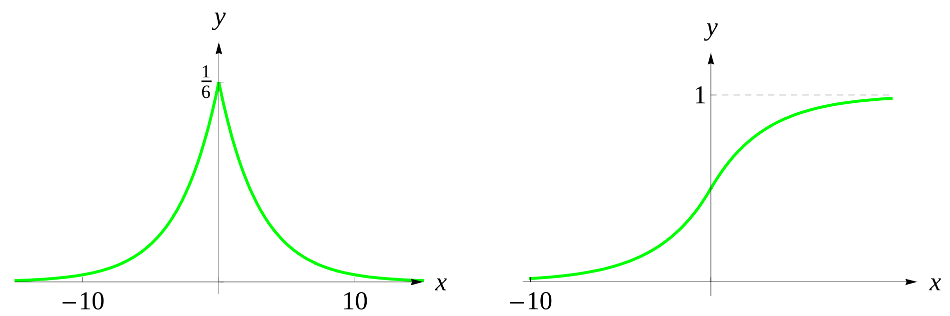
Definicija 2.10. *Neprekidna slučajna varijabla X ima* **dvostranu eksponencijalnu distribuciju** *s parametrom $\lambda > 0$ ako je funkcija gustoće dana izrazom*

$$f(x) = \frac{1}{2}\lambda e^{-\lambda|x|}, \quad x \in \mathbb{R}.$$

Funkcija distribucije te slučajne varijable definirana je sljedećim izrazom:

$$F(x) = \begin{cases} \frac{1}{2}e^{\lambda x} & , x \leq 0 \\ 1 - \frac{1}{2}e^{-\lambda x} & , x > 0 \end{cases}.$$

Primjer 2.24. *Grafovi funkcije gustoće i funkcije distribucije slučajne varijable X s dvostranom eksponencijalnom distribucijom s parametrom $\lambda = \frac{1}{3}$ prikazani su slikom 2.17.*



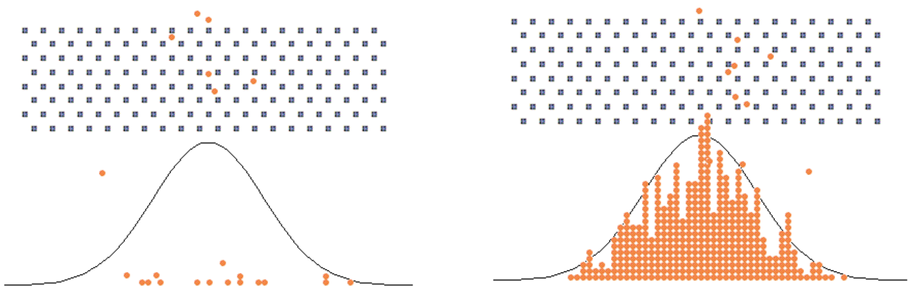
Slika 2.17: Graf funkcije gustoće i funkcije distribucije dvostrane eksponencijalne distribucije s parametrom $\lambda = 1/3$.

2.5.4 Normalna distribucija

Normalna distribucija najviše se koristi u statističkoj teoriji i primjeni. Teorija je pokazala da ona vrlo dobro opisuje pokuse čiji su ishodi posljedica sume mnogo međusobno nezavisnih i jednako distribuiranih utjecaja. Objasnimo tu činjenicu primjerom.

Primjer 2.25 (Galtonova daska[2]). *U dasku ubodemo pribadače kao što je prikazano na slici 2.18 i iz izvora pri vrhu konstrukcije pustimo da pada mnoštvo malih kuglica. U podnožju ih zaustavljamo podlogom i određujemo funkciju koja opisuje gustoću kuglica na toj podlozi. Jasno je da su kuglice mnogo puta udarile u pribadače prije negoli su pale i svaki su put promjenile smjer kretanja. Njihov pomak u stranu od mjesta s kojeg su bačene nastao je kao suma mnoštva malih jednako distribuiranih i nezavisnih pomaka uzrokovanih sudarima. Dobivena gustuća kuglica na dasci ima upravo zvonoliki oblik Gaussove krivulje koja je graf funkcije gustoće normalne distribucije.*

[2] Tim eksperimentom engleski znanstvenik Sir Francis Galton (1822-1911) demonstrirao je centralni granični teorem. Eksperiment je poznat i pod nazivima "bean machine", "quincunx" i "Galtonova kutija".



Slika 2.18: Galtonova daska.

Definicija 2.11. *Za neprekidnu slučajnu varijablu kažemo da ima* **Gaussovu** *ili* **normalnu** *distribuciju s parametrima μ i σ^2 ako je njezina funkcija gustoće dana izrazom*

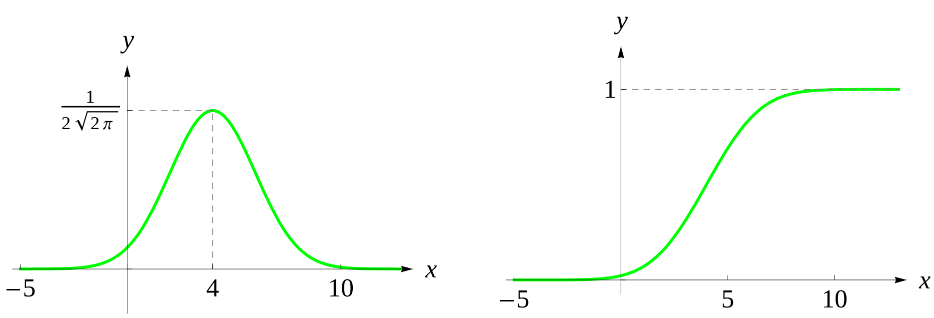
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R},$$

gdje su μ i σ realni brojevi i $\sigma > 0$. Ako slučajna varijabla X ima normalnu distribuciju s parametrima μ i σ^2, koristimo oznaku $X \sim \mathcal{N}(\mu, \sigma^2)$.

Funkcija distribucije normalne slučajne varijable (slika 2.19) definirana je izrazom

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt, \quad x \in \mathbb{R}.$$

Slika 2.19 prikazuje graf funkcije gustoće i funkcije distribucije normalne distribucije s parametrima $\mu = 4$ i $\sigma^2 = 4$.

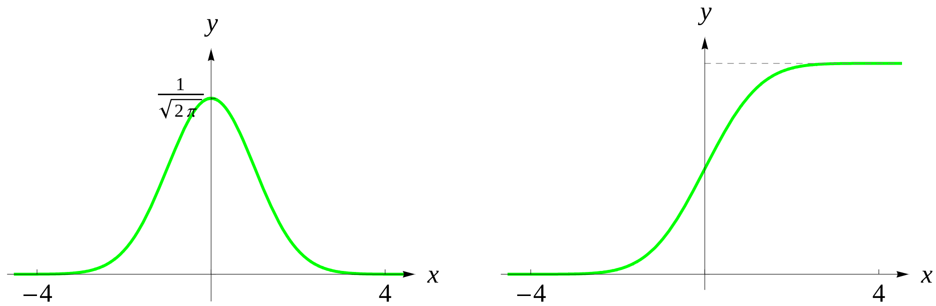


Slika 2.19: Graf funkcije gustoće i funkcije distribucije slučajne varijable $X \sim \mathcal{N}(4, 4)$.

Normalna distribucija s parametrima $\mu = 0$ i $\sigma^2 = 1$ zove se **jedinična** ili **standardna normalna distribucija**. Gustoća standardne normalne distribucije dana je izrazom

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}.$$

Slika 2.20 prikazuje graf funkcije gustoće i funkcije distribucije standardne normalne distribucije.



Slika 2.20: Graf funkcije gustoće i funkcije distribucije standardne normalne distribucije.

Vrijednosti funkcije distribucije za takve slučajne varijable moraju se računati korištenjem metoda numeričkog integriranja s obzirom da se integrali uglavnom ne daju riješiti eksplicitno. U većini knjiga koje primjenjuju statistiku dane su tablice vrijednosti funkcije distribucije standardne normalne slučajne varijable koja je definirana sljedećim izrazom:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt, \quad x \in \mathbb{R}.$$

U današnje se vrijeme za računanje vjerojatnosti po normalnoj distribuciji najčešće koriste naprednija džepna računala ili specijalizirani software na računalu.

Primjer 2.26. *Vrijeme (u satima) koje student treće godine studija matematike provede učeći u jednom danu može se opisati slučajnom varijablom X koja ima normalnu distribuciju s parametrima $\mu = 5.43$ i $\sigma = 0.7$. Korištenjem programskog paketa Mathematica i ugrađene funkcije $CDF[\ ]$ [3] slijedi:*

a) vjerojatnost da student učeći provede manje od 5 sati:

$$P\{X < 5\} = CDF[ndist, 5] \approx 0.269,$$

gdje je $ndist = NormalDistribution[5.43, 0.7]$.

[3] CDF = cumulative distribution function.

b) vjerojatnost da student učeći provede barem 3 sata:

$$P\{X \geq 3\} = 1 - P\{X < 3\} = 1 - CDF[ndist, 3] \approx 0.999.$$

c) vjerojatnost da student učeći provede između 3 i 7 sati:

$$P\{3 < X < 7\} = CDF[ndist, 7] - CDF[ndist, 3] \approx 0.987.$$

2.6 Numeričke karakteristike slučajne varijable

Iz prethodnih razmatranja jasno je da za opisivanje neke jednodimenzionalne veličine, koja ima slučajan karakter kao model može poslužiti slučajna varijabla. Za njezino zadavanje potrebno je zadati distribuciju pa su tada u potpunosti određena vjerojatnosna svojstva. Distribucije mogu biti jednostavne, ali i vrlo složenog karaktera i nije uvijek jednostavno predočiti zakonitosti ponašanja slučajne karakteristike ispisivanjem njezine distribucije. Razvojem teorije vjerojatnosti postalo je jasno da je često za slučajnu varijablu moguće definirati nekoliko karakterističnih brojeva koji mogu, zbog svojih generalnih svojstava, dodatno pomoći u opisivanju slučajne varijable. Takve karakteristične brojeve zvat ćemo **numeričke karakteristike** slučajne varijable.

2.6.1 Očekivanje diskretne slučajne varijable

Osnovna je numerička karakteristika slučajne varijable matematičko očekivanje.

Definicija 2.12. *Neka je $(\Omega, \mathcal{P}(\Omega), P)$ diskretan vjerojatnosni prostor i X slučajna varijabla na njemu. Ako red $\sum\limits_{\omega\in\Omega} X(\omega)P\{\omega\}$ apsolutno konvergira (tj. ako konvergira red $\sum\limits_{\omega\in\Omega} |X(\omega)|P\{\omega\}$), onda kažemo da slučajna varijabla X ima matematičko očekivanje i broj*

$$EX = \sum_{\omega\in\Omega} X(\omega)P\{\omega\}$$

zovemo **matematičko očekivanje (očekivanje)** *slučajne varijable X.*

Navedena definicija matematičkog očekivanja ne može se jednostavno primijeniti za njegovo računanje s obzirom da je dana zadavanjem slučajne varijable na temeljnom vjerojatnosnom prostoru. Za računanje je matematičkog očekivanja puno prikladnija formula koja se može dati na temelju tablice distribucije diskretne slučajne varijable, o čemu govori sljedeći teorem.

Teorem 2.1. *Neka je* $(\Omega, \mathcal{P}(\Omega), P)$ *diskretan vjerojatnosni prostor i*

$$X = \begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{pmatrix}$$

slučajna varijabla na njemu. Redovi $\sum_{\omega\in\Omega} X(\omega)P\{\omega\}$ *i* $\sum_{i\in\mathbb{N}} x_i p_i$ *istovremeno ili apsolutno konvergiraju ili apsolutno divergiraju. U slučaju apsolutne konvergencije sume su im jednake, tj. vrijedi*

$$EX = \sum_{\omega\in\Omega} X(\omega)P\{\omega\} = \sum_{i\in\mathbb{N}} x_i p_i.$$

Dokaz. Koristeći činjenicu da familija skupova $\{\{X = x_i\}, i \in \mathbb{N}\}$ čini particiju skupa Ω (jer je s X zadana funkcija!), vidimo da se jedan red može dobiti iz drugog "preslagivanjem članova" u smislu teorije ponovljenih redova, tj.

$$\sum_{\omega\in\Omega} X(\omega)P\{\omega\} = \sum_{i\in\mathbb{N}} \sum_{\{X(\omega)=x_i\}} X(\omega)P\{\omega\} = \sum_{i\in\mathbb{N}} \sum_{\{X(\omega)=x_i\}} x_i P\{\omega\} =$$

$$= \sum_{i\in\mathbb{N}} x_i \sum_{\{X(\omega)=x_i\}} P\{\omega\} = \sum_{i\in\mathbb{N}} x_i p_i.$$

Korištenjem rezultata poglavlja 5.3 (Dodatak) zaključujemo da ti redovi istovremeno ili apsolutno konvergiraju ili apsolutno divergiraju, a u slučaju apsolutne konvergencije sume su im jednake.

Primjer 2.27. *Neka slučajna varijabla* X *opisuje rezultate bacanja pravilno izrađene igraće kockice. Očekivanje je te slučajne varijable broj*

$$EX = \sum_{i=1}^{6} \frac{1}{6} i = \frac{21}{6}.$$

Ako je zadana slučajna varijabla X na $(\Omega, \mathcal{P}(\Omega), P)$ i funkcija $g : \mathcal{R}(X) \to \mathbb{R}$, onda je kompozicijom $g(X)$ također dana slučajna varijabla na istom vjerojatnosnom prostoru. Npr. X^2, $2X + 3$, e^X,... su tako nastale slučajne varijable. Za računanje njihovih očekivanja (ako postoje) neće biti potrebno specificirati njihove distribucije, već se može iskoristiti distribucija slučajne varijable X. Naime, ako red $\sum_{\omega\in\Omega} g(X(\omega))P\{\omega\}$ apsolutno konvergira, onda postoji očekivanje $Eg(X)$. Međutim, uobičajenim "preslagivanjem" koristeći particiju $\{\{X = x_i\}, i \in \mathbb{N}\}$ od Ω i teoriju ponovljenih redova, vidimo da vrijedi:

$$\sum_{\omega\in\Omega} g(X(\omega))P\{\omega\} = \sum_{i\in\mathbb{N}} g(x_i)p_i.$$

Time smo dokazali sljedeći koristan rezultat.

Teorem 2.2. *Neka je* $(\Omega, \mathcal{P}(\Omega), P)$ *diskretan vjerojatnosni prostor,*

$$X = \begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{pmatrix}$$

slučajna varijabla na njemu i $g : \mathcal{R}(X) \to \mathbb{R}$ *funkcija takva da postoji* $Eg(X)$*. Tada vrijedi:*

$$Eg(X) = \sum_{i \in \mathbb{N}} g(x_i) p_i.$$

Koristeći taj rezultat i definiciju očekivanja, možemo dokazati da vrijede sljedeća korisna svojstva očekivanja:

1. Neka su a i b realni brojevi, a X slučajna varijabla koja ima očekivanje. Tada i slučajna varijabla $aX + b$ ima očekivanje i vrijedi:

$$E(aX + b) = aEX + b.$$

2. Ako su X i Y dvije slučajne varijable koje imaju očekivanja i ako vrijedi $X(\omega) \leq Y(\omega)$ za sve $\omega \in \Omega$ onda je i $EX \leq EY$ (**monotonost očekivanja**).

3. Ako je X slučajna varijabla koja ima svojstvo $X(\omega) \geq 0$ i ako je red $\sum\limits_{\omega \in \Omega} X(\omega) P\{\omega\}$ konvergentan, tada je $EX \geq 0$ (**pozitivnost očekivanja**).

Osim navedenih svojstava, očekivanje ima i svojstvo linearnosti. Naime, vrijedi sljedeći teorem:

Teorem 2.3. *Neka su* X *i* Y *dvije slučajne varijable na vjerojatnosnom prostoru* $(\Omega, \mathcal{P}(\Omega), P)$ *koje imaju očekivanja* EX*, odnosno* EY*. Tada za proizvoljne* $a, b \in \mathbb{R}$*, slučajna varijabla* $aX + bY$ *također ima očekivanje i vrijedi:*

$$E(aX + bY) = aEX + bEY.$$

Dokaz. Iz apsolutne konvergencije redova $\sum\limits_{\omega \in \Omega} X(\omega) P\{\omega\}$ i $\sum\limits_{\omega \in \Omega} Y(\omega) P\{\omega\}$ slijedi apsolutna konvergencija reda $\sum\limits_{\omega \in \Omega} (aX(\omega) + bY(\omega)) P\{\omega\}$. Tvrdnja teorema slijedi iz poznatih rezultata o sumi konvergentnih redova (vidi npr. [13]).

Koristeći princip matematičke indukcije može se dokazati i generalizacija teorema 2.3 na linearnu kombinaciju konačno mnogo diskretnih slučajnih varijabli koje imaju očekivanje. Formulirajte tvrdnju i dokažite je!

Primjer 2.28. *Pretpostavimo da izvodimo slučajan pokus bacanja nepravilno izrađenog novčića kod kojega je relizacija pisma favorizirana i događa se s vjerojatnošću* 0.7 *(koristit ćemo oznaku* 1 *za realizaciju pisma i* 0 *za realizaciju glave). Tada slučajna varijabla X kojom modeliramo taj pokus ima Bernoullijevu distribuciju s parametrom $p = 0.7$, tj. zadana je tablicom distribucije*

$$X = \begin{pmatrix} 0 & 1 \\ 0.3 & 0.7 \end{pmatrix}.$$

Slučajna varijabla Y koja opisuje broj realiziranih pisama nakon 100 *nezavisnih bacanja toga novčića ima binomnu distribuciju s parametrima $n = 100$ i $p = 0.7$, tj. $Y \sim \mathcal{B}(100, 0.7)$. Kako binomnu slučajnu varijablu možemo predstaviti kao sumu n Bernoullijevih slučajnih varijabli, tj.*

$$Y = \sum_{i=1}^{n} X_i, \quad X_i \text{ jednako distribuirana kao } X, \quad \forall i,$$

računanje je matematičkog očekivanja od Y znatno pojednostavljeno i svodi se na primjenu svojstva aditivnosti očekivanja:

$$EX = \sum_{i=1}^{100} EX_i = \sum_{i=1}^{100}(0 \cdot 0.3 + 1 \cdot 0.7) = \sum_{i=1}^{100} 0.7 = 100 \cdot 0.7 = 70.$$

Za očekivanje obično kažemo da je jedna mjera centralne tendencije slučajne varijable. Naime, ako se pitamo postoji li realan broj koji je na neki način "najbliži" slučajnoj varijabli, onda je to u jednom načinu mjerenja udaljenosti upravo očekivanje. Pritom udaljenost između slučajne varijable i broja mjerimo kao očekivano kvadratno odstupanje: $E(X-a)^2$.[4] Zaista, definiramo li funkciju $g(a) = E(X-a)^2 = EX^2 - 2aEX + a^2$ na $\mathbb{R}$, vidimo da ona postiže minimum upravo u vrijednosti $a = EX$.

Primjer 2.29. *Ilustrirajmo značenje očekivanog kvadratnog odstupanja slučajne varijable od konstante na primjerima.*

a) Neka je slučajna varijabla X zadana tablicom distribucije

$$X = \begin{pmatrix} 1 & 2 & 3 & 4 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

Tada je

$$E(X-a)^2 = (1-a)^2\frac{1}{4} + (2-a)^2\frac{1}{4} + (3-a)^2\frac{1}{4} + (4-a)^2\frac{1}{4}$$

funkcija varijable a. Minimum ove funkcije je točno prosjek vrijednosti koje slučajna varijabla može postići, tj. 2.5. *Uočimo da se ovdje sve vrijednosti reliziraju s istom vjerojatnošću (slika 2.21).*

b) Neka je slučajna varijabla X zadana tablicom distribucije

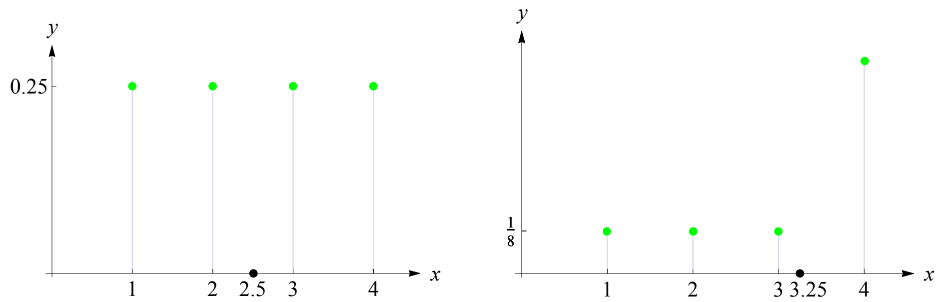
$$X = \begin{pmatrix} 1 & 2 & 3 & 4 \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{5}{8} \end{pmatrix}.$$

Tada

$$E(X-a)^2 = (1-a)^2\frac{1}{8} + (2-a)^2\frac{1}{8} + (3-a)^2\frac{1}{8} + (4-a)^2\frac{5}{8}$$

kao funkcija varijable a postiže minimum koji nije jednak prosjek vrijednosti koje slučajna varijabla može primiti jer se vrijednost 4 *realizira bitno većom vjerojatnošću nego ostale vrijednosti iz $\mathcal{R}(X)$. Ovdje je minimum za $a = 3.25$, tj. "povučen" je prema* 4 *(slika 2.21).*

[4] Više o toj temi možete pogledati u [26].



Slika 2.21: Grafovi distribucija slučajnih varijabli iz primjera 2.29.

Minimalno očekivano kvadratno odstupanje slučajne varijable od konstante, koje se postiže u očekivanju, također ima svoje značenje. Zove se varijanca i tema je sljedećeg poglavlja.

2.6.2 Varijanca i ostali momenti. Važne nejednakosti

Koristeći definiciju očekivanja, za slučajnu varijablu definiramo momente i centralne momente.

Definicija 2.13. *Neka je X slučajna varijabla na diskretnom vjerojatnosnom prostoru $(\Omega, \mathcal{P}(\Omega), P)$ i $r > 0$.*

- *Ako postoji $E(X^r)$, onda broj $\mu_r = E(X^r)$ zovemo* **moment r-tog reda** *ili* **r-ti moment** *od X.*
- *Ako postoji $E(|X|^r)$, onda broj $E(|X|^r)$ zovemo* **apsolutni moment r-tog reda** *ili* **r-ti apsolutni moment** *od X.*
- *Ako postoji EX i $E(|X - EX|^r)$ onda broj $E(|X - EX|^r)$ zovemo* **r-ti centralni moment** *od X.*
- *Ako postoji $E(X - EX)^2$, onda taj nenegativan broj zovemo* **varijanca** *slučajne varijable X i označavamo* $\operatorname{Var} X$, σ_X^2 *ili* σ^2.

Uočimo:

- Očekivanje je slučajne varijable X njezin moment prvog reda. Očekivanje uobičajeno označavamo s μ umjesto μ_1.
- Varijanca je drugi centralni moment. Ona predstavlja očekivano kvadratno odstupanje slučajne varijable od njezinog očekivanja.

Primjer 2.30. *Neka je slučajna varijabla X zadana tablicom distribucije*

$$X = \begin{pmatrix} -a & a & a^2 \\ 0.3 & 0.3 & 0.4 \end{pmatrix}.$$

Izračunajmo momente EX, EX^2, EX^3, $\operatorname{Var} X$, $E(X-E(X))^3$ te diskretne slučajne varijable:

$EX = -0.3a + 0.3a + 0.4a^2 = 0.4a^2,$

$EX^2 = 0.6a^2 + 0.4a^4 = 0.2a^2(2a^2+3),$

$EX^3 = -0.3a^3 + 0.3a^3 + 0.4a^6 = 0.4a^6,$

$\operatorname{Var} X = E\left((X-EX)^2\right) = EX^2 - (EX)^2 = 0.24a^2(a^2+2.5),$

$E(X-EX)^3 = E\left(X^3\right) - 3E\left(X^2\right) EX + 2\,(EX)^3 = 0.464a^6 - 0.6a^2(2a^2+3).$

Propozicija 2.1. *Neka je $r > 0$ i $E(|X|^r)$ postoji. Tada postoji i $E(|X|^s)$ za svaki $0 < s < r$.*

Dokaz. *Za svaki realan broj x i $0 < s < r$ vrijedi nejednakost:*

$$|x|^s < 1 + |x|^r.$$

Odavde i iz monotonosti očekivanja vidimo da vrijedi:

$$|X|^s \leq 1 + |X|^r \implies E(|X|^s) \leq 1 + E(|X|^r) < \infty,$$

pa postoji očekivanje od X^s.

Kao ilustraciju mogućih interpretacija momenata navodimo tzv. Čebiševljevu nejednakost koja daje korisnu interpretaciju očekivanja i varijance svake slučajne varijable koja ima varijancu.

Propozicija 2.2 (Čebiševljeva nejednakost)**.** *Neka je X slučajna vrijabla koja ima varijancu σ^2 te neka je dan broj $k > 0$. Tada vrijedi:*

$$P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2},$$

gdje je μ očekivanje slučajne varijable X.

U dokazu te propozicije iskoristit ćemo sljedeću nejednakost koja je općenito vrlo korisna u teoriji vjerojatnosti.

Teorem 2.4. *Neka je X slučajna varijabla na vjerojatnosnom prostoru $(\Omega, \mathcal{P}(\Omega), P)$ i g nenegativna funkcija definirana na $\mathcal{R}(X)$ takva da postoji $Eg(X)$. Tada za svaki $\varepsilon > 0$ vrijedi:*

$$P\{g(X) \geq \varepsilon\} \leq \frac{Eg(X)}{\varepsilon}.$$

Dokaz. *Prije svega, uočimo jednu općenito korisnu činjenicu: ako je* $A \in \mathcal{P}(\Omega)$ *neki skup, označimo li*

$$I_A(\omega) = \begin{cases} 1 \ , \ \omega \in A \\ 0 \ , \ \omega \notin A \end{cases},$$

I_A *je slučajna varijabla na* Ω *koju možemo iskoristiti za povezivanje vjerojatnosti skupa* A *i pojma očekivanja slučajne varijable. Naime, distribucija te slučajne varijable dana je tablicom*

$$I_A = \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}, \quad p = P\{I_A = 1\} = P\{A\}.$$

Dakle, vrijedi:

$$EI_A = P(A).$$

Korištenjem tog načina označavanja, monotonosti i linearnosti očekivanja te nenegativnosti funkcije g *vidimo da za svaki* $\varepsilon > 0$ *vrijedi:*

$$\begin{aligned} Eg(X) &= E(g(X)I_{\{g(X)\geq\varepsilon\}} + g(X)I_{\{g(X)<\varepsilon\}}) \\ &= E(g(X)I_{\{g(X)\geq\varepsilon\}}) + E(g(X)I_{\{g(X)<\varepsilon\}}) \\ &\geq E(g(X)I_{\{g(X)\geq\varepsilon\}}) \geq \varepsilon EI_{\{g(X)\geq\varepsilon\}} \\ &= \varepsilon P\{g(X) \geq \varepsilon\}, \end{aligned}$$

što dokazuje tvrdnju.

Dokaz (Čebiševljeva nejednakost)**.** *Iz prethodnog teorema slijedi:*

$$\begin{aligned} P\{|X - EX| \geq k\sigma\} &= P\{|X - EX|^2 \geq k^2\sigma^2\} \\ &\leq \frac{E|X - EX|^2}{k^2\sigma^2} = \frac{\operatorname{Var} X}{k^2\sigma^2} = \frac{1}{k^2}. \end{aligned}$$

Interpretacija se varijance i očekivanja korištenjem Čebiševljeve nejednakosti zapravo temelji na drugom korijenu iz varijance. Tu veličinu, tj. $\sqrt{\operatorname{Var} X}$ zovemo **standardna devijacija** slučajne varijable i označavamo sa σ_X ili σ.

Dakle, vidimo da Čebiševljeva nejednakost tvrdi: **vjerojatnost da odstupanje slučajne varijable X od njezinog očekivanja bude po apsolutnoj vrijednosti veće ili jednako k standardnih devijacija, manja je ili jednaka od $1/k^2$.**

Na primjer, vjerojatnost je manja ili jednaka $1/9 \approx 11.1\%$ da slučajna varijabla odstupa od svog očekivanja za više od 3 standardne devijacije. Korištenjem statističke interpretacije vjerojatnosti možemo tada zaključiti da će se, prilikom puno nezavisnih realizacija slučajne varijable X, u ne više od 11.1% slučajeva realizirati vrijednosti izvan intervala $[\mu - 3\sigma, \mu + 3\sigma]$, a u barem 88.9% slučajeva vrijednosti unutar tog intervala. Kao što se može vidjeti iz navedenih objašnjenja, ima smisla standardnu devijaciju smatrati jednom od mjera koja govori o raspršenosti realizacija slučajne varijable oko očekivanja, tj. mjerom raspršenja.

Primjer 2.31. *Neka je X diskretna slučajna varijabla kojom je modeliran broj kišnih dana u Osijeku tijekom jedne godine. Poznato je da je očekivani broj kišnih dana 128, a standardna devijacija 4. Ako želimo ocijeniti $P\{|X - EX| < 5\}$, tj. vjerojatnost da broj kišnih dana u Osijeku tijekom jedne godine odstupa od očekivanog broja za manje od pet dana, koristimo Čebiševljevu nejednakost. Iz Čebiševljeve nejednakosti dane u propoziciji 2.2 primjenom svojstva vjerojatnosti suprotnog događaja slijedi da je*

$$P\{|X - EX| < k\sigma\} \geq 1 - \frac{1}{k^2}.$$

Budući da je u tom primjeru $EX = 128$, $\sigma = 4$ i $k\sigma = 5$, tj. $k = 5/4$, slijedi da je

$$P\{|X - 128| < 5\} \geq 0.36.$$

Primjer 2.32. *Neka je X diskretna slučajna varijabla kojom je modeliran broj komaraca uhvaćen u danu klopku u jednom satu tijekom lipnja u Osijeku. Pretpostavimo da je poznato da je u opisanim uvjetima očekivani broj komaraca uhvaćenih u tu klopku po satu jednak 50, a varijanca 64 te da je*

$$P\{|X - 50| < 8k\} \geq 0.5$$

za neki neki $k \in \langle 0, 5\rangle$. Pomoću zadanih veličina možemo odrediti granice intervala $\langle 50 - 8k, 50 + 8k\rangle$ čijim se cjelobrojnim elementima, prema Čebiševljevoj nejednakosti, X realizira vjerojatnošću većom ili jednakom 0.5. Rješavanjem jednadžbe

$$1 - \frac{1}{k^2} = 0.5$$

slijedi da je parametar k iz Čebiševljeve nejednakosti jednak $\sqrt{2}$. Prema tome je

$$P\left\{50 - 8\sqrt{2} < X < 50 + 8\sqrt{2}\right\} \geq 0.5.$$

Budući da je $(50 - 8\sqrt{2}) \approx 38.686$, a $(50 + 8\sqrt{2}) \approx 61.314$, slijedi da je vjerojatnost da se X realizira brojem iz skupa $\{39, 40, \ldots, 61\}$ veća ili jednaka 0.5. U kontekstu našeg promatranja to znači da je vjerojatnost da u opisanu klopku bude ulovljeno n, $n \in \{39, 40, \ldots, 61\}$, komaraca veća ili jednaka 0.5.

Primjer 2.33. *Neka je dana slučajna varijabla X koja ima očekivanje 0.7 i standardnu devijaciju 0.458. Tada Čebiševljeva nejednakost garantira: vjerojatnost odstupanja te slučajne varijable od 0.7 za više od dvije standardne devijacije iznosi najviše 0.25. Naime*

$$P\{|X - 0.7| \geq 2\sigma\} \leq \frac{1}{4}.$$

Korištenjem statističke interpretacije vjerojatnosti možemo zaključiti da će se, prilikom puno nezavisnih realizacija te slučajne varijable, najviše u 25% slučjeva realizirati vrijednosti izvan intervala $[0.7 - 2 \cdot 0.458, 0.7 + 2 \cdot 0.458] = [0.2417, 1.1583]$, a u barem 75% slučajeva vrijednosti unutar tog intervala. (Odredite što kraći interval koji će sadržavati barem 88% realizacija!)

Dakako, ta ocjena nije precizna s obzirom da se odnosi na sve slučajne varijable sa zadanim iznosom očekivanja i standardne devijacije bez obzira na tip distribucije. Ukoliko je točno poznata distribucija slučajne varijable, takve se vjerojatnosti se mogu izračunati i puno preciznije, što će biti ilustrirano u narednim poglavljima.

S obzirom da je varijanca moment koji ćemo često koristiti, dokažimo nekoliko korisnih svojstava.

1. Neka je X slučajna varijabla koja ima varijancu te a i b proizvoljni realni brojevi. Tada vrijedi
$$\operatorname{Var}(aX+b) = a^2 \operatorname{Var} X.$$

   Dokaz.
$$\begin{aligned}\operatorname{Var}(aX+b) &= E\left((aX+b)-(aEX+b)\right)^2 = \\ &= E\left(aX-aEX\right)^2 = a^2E(X-EX)^2 = a^2\operatorname{Var} X.\end{aligned}$$

2. Ako za slučajnu varijablu X vrijedi $\operatorname{Var} X = 0$, onda ona zapravo nema karakter slučajnosti, tj. $P\{X = konst.\} = 1$. Očigledno je da vrijedi i drugi smjer te tvrdnje, tj. ako za slučajnu varijablu X vrijedi $P\{X = konst.\} = 1$, onda je $\operatorname{Var} X = 0$.

   Dokaz. *Neka je* $\operatorname{Var} X = 0$. *Tada vrijedi:* $E(X-\mu)^2 = 0$. *S obzirom da je* $(X-\mu)^2 \geq 0$, *na skupu na kojemu je* $(X-\mu)^2 > 0$ *mora biti pripadna vjerojatnost* 0, *tj.* $P\{(X-\mu)^2 > 0\} = 0$. *Odavde slijedi da je* $P\{(X-\mu)^2 = 0\} = 1$, *tj.* $P\{X = \mu\} = 1$, *što dokazuje prvu tvrdnju. Za dokaz druge tvrdnje*

   pretpostavimo $P\{X = c\} = 1$. *Tada je* $EX = c$, *a* $\operatorname{Var} X = E(X-c)^2 = 0$.

3. Varijancu možemo računati također primjenom formule
$$\operatorname{Var} X = EX^2 - (EX)^2.$$

   Dokaz. *Neka je* $EX = \mu$. *Tada zbog linearnosti očekivanja vrijedi:*
$$\operatorname{Var} X = E(X-\mu)^2 = E(X^2 - 2\mu EX + \mu^2) = EX^2 - \mu^2.$$

2.6.3 Očekivanje i varijanca nekih parametarskih diskretnih distribucija

Parametarski se zadane familije distribucija često koriste u praksi, a njihovi se parametri nerijetko mogu iskazati u terminima momenata. U ovom poglavlju izračunat ćemo očekivanje i varijancu za diskretne parametarske familije definirane u poglavlju 2.4.

Diskretna uniformna distribucija

Za slučajnu varijablu X rekli smo da ima diskretnu uniformnu distribuciju ako je njezin skup vrijednosti konačan podskup skupa realnih brojava , tj. $\mathcal{R}(X) = \{x_1, x_2, \ldots, x_n\} \subseteq \mathbb{R}$, a pripadni niz vjerojatnosti definiran je na sljedeći način:

$$p_i = P\{X = x_i\} = \frac{1}{n}, \quad i = 1, 2, \ldots, n.$$

Očekivanje je te slučajne varijable aritmetička sredina elemenata skupa $\mathcal{R}(X)$, tj.

$$EX = \sum_{i=1}^{n} x_i p_i = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}_n.$$

Za varijancu tada vrijedi:

$$\operatorname{Var} X = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x}_n)^2.$$

Specijalno, ako je skup vrijednosti $\{1, 2, 3, \ldots, n\}$, tada je:

$$EX = \frac{n+1}{2}, \quad \operatorname{Var} X = \frac{n^2 - 1}{12}.$$

Bernoullijeva distribucija

Za slučajnu varijablu rekli smo da ima Bernoullijevu distribuciju ako je zadana tablicom distribucije

$$X = \begin{pmatrix} 0 & 1 \\ 1 - p & p \end{pmatrix}, \quad p \in \langle 0, 1 \rangle.$$

Očekivanje i varijanca te distribucije dani su sljedećim izrazima:

$$EX = p, \ \operatorname{Var} X = p(1 - p).$$

Uočimo da je slučajna varijabla I_A iz dokaza teorema 2.4 Bernoullijeva slučajna varijabla s parametrom $p = P(A)$.

Binomna distribucija

Binomna slučajna varijabla $\mathcal{B}(n,p)$ s parametrima $n \in \mathbb{N}$ i $p \in \langle 0,1\rangle$ definirana je skupom vrijednosti $\{0,1,2,...,n\}$ s pripadnim vjerojatnostima

$$p_i = P\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i}.$$

Označimo $q = 1 - p$.

Prije negoli izračunamo očekivanje i varijancu binomne slučajne varijable uočimo da za fiksne realne brojeva a i b vrijedi sljedeće:

$$g(t) = (at+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i t^i b^{n-i}$$

$$g'(t) = n(at+b)^{n-1}a = \sum_{i=1}^{n} i \binom{n}{i} a^i t^{i-1} b^{n-i}$$

$$g'(1) = n(a+b)^{n-1}a = \sum_{i=1}^{n} i \binom{n}{i} a^i b^{n-i} \tag{2.3}$$

$$g''(t) = n(n-1)(at+b)^{n-2}a^2 = \sum_{i=2}^{n} i(i-1) \binom{n}{i} a^i t^{i-2} b^{n-i}$$

$$g''(1) = n(n-1)(a+b)^{n-2}a^2 = \sum_{i=2}^{n} i(i-1) \binom{n}{i} a^i b^{n-i} \tag{2.4}$$

Primjenom izraza (2.3) dobivamo da je

$$EX = \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} = np(p+1-p)^{n-1} = np.$$

Primjenom izaza (2.3) i (2.4) dobivamo da je

$$\begin{aligned} \operatorname{Var} X &= EX^2 - (EX)^2 = \sum_{i=0}^{n} i^2 \binom{n}{i} p^i (1-p)^{n-i} - (np)^2 \\ &= n(n-1)p^2 + np - (np)^2 = npq. \end{aligned}$$

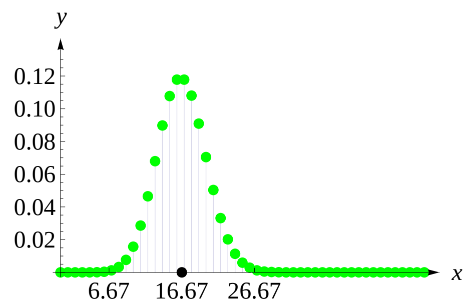
Dakle, očekivanje i varijanca binomne slučajne varijable dani su sljedećim izrazima:

$$E(X) = np, \quad \operatorname{Var} X = npq.$$

Primjer 2.34. *Očekivanje i varijanca binomne slučajne varijable X s parametrima $n = 50$ i $p = \frac{1}{3}$ jesu*

$$\mu = np = \frac{50}{3} \approx 16.67, \quad \sigma^2 = np(1-p) = \frac{100}{9} \approx 11.11,$$

dok je standardna devijacija $\sigma = 10/3 \approx 3.33$. Slikom 2.22 prikazana je distribucija te slučajne varijable s označenim očekivanjem i granicama intervala $[\mu - 3\sigma, \mu + 3\sigma] = [6.67, 26.67]$.



Slika 2.22: Distribucija slučajne varijable $X \sim \mathcal{B}(50, \frac{1}{3})$.

Odredimo vjerojatnost realizacije te slučajne varijable unutar intervala $\langle 6.67, 26.67\rangle$, tj. vjerojatnost realizacije slučajne varijable udaljene od očekivanja za manje od tri standardne devijacije:

$$P\{X \in \langle 6.67, 26.67\rangle\} = \sum_{i=7}^{26} \binom{50}{i} \left(\frac{1}{3}\right)^i \left(\frac{2}{3}\right)^{50-i} \approx 0.997.$$

Dakle, ta će se slučajna varijabla s vjerojatnošću približno 0.997 realizirati unutar intervala $[\mu - 3\sigma, \mu + 3\sigma]$. Korištenjem statističke interpretacije vjerojatnosti možemo zaključiti da će, prilikom puno nezavisnih realizacija te slučajne varijable, njih oko 99.7% pasti unutar tog intervala.

Korištenjem računala odredite najkraći simetričan interval oko očekivanja za koji možemo tvrditi da će sadržati barem 95% realizacije ove slučajne varijable prilikom puno nezavisnih ponavljanja pokusa.

Poissonova distribucija

Slučajna varijabla X ima Poissonovu distribuciju s parametrom $\lambda > 0$ ako prima vrijednosti iz skupa $\{0, 1, 2, \dots\}$ s vjerojatnostima

$$p_i = P\{X = i\} = e^{-\lambda}\frac{\lambda^i}{i!}.$$

Izračunajmo očekivanje i varijancu Poissonove slučajne varijable:

$$EX = \sum_{i=0}^{\infty} i\frac{e^{-\lambda}\lambda^i}{i!} = e^{-\lambda}\sum_{i=0}^{\infty} i\frac{\lambda^i}{i!} = e^{-\lambda}\sum_{i=1}^{\infty} \lambda\frac{\lambda^{i-1}}{(i-1)!} = \lambda e^{-\lambda}e^{\lambda} = \lambda.$$

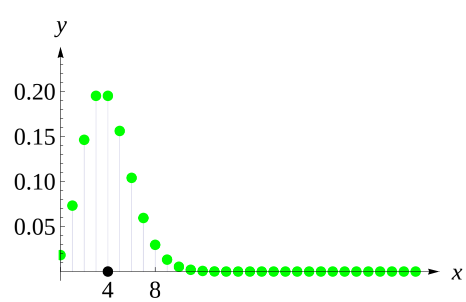
$$
\begin{aligned}
EX^2 &= \sum_{i=0}^{\infty} i^2 \frac{e^{-\lambda}\lambda^i}{i!} = e^{-\lambda}\sum_{i=1}^{\infty} i \frac{\lambda^i}{(i-1)!} = \\
&= \lambda e^{-\lambda}\sum_{i=1}^{\infty}(i-1+1)\frac{\lambda^{i-1}}{(i-1)!} = \\
&= \lambda e^{-\lambda}\left(\lambda\sum_{i=2}^{\infty}\frac{\lambda^{i-2}}{(i-2)!} + \sum_{i=1}^{\infty}\frac{\lambda^{i-1}}{(i-1)!}\right) = \\
&= \lambda e^{-\lambda}[\lambda e^{\lambda} + e^{\lambda}] = \lambda(\lambda+1),
\end{aligned}
$$

$$
\operatorname{Var} X = EX^2 - (EX)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.
$$

Dakle, očekivanje i varijanca Poissonove slučajne varijable dani su sljedećim izrazom:

$$
EX = \operatorname{Var} X = \lambda.
$$

Primjer 2.35. *Očekivanje i varijanca Poissonove slučajne varijable X s parametrom $\lambda = 4$ jednaki su vrijednosti parametra λ, tj. 4, dok je standardna devijacija $\sigma = 2$. Slikom 2.23 prikazana je distribucija te slučajne varijable s označenim očekivanjem i granicama intervala $[\mu - 2\sigma, \mu + 2\sigma] = [0, 8]$.*



Slika 2.23: Distribucija slučajne varijable $X \sim \mathcal{P}(4)$.

Odredimo vjerojatnost realizacije te slučajne varijable unutar intervala $[0, 8]$, tj. vjerojatnost realizacije slučajne varijable udaljene od očekivanja za manje ili točno dvije standardne devijacije:

$$
P\{X \in [0, 8]\} = \sum_{i=0}^{8} e^{-4}\frac{4^i}{i!} \approx 0.979.
$$

Dakle, ta će se slučajna varijabla s vjerojatnošću približno 0.979 realizirati unutar intervala $[\mu - 2\sigma, \mu + 2\sigma]$. Korištenjem statističke interpretacije vjerojatnosti možemo zaključiti da će, prilikom puno nezavisnih realizacija te slučajne varijable, njih oko 97.8% pasti unutar tog intervala.

Odredite najkraći simetričan interval oko očekivanja za koji možemo tvrditi da će sadržati barem 95% realizacija te slučajne varijable prilikom puno nezavisnih ponavljanja pokusa. (Iskoristite računalo!)

Geometrijska distribucija

Slučajna varijabla X ima geometrijsku distribuciju s parametrom $p \in \langle 0,1\rangle$ ako prima vrijednosti iz skupa $\{1,2,3,\ldots\}$ s vjerojatnostima

$$p_k = P\{X=k\} = p(1-p)^{k-1}.$$

Izračunajmo očekivanje i varijancu te distribucije.

Deriviranjem jednakosti

$$\frac{1}{1-x} = \sum_{i=0}^{\infty} x^i, \quad |x|<1, \tag{2.5}$$

po x i primjenom dobivenog izraza uz $x = 1-p$ slijedi

$$EX = \sum_{i=1}^{\infty} ip(1-p)^{i-1} = \frac{1}{p}.$$

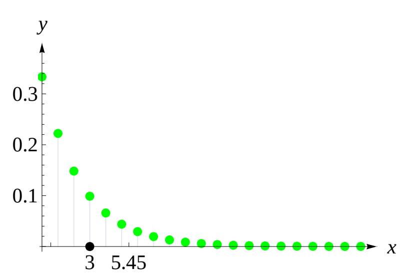
Primjenom druge derivacije izraza (2.5) uz $x = 1-p$ dobivamo da je

$$\operatorname{Var} X = \frac{1-p}{p^2}.$$

Primjer 2.36. *Očekivanje i varijanca geometrijske slučajne varijable X s parametrom $p = 1/3$ jesu*

$$\mu = \frac{1}{p} = 3, \quad \sigma^2 = \frac{1-p}{p^2} = 6,$$

dok je standardna devijacija $\sigma = \sqrt{6} \approx 2.45$. Slikom 2.24 prikazana je distribucija te slučajne varijable s označenim očekivanjem i granicama intervala $[\mu-\sigma, \mu+\sigma] = [0.55, 5.45]$.



Slika 2.24: Distribucija geometrijske slučajne varijable s parametrom $p = \frac{1}{3}$.

Odredimo vjerojatnost realizacije te slučajne varijable unutar intervala $\langle 0.55, 5.45\rangle$, *tj. vjerojatnost realizacije slučajne varijable udaljene od očekivanja za manje od jedne standardne devijacije:*

$$P\{X \in \langle 0.55, 5.45\rangle\} = \sum_{i=1}^{5} ip(1-p)^{i-1} \approx 0.579.$$

Dakle, ta će se slučajna varijabla s vjerojatnošću približno 0.579 *realizirati unutar intervala* $[\mu - \sigma, \mu + \sigma]$. *Korištenjem statističke interpretacije vjerojatnosti možemo zaključiti da će, prilikom puno nezavisnih realizacija te slučajne varijable, njih oko* 57.8% *pasti unutar tog intervala.*

Korištenjem računala odredite najkraći simetričan interval oko očekivanja za koji možemo tvrditi da će sadržati barem 95% realizacije te slučajne varijable prilikom puno nezavisnih ponavljanja pokusa.

2.6.4 Očekivanje i momenti neprekidne slučajne varijable

Za neprekidne slučajne varijable očekivanje se definira korištenjem pripadne funkcije gustoće.

Definicija 2.14. *Neka je X neprekidna slučajna varijabla s funkcijom gustoće f. Ako je konačan integral*

$$\int_{-\infty}^{\infty} |x| f(x)\, dx,$$

onda kažemo da slučajna varijabla X ima očekivanje i broj

$$\mu = EX = \int_{-\infty}^{\infty} x f(x)\, dx$$

zovemo **matematičko očekivanje neprekidne slučajne varijable** *X.*

Valja naglasiti da sva spomenuta svojstva očekivanja (vidi poglavlje 2.29) koja vrijede za diskretne slučajne varijable vrijede i za neprekidne slučajne varijable. Također, ako je dana realna funkcija realne varijable g, onda očekivanje slučajne varijable[5] $g(X)$ možemo računati analogno diskretnom slučaju, ali koristeći funkciju gustoće i integral umjesto sume, tj.

$$Eg(X) = \int_{-\infty}^{\infty} g(x) f(x)\, dx.$$

[5] za uvjete na funkciju g koji osiguravaju da je $g(X)$ neprekidna slučajna vrijabla pogledati primjerice [26]

Definicija je momenata onda identična definiciji momenata u diskretnom slučaju, ali se momenti drugačije računaju s obzirom da se definicije očekivanja razlikuju. Tako je npr. drugi centralni moment, odnosno varijanca, definiran na sljedeći način:

$$Var\,X = \int_{-\infty}^{\infty} (x - EX)^2 f(x)\,dx.$$

Također se može pokazati da pri toj definiciji ostaju sačuvana sva svojstva varijance dokazana u poglavlju 4.19 za diskretan slučaj, kao i navedene korisne nejednakosti (vidi npr. [26]).

Primjer 2.37. *Funkcija gustoće slučajne varijable definirana je pravilom*

$$f(x) = \begin{cases} \sin x & , \ x \in \langle 0, \pi/2] \\ 0 & , \ x \notin \langle 0, \pi/2] \end{cases}.$$

Izračunajmo momente EX, EX^2, EX^3, $\operatorname{Var} X$, $E[X - E(X)]^3$ te neprekidne slučajne varijable:

$$EX = \int_0^{\pi/2} x \sin x\,dx = 1, \qquad EX^2 = \int_0^{\pi/2} x^2 \sin x\,dx = \pi - 2,$$

$$EX^3 = \int_0^{\pi/2} x^3 \sin x\,dx = \frac{3}{4}(\pi - 8), \qquad \operatorname{Var} X = EX^2 - (EX)^2 = \pi - 3,$$

$$E\,(X - EX)^3 = \int_0^{\pi/2} (x - EX)^3 \sin x\,dx = \frac{3\pi^2}{4} - 3\pi + 2.$$

2.6.5 Očekivanje i varijanca nekih parametarskih neprekidnih distribucija

Analogno parametarskim diskretnim distribucijama, za nastavak će kolegija biti korisno izračunati i zapamtiti izraze za očekivanje i varijancu nekih parametarskih neprekidnih distribucija koje se često koriste. U tu je svrhu samo potrebno riješiti zadane integrale. S obzirom da je normalna slučajna vrijabla najvažnija u klasi neprekidnih slučajnih varijabli, ovdje ćemo navesti samo izračun očekivanja i varijance te distribucije, a za ostale navodimo eksplicitne izraze u tablici 2.1.

Očekivanje i varijanca normalne distribucije

Neka je $X \sim \mathcal{N}(\mu, \sigma^2)$, tj.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Da bismo lakše riješili potrebne integrale, uočimo prvo da je funkcija

$$g(x) = x \, \frac{1}{\sqrt{2\pi}} \, e^{-\frac{x^2}{2}}$$

neparna, na $\mathbb{R}$ integrabilna funkcija, pa je

$$\int\limits_{-\infty}^{\infty} g(x)\, dx = 0.$$

Osim toga, zbog normiranosti funkcije gustoće neprekidne slučajne varijable znamo da je

$$\int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \, e^{-\frac{x^2}{2}} \, dx = 1.$$

Primjenom supstituciije $t = (x-\mu)/\sigma$ slijedi:

$$\begin{aligned} EX &= \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\infty} x \, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} (\sigma t + \mu) \, e^{-\frac{t^2}{2}} \, dt \\ &= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \sigma t \, e^{-\frac{t^2}{2}} \, dt + \mu \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{-\frac{t^2}{2}} \, dt \\ &= \mu, \end{aligned}$$

$$\begin{aligned} EX^2 &= \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\infty} x^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} (\sigma t + \mu)^2 e^{-\frac{t^2}{2}} \, dt \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}} \, dt + \frac{2\sigma\mu}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} t e^{-\frac{t^2}{2}} \, dt + \frac{\mu^2}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{-\frac{t^2}{2}} \, dt \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}} \, dt + \mu^2 = \sigma^2 \sqrt{\frac{2}{\pi}} \int\limits_{0}^{\infty} t^2 e^{-\frac{t^2}{2}} \, dt + \mu^2. \end{aligned}$$

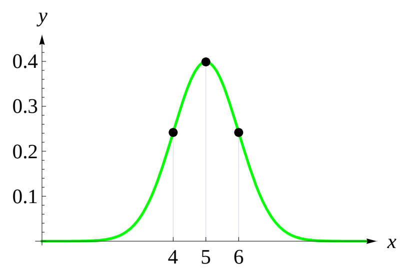
Gornji integral rješavamo parcijalnom integracijom. Slijedi da je

$$\sigma^2 \sqrt{\frac{2}{\pi}} \int\limits_{0}^{\infty} t^2 e^{-\frac{t^2}{2}} \, dt = \sigma^2 \sqrt{\frac{2}{\pi}} \int\limits_{0}^{\infty} e^{-\frac{t^2}{2}} \, dt = \sigma^2.$$

Dakle, $\operatorname{Var} X = \sigma^2$.

Uočimo da su parametri μ i σ^2 normalne distribucije upravo njezino matematičko očekivanje i varijanca, redom.

Primjer 2.38. *Graf funkcije gustoće normalne slučajne varijable s očekivanjem $\mu = 5$ i varijancom $\sigma^2 = 1$ prikazan je slikom 2.25.*



Slika 2.25: Graf funkcije gustoće normalne distribucije $\mathcal{N}(5, 1)$.

Vidimo da funkcija gustoće ima maksimum u očekivanju, a za vrijednosti nezavisne varijable $x_1 = \mu - \sigma$ i $x_2 = \mu + \sigma$ ima točke infleksije. Odredimo vjerojatnost realizacije te slučajne varijable unutar intervala $[\mu - \sigma, \mu + \sigma] = [4, 6]$, tj. vjerojatnost realizacije slučajne varijable udaljene od očekivanja za manje ili točno jednu standardnu devijaciju (koristite računalo!).

$$P\{X \in [4, 6]\} \approx 0.682.$$

Provjerite da je $P\{X \in [\mu - 2\sigma, \mu + 2\sigma]\} \approx 0.955$ i $P\{X \in [\mu - 3\sigma, \mu + 3\sigma]\} \approx 0.997$.

Kratak pregled obrađenih distribucija dan je u tablici 2.1.

Naziv distribucije	**Distribucija (gustoća)**	**Očekivanje**	**Varijanca**
Bernoullijeva, $p \in \langle 0, 1\rangle$	$\begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$	p	$p(1-p)$
Binomna $\mathcal{B}(n, p)$, $n \in \mathbb{N}$, $p \in \langle 0, 1\rangle$	$\mathcal{R}(X) = \{0, 1, 2, \ldots, n\}$ $p_i = \binom{n}{i} p^i (1-p)^{n-i}$	np	$np(1-p)$
Poissonova $\mathcal{P}(\lambda)$, $\lambda > 0$	$\mathcal{R}(X) = \mathbb{N}_0$ $p_i = \frac{\lambda^i}{i!} e^{-\lambda}$	λ	λ
Geometrijska $\mathcal{G}(p)$, $p \in \langle 0, 1\rangle$	$\mathcal{R}(X) = \mathbb{N}$ $p_i = p(1-p)^{i-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$

Naziv distribucije	**Distribucija (gustoća)**	**Očekivanje**	**Varijanca**
Uniformna na $\langle a, b\rangle$ $\mathcal{U}(a,b)$, $a < b$	$f(x) = \frac{1}{b-a} I_{(a,b)}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Eksponencijalna $\mathcal{E}(\lambda)$, $\lambda > 0$	$f(x) = \lambda e^{-\lambda x} I_{[0,\infty)}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Laplaceova, $\lambda > 0$	$f(x) = \frac{1}{2}\lambda e^{-\lambda\lvert x\rvert}$	0	$\frac{2}{\lambda^2}$
Normalna $\mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2

Tablica 2.1: Tablica distribucija ili funkcija gustoća te očekivanja i varijanci važnijih parametarskih familija diskretnih i neprekidnih distribucija.

2.7 Transformacija slučajne varijable

2.7.1 Postupak standardizacije

Korištenjem svojstava očekivanja i varijance dokazat ćemo da svaku slučajnu varijablu koja ima varijancu možemo afino transformirati (dodavanjem konstante i množenjem s konstantom različitom od 0) tako da novonastala slučajna varijabla ima očekivanje 0 i varijancu 1. Takav se postupak transformiranja slučajne varijable u statistici naziva **postupak standardizacije**.

Propozicija 2.3. *Neka je X slučajna varijabla s očekivanjem $\mu \in \mathbb{R}$ i varijancom $\sigma^2 > 0$. Tada slučajna varijabla*

$$Y = \frac{X - \mu}{\sigma}, \quad \sigma = \sqrt{\sigma^2},$$

ima očekivanje 0 i varijancu 1.

Dokaz. *Korištenjem rezultata poglavlja 2.6 vidimo da vrijedi:*

$$EY = \frac{1}{\sigma}(EX - \mu) = 0,$$

$$\operatorname{Var} Y = \frac{1}{\sigma^2}\operatorname{Var} X = 1.$$

Vidimo da smo postupkom standardizacije lako promijenili varijancu i očekivanje slučajne varijable, ali je ovdje ostalo nejasno što se pri toj transformaciji dogodilo s njezinom funkcijom distribucije, tj. na koji je način ona promijenjena.

Propozicija 2.4. *Neka je X slučajna varijabla s funkcijom distribucije $F_X(x)$, očekivanjem $\mu \in \mathbb{R}$ i varijancom $\sigma^2 > 0$. Tada slučajna varijabla*

$$Y = \frac{X - \mu}{\sigma}, \quad \sigma = \sqrt{\sigma^2},$$

ima funkciju distribucije $F_Y(x) = F_X(\sigma x + \mu)$.

Dokaz. *S obzirom da je $\sigma > 0$, vrijedi:*

$$F_Y(x) = P\{Y \leq x\} = P\left\{\frac{X - \mu}{\sigma} \leq x\right\} = P\left\{X \leq \sigma x + \mu\right\} = F_X(\sigma x + \mu).$$

Primjer 2.39. *Neka je $X \sim \mathcal{B}(n,p)$, $n \in \mathbb{N}$, $p \in \langle 0, 1\rangle$. Postupkom standardizacije dobivamo slučajnu varijablu*

$$Y = \frac{X - np}{\sqrt{np(1-p)}}.$$

Uočimo da Y više nema distribuciju u binomnoj familiji!

Primjer 2.40. *Neka je $X \sim \mathcal{P}(\lambda)$, $\lambda > 0$. Postupkom standardizacije dobivamo slučajnu varijablu*

$$Y = \frac{X - \lambda}{\sqrt{\lambda}}.$$

Uočimo da Y više nema distribuciju u Poissonovoj familiji!

Primjer 2.41. *Neka je $X \sim \mathcal{U}(a,b)$. Funkcija distribucije slučajne varijable X dana je izrazom*

$$F_X(x) = \int_{-\infty}^{x} f(x)\,dx = \begin{cases} 0 & , \ x < a \\ \dfrac{x-a}{b-a} & , \ a \leq x < b \\ 1 & , \ x \geq b \end{cases}.$$

Postupkom standardizacije dobivamo slučajnu varijablu Y sa sljedećom funkcijom distribucije

$$F_Y(x) = F_X(\sigma x + \mu) = \begin{cases} 0 & , \ \sigma x + \mu < a \\ \dfrac{\sigma x + \mu - a}{b-a} & , \ a \leq \sigma x + \mu < b \\ 1 & , \ \sigma x + \mu \geq b \end{cases},$$

tj.

$$F_Y(x) = \begin{cases} 0 & , \ x < \frac{a-\mu}{\sigma} \\ \dfrac{x - \frac{a-\mu}{\sigma}}{\frac{b-\mu}{\sigma} - \frac{a-\mu}{\sigma}} & , \ \frac{a-\mu}{\sigma} \leq x < \frac{b-\mu}{\sigma} \\ 1 & , \ x \geq \dfrac{b-\mu}{\sigma} \end{cases}.$$

Odavde vidimo da je $Y \sim \mathcal{U}\left(\dfrac{a-\mu}{\sigma}, \dfrac{b-\mu}{\sigma}\right)$, tj. postupak standardizacije zadržao je slučajnu varijablu u istoj klasi, ali s drugim parametrima.

Primjer 2.42. *Neka je $X \sim \mathcal{N}(\mu, \sigma^2)$. Postupkom standardizacije dobivamo slučajnu varijablu*

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0,1).$$

Provjerite ovu tvrdnju!

2.7.2 Bijektivna transformacija slučajne varijable

Postupkom standardizacije transformiramo slučajnu varijablu afinom funkcijom s pozitivnim koeficijentom smjera. Naime, u postupku standardizacije novonastala slučajna varijabla Y kompozicija je afine funkcije $g(x) = \dfrac{x-\mu}{\sigma}$ i slučajne varijable X, tj. $Y = g(X)$. Pokazali smo da se, kod ovako jednostavne funkcije g, može lako odrediti funkcija distribucije novonastale slučajne varijable. Sličan postupak može se primijeniti kod svih transformacija bijektivnim funkcijama g. Ako je slučajna varijabla X diskretna, opisanim postupkom dobivamo slučajnu varijablu $Y = g(X)$ koja je također diskretnog tipa. Međutim, ako je X apsolutno neprekidna, $Y = g(X)$ će biti apsolutno neprekidna samo ako je moguće izraziti njezinu funkciju distribucije pomoću funkcije gustoće.

Primjer 2.43. *Neka je dana diskretna slučajna varijabla X tablicom distribucije*

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}, \quad 0 \le p_i \le 1, \quad \sum_{i\in I} p_i = 1, \quad I \subseteq \mathbb{N},$$

i neka je $g : \mathcal{R}(X) \to \mathbb{R}$ bijekcija. Tada je slučajna varijabla $Y = g(X)$ zadana tablicom distribucije

$$Y = \begin{pmatrix} g(x_1) & g(x_2) & \dots & g(x_n) & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}, \quad p_i \ge 1, \quad \sum_{i\in I} p_i = 1 \quad I \subseteq \mathbb{N}.$$

Primjer 2.44. *Neka je dana neprekidna slučajna varijabla X funkcijom gustoće f i neka je bijekcija $g\colon \mathbb{R} \to \mathbb{R}$ definirana pravilom $g(x) = e^x$. Tada je $Y = g(X) = e^X$ neprekidna slučajna varijabla s funkcijom gustoće $h(x) = f(\ln x)\,\dfrac{1}{x}$ za $x > 0$ i $h(x) = 0$ za $x \le 0$.*
Zaista,

$$F_Y(x) = P\{Y \le x\} = P\{g(X) \le x\} = \begin{cases} 0 & , \ x \le 0 \\ P\{X \le g^{-1}(x)\} & , \ x > 0 \end{cases}.$$

Nadalje, za $x > 0$ vrijedi:

$$P\{X \le g^{-1}(x)\} = \int_{-\infty}^{g^{-1}(x)} f(t)\,dt = \int_{-\infty}^{\ln x} f(t)\,dt.$$

Uz supstituciju $t = \ln u$ slijedi:

$$P\{X \le g^{-1}(x)\} = \int_0^x f(\ln u)\frac{1}{u}\,du.$$

Definiranjem funkcije

$$h(u) = \begin{cases} 0 & , \ u \le 0 \\ f(\ln u)\dfrac{1}{u} & , \ u > 0 \end{cases}$$

vidimo da je ona nenegativna i vrijedi

$$P\{Y \le x\} = \int_{-\infty}^{x} h(u)\,du,$$

pa je time pokazano da je $h(u)$ funkcija gustoće slučajne varijable Y.

Primjer 2.45. *Neka je neprekidna slučajna varijabla X zadana funkcijom gustoće f i neka je $g(x) = e^{-x}$. Tada se analognim postupkom kao u primjeru 2.44 vidi da je $Y = g(X) = e^{-X}$ neprekidna slučajna varijabla s funkcijom gustoće $h(x) = f(-\ln(x))\frac{1}{x}$ za $x > 0$ i $h(x) = 0$ za $x \leq 0$.*

Za neprekidnu slučajnu varijablu X i bijekciju $g\colon \mathbb{R} \to \mathcal{R}(g) \subseteq \mathbb{R}$ može se odrediti općeniti izraz za funkciju gustoće slučajne varijable $Y = g(X)$ ako je ona neprekidna slučajna varijabla. O tome govori teorem 2.5.

Teorem 2.5. *Neka je X neprekidna slučajna varijabla, f_X njezina funkcija gustoće, a F_X funkcija distribucije. Neka je, nadalje, $g\colon \mathbb{R} \to \mathcal{R}(g) \subseteq \mathbb{R}$ bijekcija. Ako je funkcija g derivabilna na $\mathbb{R}$, onda je $Y = g(X)$ neprekidna slučajna varijabla s funkcijom gustoće*

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))|[g^{-1}(y)]'| & , y \in \mathcal{R}(g) \\ 0 & , y \in (\mathcal{R}(g))^c \end{cases}.$$

Dokaz. Neka je X neprekidna slučajna varijabla i g bijekcija kao u iskazu teorema. Tražimo funkciju distribucije F_Y slučajne varijable $Y = g(X)$. Uočimo da vrijedi:

$$F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\}.$$

Da bismo izrazili $P\{g(X) \leq y\}$ u terminima $P\{X \in A\}$ za neki skup $A \subseteq \mathbb{R}$ (tj. u terminima funkcije distribucije slučajne varijable X), potrebno je riješiti nejednadžbu

$$g(x) \leq y.$$

S obzirom da je g bijekcija s $\mathbb{R}$ na $\mathcal{R}(g) \subseteq \mathbb{R}$, postupak rješavanja ovisi samo o tome da li je g monotono rastuća ili monotono padajuća funkcija.

- Neka je $g : \mathbb{R} \to \mathcal{R}(g)$ monotono rastuća funkcija.

Tada za sve $y \in \mathcal{R}(g)$ vrijedi:

$$F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\} = P\{X \leq g^{-1}(y)\} = F_X(g^{-1}(y)).$$

Ako je funkcija g derivabilna na $\mathbb{R}$, onda je i funkcija $F_X(g^{-1}(y))$ derivabilna na $\mathcal{R}(g)$, pa možemo definirati funkciju

$$f_Y(y) = \frac{dF_X(g^{-1}(y))}{dy} = f_X(g^{-1}(y))[g^{-1}(y)]', \quad y \in \mathcal{R}(g).$$

Uočimo da je f_Y nenegativna funkcija s obzirom da su F_X i g monotono rastuće funkcije. Također vrijedi:

$$\int_{\mathcal{R}(g)} f_X(g^{-1}(y))[g^{-1}(y)]'\,dy = \int_{\mathbb{R}} f_X(x)\,dx = 1.$$

Sada možemo proširiti funkciju $f_Y(y)$ na cijeli $\mathbb{R}$ tako da definiramo:

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))[g^{-1}(y)]' & , y \in \mathcal{R}(g) \\ 0 & , y \in (\mathcal{R}(g))^c \end{cases}.$$

- Neka je $g\colon \mathbb{R} \to \mathcal{R}(g)$ monotono padajuća funkcija.

Tada za sve $y \in \mathcal{R}(g)$ vrijedi:

$$F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\} = P\{X \geq g^{-1}(y)\} =$$

$$= 1 - P\{X < g^{-1}(y)\} = 1 - F_X(g^{-1}(y)).$$

Ako je funkcija g derivabilna na $\mathbb{R}$ onda je i funkcija $(1 - F_X(g^{-1}(y)))$ derivabilna na $\mathcal{R}(g)$, pa možemo definirati funkciju

$$f_Y(y) = \frac{d(1 - F_X(g^{-1}(y)))}{dy} = -f_X(g^{-1}(y))[g^{-1}(y)]', \quad y \in \mathcal{R}(g).$$

Uočimo da je f_Y nenegativna funkcija jer je $f_X(g^{-1}(y)) \geq 0$ zbog nenegativnosti funkcije f_X, a $[g^{-1}(y)]' < 0$ s obzirom da je g monotono padajuća funkcija. Također vrijedi:

$$\int_{\mathcal{R}(g)} \left(-f_X(g^{-1}(y))[g^{-1}(y)]'\right) dy = \int_{\mathbb{R}} f_X(x)\, dx = 1.$$

Sada možemo proširiti funkciju $f_Y(y)$ na cijeli $\mathbb{R}$ tako da definiramo:

$$f_Y(y) = \begin{cases} -f_X(g^{-1}(y))[g^{-1}(y)]' & , y \in \mathcal{R}(g) \\ 0 & , y \in (\mathcal{R}(g))^c \end{cases}.$$

Prema prethodnim rezultatima slijedi da za derivabilnu bijekciju $g\colon \mathbb{R} \to \mathcal{R}(g)$ možemo definirati funkciju $f_Y : \mathbb{R} \to \mathbb{R}$ izrazom:

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left|[g^{-1}(y)]'\right| & , \quad y \in \mathcal{R}(g) \\ 0 & , y \in (\mathcal{R}(g))^c \end{cases}.$$

Ta je funkcija nenagativna i vrijedi:

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)\, dt.$$

Dakle, Y je neprekidna slučajna varijabla s funkcijom gustoće f_Y.

2.7.3 Primjeri transformacija koje nisu bijektivne

Ako transformacija slučajne varijable nije bijekcija, također je moguće odrediti distribuciju novonastale slučajne varijable, ali je postupak tehnički zahtjevniji. Za ilustraciju ćemo se poslužiti primjerima.

Primjer 2.46. *Pretpostavimo da je diskretna slučajna varijabla X dana tablicom distribucije*

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}, \quad p_i \geq 1, \quad \sum_i p_i = 1.$$

Neka je $g : \mathcal{R}(X) \to \mathbb{R}$ proizvoljna funkcija te neka je slučajna varijabla Y definirana kao $Y = g(X)$. Označimo $\mathcal{R}(Y) = \{y_1, y_2, \dots, y_k, \dots\}$. Tada slučajna varijabla Y ima tablicu distribucije

$$Y = \begin{pmatrix} y_1 & y_2 & \dots & y_k & \dots \\ q_1 & q_2 & \dots & q_k & \dots \end{pmatrix}, \quad q_k \geq 1, \quad \sum_k q_k = 1,$$

gdje je

$$q_k = \sum_{\{\omega \in \Omega | g(X(\omega)) = y_k\}} P\{\omega\} = \sum_{\{i | g(x_i) = y_k\}} p_i.$$

Primjer 2.47. *Pretpostavimo da je slučajna varijabla X zadana tablicom distribucije*

$$X = \begin{pmatrix} -a & 0 & a \\ 0.4 & 0.2 & 0.4 \end{pmatrix}.$$

Tada slučajna varijabla $Y = X^2$, koja je dobivena kompozicijom slučajne varijable X i nebijektivne funkcije $g(t) = t^2$, $g \colon \mathbb{R} \to \mathbb{R}$, ima distribuciju zadanu tablicom

$$X = \begin{pmatrix} 0 & a^2 \\ 0.2 & 0.8 \end{pmatrix}.$$

Primjer 2.48. *Neka je X neprekidna slučajna varijabla zadana funkcijom gustoće f. Tada je $Y = X^2$ neprekidna slučajna varijabla s funkcijom gustoće*

$$h(x) = \begin{cases} 0 & , \ x < 0 \\ \dfrac{1}{2\sqrt{x}} \left(f(\sqrt{x}) + f(-\sqrt{x}) \right) & , \ x \geq 0 \end{cases}.$$

Zaista, za $x < 0$ je $P\{Y \leq x\}$ očigledno jednako 0. Za $x \geq 0$ je

$$P\{Y \leq x\} = P\{X^2 \leq x\} = P\{-\sqrt{x} \leq X \leq \sqrt{x}\} = \int_{-\sqrt{x}}^{\sqrt{x}} f(t)\,dt = \int_0^{\sqrt{x}} (f(-t) + f(t))\,dt.$$

Supstitucijom $t = \sqrt{u}$ u tom integralu vidimo da za $x \geq 0$ vrijedi:

$$P\{Y \leq x\} = \int_0^x \frac{1}{2\sqrt{u}} \left(f(\sqrt{u}) + f(-\sqrt{u}) \right) du = \int_0^x h(u)\,du.$$

Dakle, h je funkcija gustoće neprekidne slučajne varijable $Y = X^2$.

Odredite funkciju gustoće neprekidne slučajne varijable $Y = X^2$ ako je X standardna normalna slučajna varijabla, tj. $X \sim \mathcal{N}(0, 1)$.

2.8 Generiranje slučajnih varijabli

Važan je način ispitivanja istinitosti tvrdnji koje se odnose na slučajne varijable provjera tih tvrdnji na podacima. Međutim, podaci koji se u tu svrhu trebaju koristiti realizacije su slučajne varijable. Pitanje je kako prikupiti podatke iz slučajne varijable zadane distribucije. Radi ilustracije problema zamislimo da želimo dobiti 30 podataka koji su realizacije Bernoullijeve slučajne varijable s tablicom distribucije

$$X = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Za to postoji nekoliko jednostavnih načina. Primjerice, bacimo pravilno izrađen novčić 30 puta i pri tome bilježimo 0 ako je palo pismo, a 1 ako se okrenula glava. Ili uzmemo kutijicu s istim brojem bijelih i crnih kuglica pa iz nje na slučajan način izvlačimo jednu kuglicu 30 puta, ali tako da svaki puta izvučenu kuglicu vratimo u kutijicu i promiješamo. Pritom bilježimo 0 ako je izvučena bijela kuglica, a 1 ako je izvučena crna kuglica.

Međutim, kako ćemo dobiti podatke iz, primjerice, normalne distribucije sa zadanim očekivanjem i varijancom?

Posebno koristan rezultat koji se koristi u tu svrhu odnosi se na distribuciju slučajne varijable F, gdje je F funkcija distribucije slučajne varijable X.

Naime, pretpostavimo da je X neprekidna slučajna varijabla s funkcijom distribucije F. S obzirom da je X neprekidna slučajna varijabla, njezina je funkcija distribucije na nosaču funkcije gustoće (tj. na skupu $\{x \in \mathbb{R} : f(x) > 0\}$) neprekidna rastuća bijekcija. Za funkciju distribucije F_U slučajne varijable $U = F(X)$ vrijedi:

$$F_U(u) = P\{U \leq u\} = P\{F(X) \leq u\} = \begin{cases} 0 & , u < 0 \\ P\{X \leq F^{-1}(u)\} & , u \in [0,1) \\ 1 & , , u \geq 1 \end{cases}.$$

S obzirom da je $P\{X \leq F^{-1}(u)\} = F(F^{-1}(u)) = u$, vidimo da je $U \sim \mathcal{U}(0,1)$, i to neovisno o funkciji distribucije F sve dok je ona funkcija distribucije neprekidne slučajne varijable.

Pogledajmo i drugi smisao toga rezultata. Pođimo od slučajne varijable $U \sim \mathcal{U}(0,1)$. Odaberimo funkciju distribucije F neprekidne slučajne varijable X koju želimo modelirati. Očigledno je da se distribucija slučajne varijable X može dobiti kao distribucija kompozicije $F^{-1}(U)$. Naime, vrijedi:

$$P\{F^{-1}(U) \leq x\} = P\{U \leq F(x)\} = F(x).$$

Dakle, ako trebamo generirati n podataka iz neprekidne slučajne varijable s distribucijom F, dovoljno je da imamo niz podataka $(u_i, i = 1, \ldots, n)$ iz $\mathcal{U}(0,1)$. Tražene podatke tada dobijemo kao $(F^{-1}(u_i), i = 1, \ldots, n)$.

Za generiranje se realizacija uniformne distribucije na $\langle 0, 1\rangle$ danas koriste takozvani generatori slučajnih brojeva koji su ugrađeni u sve naprednije računalne programe.

Uniformna distribucija na intervalu $\langle 0, 1\rangle$ može se također iskoristiti i za generiranje realizacija diskretnih distribucija. Zainteresirani čitatelj može više informacija o tome pronaći u [8].

2.9 Zadaci

Zadatak 2.1. Promotrimo slučajan pokus koji se sastoji od bacanja pravilno izrađenog novčića dva puta zaredom. Neka je X slučajna varijabla čija je vrijednost broj realiziranih pisama. Odredite tablicu distribucije slučajne varijable X.

Rješenje:

$$X = \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}$$

Zadatak 2.2. Strijelac na raspolaganju ima tri metka i gađa metu dok je ne pogodi ili dok ne potroši sva tri metka. Neka je X slučajna varijabla čija je vrijednost broj potrošenih metaka. Uz pretpostavku o nezavisnosti gađanja tablicu distribucije od X ako je poznato da je vjerojatnost pogotka mete pri svakom gađanju jednaka 0.8.

Rješenje:

$$X = \begin{pmatrix} 1 & 2 & 3 \\ 0.8 & 0.16 & 0.04 \end{pmatrix}$$

Zadatak 2.3. Promotrimo slučajan pokus koji se sastoji od bacanja pravilno izrađene igraće kockice dva puta zaredom. Neka je X slučajna varijabla čija je vrijednost zbroj realiziranih brojeva u oba bacanja. Odredite tablicu distribucije slučajne varijable X.

Rješenje:

$$X = \begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{1}{18} & \frac{1}{12} & \frac{1}{9} & \frac{5}{36} & \frac{1}{6} & \frac{5}{36} & \frac{1}{9} & \frac{1}{12} & \frac{1}{18} & \frac{1}{36} \end{pmatrix}$$

Zadatak 2.4. Iz kutije u kojoj se nalazi 7 kuglica numeriranih brojevima od 1 do 7 izvlače se istovremeno tri kuglice $\{i,j,k\}$. Odredite tablicu distribucije i funkciju distribucije slučajne varijable X definirane na sljedeći način:

$$X(\{i,j,k\}) = \max(i,j,k), \;\; i,j,k \in \{1,\ldots,7\}.$$

Rješenje:
Tablica distribucije:

$$X = \begin{pmatrix} 3 & 4 & 5 & 6 & 7 \\ \frac{1}{35} & \frac{3}{35} & \frac{6}{35} & \frac{10}{35} & \frac{15}{35} \end{pmatrix}$$

Funkcija distribucije:

$$F(x) = \begin{cases} 0 & , \; x \in (-\infty, 3) \\ 1/35 & , \; x \in [3,4) \\ 4/35 & , \; x \in [4,5) \\ 10/35 & , \; x \in [5,6) \\ 20/35 & , \; x \in [6,7) \\ 1 & , \; x \in [7,\infty) \end{cases}$$

Zadatak 2.5. Iz kutije u kojoj se nalazi 5 kuglica numeriranih brojevima od 1 do 5 izvlače se istovremeno tri kuglice (koje su numerirane brojevima i, j i k gdje su $i,j,k \in \{1,\ldots,5\}$). Odredite tablicu distribucije i funkciju distribucije slučajne varijable X definirane na sljedeći način:

$$X(\{i,j,k\}) = \min\{i,j,k\}, \;\; i,j,k \in \{1,\ldots,5\}.$$

Zadatak 2.6. U smjeru kretanja automobila nalaze se redom tri semafora koji rade nezavisno jedan od drugog. Na svakomu se s vjerojatnošću $p = 0.5$ pojavljuje crveno i s vjerojatnošću $q = 0.5$ zeleno svjetlo. Slučajna varijabla X predstavlja broj semafora pored kojih prolazi automobil do prvog zaustavljanja. Odredite tablicu distribucije i funkciju distribucije slučajne varijable X te nacrtajte graf funkcije distribucije.

Rješenje:
Tablica distribucije:

$$X = \begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \end{pmatrix}$$

Funkcija distribucije:

$$F(x) = \begin{cases} 0 & , \; x \in (-\infty, 0) \\ 1/2 & , \; x \in [0,1) \\ 3/4 & , \; x \in [1,2) \\ 7/8 & , \; x \in [2,3) \\ 1 & , \; x \in [3,\infty) \end{cases}$$

Zadatak 2.7. Odredite matematičko očekivanje i varijancu slučajne varijable zadane sljedećom tablicom distribucije

$$X = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1/2 & 1/4 & 1/8 & 1/16 & 1/16 \end{pmatrix}.$$

Rješenje:

$$E(X) = \frac{15}{16}, \quad \operatorname{Var} X = \frac{367}{256}.$$

Zadatak 2.8. Slučajna varijabla X zadana je tablicom distribucije

$$X = \begin{pmatrix} -1 & 0 & 1 & 3 & 4 & 8 \\ a & 1/8 & a - b^2 & b^2 & 1/4 & b \end{pmatrix},$$

gdje su a i b nepoznati parametri. Ako je očekivanje od X 25/8, odredite:

a) vrijednosti parametara a i b,

b) varijancu slučajne varijable X,

c) funkciju distribucije slučajne varijable X.

Rješenje: a) $a = \frac{3}{16}$, $b = \frac{1}{4}$ b) $\operatorname{Var} X = \frac{719}{64}$

c) Funkcija distribucije:

$$F(x) = \begin{cases} 0 & , x \in (-\infty, -1) \\ 3/16 & , x \in [-1, 0) \\ 5/16 & , x \in [0, 1) \\ 7/16 & , x \in [1, 3) \\ 1/2 & , x \in [3, 4) \\ 3/4 & , x \in [4, 8) \\ 1 & , x \in [8, \infty) \end{cases}$$

Zadatak 2.9. Ispitujemo ispravnost sustava koji se sastoji od tri komponente. Otkazivanje komponenti događa se nezavisno i vjerojatnost je otkazivanja n-te komponente

$$p_n = 0.2 + (n-1)\,0.1, \quad n \in \{1, 2, 3\}.$$

Odredite distribuciju, matematičko očekivanje i varijancu slučajne varijable kojom je modeliran broj komponenti koje su otkazale.

Rješenje:

$$X = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0.336 & 0.452 & 0.188 & 0.024 \end{pmatrix}, \quad EX = 0.9, \quad \mathrm{Var\,X} = 0.61.$$

Zadatak 2.10. U spremniku se nalazi pet kutija: dvije su prazne, a u preostalim trima nalazi se po deset bombona. Izvlačimo jednu po jednu kutiju sve dok ne izvučemo praznu kutiju (izvlačenje se vrši bez vraćanja prethodno izvučenih kutija u spremnik). Kolika je vjerojatnost da izvučemo svih 30 bombona? Odredite distribuciju, matematičko očekivanje i varijancu slučajne varijable kojom je modeliran broj bombona koje smo izvukli.

Rješenje:

$$Y = \begin{pmatrix} 0 & 10 & 20 & 30 \\ 2/5 & 3/10 & 1/5 & 1/10 \end{pmatrix}, \quad EY = 10, \quad \mathrm{Var\,Y} = 100.$$

Zadatak 2.11. Diskretna slučajna varijabla zadana je tablicom distribucije

$$X = \begin{pmatrix} 1-\alpha & 1 & 1+\alpha \\ p & 2p & 3p \end{pmatrix}, \quad \alpha > 0.$$

a) Odredite vrijednost parametra p.

b) Izračunajte matematičko očekivanje, drugi moment i treći apsolutni moment slučajne varijable X.

c) Izračunajte

$$P\left\{|X| \leq 1 + \frac{\alpha}{2}\right\}.$$

Rješenje:

a) $p = 1/6$,

b) $EX = \frac{\alpha+3}{3}$, $EX^2 = 1 + \frac{2}{3}\alpha(1+\alpha)$, $E|X|^3 = \begin{cases} (\alpha^3 + 6\alpha^2 + 3\alpha + 3)/3 & , \ 0 < \alpha \leq 1 \\ (2\alpha^3 + 3\alpha^2 + 3\alpha + 2)/3 & , \ \alpha > 1 \end{cases}$,

c) $P\{|X| \leq 1 + \alpha/2\} = 1/3$.

Zadatak 2.12. Diskretna slučajna varijabla zadana je tablicom distribucije

$$X = \begin{pmatrix} 1 & 3 & 5 & 7 & 9 & 11 \\ 3p & 1/8 & 1/8 & 2p & 1/8 & p \end{pmatrix}.$$

Izračunajte:

a) vrijednost parametra p,

b) matematičko očekivanje i varijancu slučajne varijable $Y = \min\{8, X\}$.

Rješenje:

a) $p = 5/48$,

b)

$$Y = \begin{pmatrix} 1 & 3 & 5 & 7 & 8 \\ 5/16 & 1/8 & 1/8 & 5/24 & 11/48 \end{pmatrix}, \quad EY = 221/48, \quad \operatorname{Var} Y = 18983/2304.$$

Zadatak 2.13. Diskretna slučajna varijabla zadana je tablicom distribucije

$$X = \begin{pmatrix} 2 & 4 & 6 & 8 & 10 & 12 \\ 2p & 1/6 & 1/4 & 3p & 1/8 & p \end{pmatrix}.$$

Izračunajte:

a) vrijednost parametra p,

b) matematičko očekivanje i varijancu slučajne varijable $Y = \max\{5, X\} - 1$.

Zadatak 2.14. Diskretna slučajna varijabla zadana je tablicom distribucije

$$X = \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 0.1 & 0.1 & 0.2 & 0.3 & 0.3 \end{pmatrix}.$$

Odredite distribuciju, funkciju distribucije, očekivanje i varijancu slučajne varijable $Y = 3X^2 - 3$.

Zadatak 2.15. Diskretna slučajna varijabla zadana je tablicom distribucije

$$X = \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 0.1 & p & 0.3 & 3p & 0.2 \end{pmatrix}.$$

Odredite vrijednost parametra p, distribuciju, funkciju distribucije, očekivanje i varijancu slučajne varijable $Y = \left|X^3 - 1\right|$.

Zadatak 2.16. Diskretna slučajna varijabla zadana je tablicom distribucije:

$$X = \begin{pmatrix} \frac{\pi}{6} & \pi/3 & \pi/2 & 2\pi/3 & 5\pi/6 \\ 0.25 & 0.2 & 0.3 & 0.15 & 0.1 \end{pmatrix}.$$

Odredite distribuciju, funkciju distribucije, očekivanje i varijancu slučajne varijable $Y = 3 - 2\cos^2 X$.

Zadatak 2.17. Diskretna slučajna varijabla zadana je tablicom distribucije

$$X = \begin{pmatrix} 0 & \pi/4 & \pi/2 & 3\pi/4 & \pi \\ p & 0.2 & 0.1 & 0.3 & 0.2 \end{pmatrix}.$$

Odredite vrijednost parametra p, distribuciju, funkciju distribucije, očekivanje i varijancu slučajne varijable $Y = \sin X$ te izračunajte $P\{0 < Y \leq \sqrt{2}/2\}$.

Zadatak 2.18. Diskretna slučajna varijabla zadana je tablicom distribucije

$$X = \begin{pmatrix} -\pi/2 & -\pi/3 & 0 & \pi/3 & \pi/2 \\ 0.25 & 0.25 & p & 0.15 & 0.15 \end{pmatrix}.$$

Odredite vrijednost parametra p, distribuciju, funkciju distribucije, očekivanje i varijancu slučajne varijable $Y = \cos X$.

Zadatak 2.19. Diskretna slučajna varijabla zadana je tablicom distribucije

$$X = \begin{pmatrix} -\pi/2 & 0 & \pi/2 & \pi \\ p & p/2 & 2p & p/2 \end{pmatrix}.$$

Odredite vrijednost parametra p, distribuciju, funkciju distribucije, očekivanje i varijancu slučajne varijable $Y = e^{|\sin X|}$.

Zadatak 2.20. Diskretna slučajna varijabla zadana je tablicom distribucije

$$X = \begin{pmatrix} e^{-2} & e^{-1} & 1 & e & e^2 \\ 0.2 & 0.1 & p & 0.2 & 0.3 \end{pmatrix}.$$

Odredite vrijednost parametra p, distribuciju, funkciju distribucije, očekivanje i varijancu slučajne varijable $Y = (\ln X)^2$ te izračunajte $P\{Y \leq 1\}$.

Zadatak 2.21. Diskretna slučajna varijabla zadana je tablicom distribucije

$$X = \begin{pmatrix} e^{-3} & e^{-1} & 1 & e^3 & e^5 \\ p & 2p & 3p & 2p & 2p \end{pmatrix}.$$

Odredite vrijednost parametra p, distribuciju, funkciju distribucije, očekivanje i varijancu slučajne varijable $Y = |\ln X|$.

Zadatak 2.22. Diskretna slučajna varijabla zadana je tablicom distribucije

$$X = \begin{pmatrix} 3^{-3} & 3^{-1} & 1 & 3 & 3^3 \\ 0.3 & p & 2p & 0.2 & 0.2 \end{pmatrix}.$$

Odredite vrijednost parametra p, distribuciju, funkciju distribucije, očekivanje i varijancu slučajne varijable $Y = (\log_3 X)^2$.

Zadatak 2.23. Diskretna slučajna varijabla zadana je tablicom distribucije

$$X = \begin{pmatrix} 0 & \ln 2 & 2\ln 2 & 3\ln 2 & 4\ln 2 \\ 0.2 & p & 0.15 & 0.25 & 0.15 \end{pmatrix}.$$

Odredite vrijednost parametra p, distribuciju, funkciju distribucije, očekivanje i varijancu slučajne varijable $Y = e^{-X}$.

Zadatak 2.24. Slučajna varijabla X zadana je distribucijom

$$P\{X = k\} = \frac{1}{2^k}, \quad k \in \mathbb{N}.$$

Odredite distribuciju slučajne varijable $Y = \cos(\pi X)$.

Rješenje: $\mathcal{R}(Y) = \{-1, 1\}$, $P\{Y = -1\} = 2/3$, $P\{Y = 1\} = 1/3$.

Zadatak 2.25. Poznato je da je u velikom skladištu trgovine informatičkom opremom vjerojatnost pojavljivanja prijenosnog računala s greškom nastalom u proizvodnji jednaka 0.02. Pretpostavimo da iz tog skladišta biramo 10 prijenosnih računala. Odredite sljedeće vjerojatnosti:

a) vjerojatnost da je točno 5 prijenosnih računala s greškom,

b) vjerojatnost da su s greškom najviše 3 prijenosna računala;

c) vjerojatnost da je s greškom barem 6 prijenosnih računala.

Rješenje: a) $P\{X = 5\} = 0.000000728$.

Zadatak 2.26. U Bernoullijevoj shemi (n nezavisnih ponavljanja Bernoullijevog pokusa) zadana vjerojatnost uspjeha iznosi p, $p \in \langle 0, 1\rangle$. Odredite najmanji potreban broj pokusa koje treba provesti da bismo s vjerojatnošću r imali barem jedan uspjeh.

Rješenje: $n = \dfrac{\ln(1-r)}{\ln(1-p)}$.

Zadatak 2.27. Poznata osječka pizzeria primi u prosjeku 240 narudžbi tijekom jednog sata. Odredite vjerojatnost da u vremenskom periodu od jedne minute:

a) nije primljena niti jedna narudžba;

b) primljene su barem dvije narudžbe.

Rješenje: a) $P\{X = 0\} = e^{-4}$, b) $P\{X \geq 2\} = 1 - 5e^{-4}$.

Zadatak 2.28. Neka telefonska centrala primi u prosjeku 120 poziva tijekom jednog sata. Zbog iznenadnog kvara na centrali tijekom jedne minute pozivi nisu primani. Kolika je vjerojatnost da u tom periodu centrali nije bilo upućeno više od 4 poziva?

Rješenje: $P\{X < 4\} = \dfrac{19e^{-2}}{3}$.

Zadatak 2.29. Proizvodi u tvornici, među kojima se neispravan proizvod nalazi s vjerojatnošću 0.007, pakuju se u kutije po 100 komada. Kolika je vjerojatnost da kutija ne sadrži niti jedan pokvaren proizvod, a kolika da sadrži dva ili više pokvarenih proizvoda?

Rješenje: $P\{X = 0\} = 0.4954$, $P\{X \geq 2\} = 0.1554$.

Zadatak 2.30. Slučajnom pokusu bacanja nepravilno izrađenog novčića pripada dvočlani prostor elementarnih događaja $\{P, G\}$. Vjerojatnost realizacije pisma u jednom provođenju tog pokusa iznosi $p = 0.3$. Koliko je puta potrebno ponoviti taj slučajni pokus da bismo s vjerojatnošću 0.9 mogli tvrditi da se pismo realiziralo barem jednom?

Zadatak 2.31. Slučajan pokus sastoji se od bacanja pravilno izrađene igraće kockice četiri puta. Neka je broj pojavljivanja broja 5 vrijednost slučajne varijable X. Odredite zakon razdiobe slučajne varijable X, njezino očekivanje, varijancu te vjerojatnost da je broj 5 pao barem jednom.

Zadatak 2.32. Promotrimo sljedeću igru na sreću:

- u svakoj partiji igre jednom se bacaju četiri pravilno izrađene igraće kockice,
- igrač u svakoj partiji igre ulaže jednu kunu ("kladi" se na jednu kunu),
- u svakoj partiji igrač ili gubi ili nešto zarađuje. Ako se niti na jednoj kockici nije okrenula šestica, gubi uloženu kunu. Zarađuje prema sljedećoj shemi: dobiva jednu kunu za svaku šesticu koja se okrenula na bačenim kockicama i u tom slučaju zadržava i svoju uloženu kunu.

Izračunajte očekivani iznos igračevog dobitka (gubitka).

Rješenje: očekivanje je slučajne varijable kojom modeliramo igračev dobitak/gubitak -0.08 (ako igrač u svakoj partiji igre ulaže 1 kn za svakih sto odigranih partija, očekuje se gubitak od 8 kn).

Zadatak 2.33. Telefonist zaposlen u službi za korisnike nekog poduzeća zarađuje 10 kuna po primljenom pozivu klijenta. Poznato je da telefonist u tom poduzeću u danu primi u prosjeku 20 takvih poziva.

a) Odredite distribuciju slučajne varijable kojom modeliramo dnevnu zaradu telefonosta u tom poduzeću i odredite njegovu očekivanu dnevnu zaradu.

b) Izračunajte vjerojatnost da telefonist u jednom danu zaradi barem 180 kuna.

Rješenje:

a) $X \sim \mathcal{P}(20)$ - slučajna varijabla kojom se modelira broj poziva koje telefonist primi u jednom danu; $Y = 10\,X$ - slučajna varijabla kojom se modelira dnevna zarada telefonista; $EY = 200$.

b) $P\{Y \geq 180\} = 1 - \sum\limits_{k=0}^{17} \frac{20^k}{k!}\, e^{-20}$.

Zadatak 2.34. U pošiljci audio opreme nalazi se 100 mikrofona od kojih je 20 neispravno. Vlasnik trgovine audio opremom, koji ne zna broj neispravnih mikrofona u pošiljci, odlučuje da će za svoju trgovinu uzeti pošiljku ako među 10 slučajno odabranih mikrofona ne bude više od 3 neispravna.

a) Kolika je vjerojatnost da će trgovac prihvatiti pošiljku?

b) Što zaključujete o vjerojatnosti prihvaćanja pošiljke u slučaju da pošiljka sadrži jednak broj ispravnih i neispravnih mikrofona?

Rješenje:

a) $P\{X \leq 3\} = 0.89$,

b) pošiljka će s većim brojem neispravnih mikrofona biti prihvaćena s manjom vjerojatnošću.

Zadatak 2.35. Posuda sadrži 1000 lukovica tulipana, od kojih je 400 lukovica crvenih tulipana i 600 lukovica tulipana drugih boja. Iz te posude na slučajan način odaberemo 10 lukovica.

a) Kolika je vjerojatnost da smo odabrali točno pet lukovica crvenih tulipana?

b) Koliki je očekivani broj izvučenih lukovica crvenih tulipana?

Zadatke riješite pomoću hipergeometrijske i binomne distribucije.

Rješenje: primjenom hipergeometrijske distribucije dobivamo

a) $P\{X = 5\} = 0.201$,

b) $EX = 4$.

Zadatak 2.36. Student rješava pismeni ispit koji se sastoji od trideset pitanja od kojih svako nosi 5 bodova. Na svako pitanje ponuđena su tri odgovora od kojih je samo jedan točan. Pretpostavimo da odgovore na sva pitanja student odabire na slučajan način.

a) Odredite distribuciju slučajne varijable kojom modeliramo broj bodova koje student ostvaruje na tom ispitu te odredite očekivani broj ostvarenih bodova.

b) Izračunajte vjerojatnost da takvim rješavanjem ispita student ostvari barem 80 bodova.

Rješenje:

a) $X \sim \mathcal{B}(30, 1/3)$ - slučajna varijabla kojom se modelira broj točnih odgovora na ispitu; $Y = 5\,X$ - slučajna varijabla kojom se modelira broj bodova na ispitu; $E\,Y = 50$.

b) $P\{Y \geq 80\} = 1 - \sum\limits_{k=1}^{15} \binom{30}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{30-k}$.

Zadatak 2.37. U bubnju se nalazi sedam omotnica: tri su prazne, a u preostalim četirima nalazi se po 100 kuna. Igrač izvlači jednu po jednu omotnicu sve dok ne izvuče praznu (izvlačenje se vrši bez vraćanja prethodno izvučenih omotnica u bubanj).

a) Odredite distribuciju slučajne varijable X kojom je modeliran broj izvučenih omotnica.

b) Odredite transformaciju slučajne varijable X kojom je modeliran osvojeni iznos kuna te izračunajte njezino matematičko očekivanje i varijancu. Interpretirajte dobivene rezultate.

c) Kolika je vjerojatnost da igrač osvoji barem 200 kuna?

Rješenje:

a) $X = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3/7 & 2/7 & 6/35 & 3/35 & 1/35 \end{pmatrix}$.

b) $Y = 100(X-1)$ - slučajna varijabla kojom je modeliran osvojeni iznos kuna; $EY = 100$; $\operatorname{Var} Y = 100^2(51/35)$.

c) $P\{Y \geq 200\} = P\{X \geq 3\} = 2/7$.

Zadatak 2.38. U šeširu se nalazi osam kutijica: četiri su prazne, a u preostalim četirima nalaze se ključevi četiriju sefova od kojih svaki sadrži po jedan dijamant. Igrač izvlači jednu po jednu kutijicu sve dok ne izvuče praznu (izvlačenje se vrši bez vraćanja prethodno izvučenih kutijica u šešir).

a) Odredite distribuciju slučajne varijable X kojom je modeliran broj izvučenih kutijica.

b) Odredite transformaciju slučajne varijable X kojom je modeliran broj osvojenih dijamanata te izračunajte njezino matematičko očekivanje i varijancu. Interpretirajte dobivene rezultate.

c) Kolika je vjerojatnost da igrač osvoji barem dva dijamanta?

Rješenje:

a) $X = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1/2 & 2/7 & 1/7 & 2/35 & 1/70 \end{pmatrix}$.

b) $Y = X - 1$ - slučajna varijabla kojom je modeliran broj osvojenih dijamanata; $EY = 4/5$; $\operatorname{Var} Y = 118/175$.

c) $P\{Y \geq 2\} = P\{X \geq 3\} = 4/7$.

Zadatak 2.39. Automat za igre na sreću u nekoj kockarnici programiran je tako da se prirodan broj n realizira s vjerojatnošću $2/3^n$, tj.

$$P\{X = n\} = \frac{2}{3^n}, \quad n \in \mathbb{N}.$$

Izračunajte matematičko očekivanje i varijancu slučajne varijable X te pomoću Čebiševljeve nejednakosti ocijenite vjerojatnost da realizacija slučajne varijable X od njezinog očekivanja odstupa za barem dvije standardne devijacije.

Rješenje: $EX = 3/2$, $\mathrm{Var}\,X = 3/4$, $P\left\{\left|X - \frac{3}{4}\right| \geq \sqrt{3}\right\} \leq \frac{1}{4}$.

Zadatak 2.40. Slučajna varijabla X ima binomnu distribuciju s parametrima n i p, tj. $X \sim \mathcal{B}(n,p)$. Odredite distribuciju slučajne varijable $Y = 3X + 1$ te izračunajte njezino matematičko očekivanje i varijancu.

Rješenje: $EY = 3np + 1$, $\mathrm{Var}\,X = 9np(1-p)$.

Zadatak 2.41. Za zadane realne funkcije realne varijable odredite vrijednost nepoznate konstante tako da svaka od njih bude funkcija gustoće neke neprekidne slučajne varijable:

a) $f_X(x) = \begin{cases} ax & , \ 0 \leq x < 1 \\ 0 & , \ \text{inače} \end{cases}$,

b $g_X(x) = \begin{cases} b\cos 2x & , \ x \in [-\pi/4, \pi/4\rangle \\ 0 & , \ \text{inače} \end{cases}$,

c) skicirajte grafove funkcija f i g.

Rješenje: a) $a = 2$, b) $b = 1$.

Zadatak 2.42. Neka je X neprekidna slučajna varijabla zadana funkcijom gustoće

$$f_X(x) = \begin{cases} 2x & , \ 0 \leq x < 1 \\ 0 & , \ \text{inače} \end{cases}.$$

a) Odredite funkciju distribucije slučajne varijable X i skicirajte njezin graf.

b) Izračunajte $P\{0.25 < X \leq 2\}$.

Rješenje:

a) $F_X(x) = \begin{cases} 0 & , \ x < 0 \\ x^2 & , \ 0 \leq x < 1 \\ 1 & , \ x \geq 1 \end{cases}$.

b) $P\{0.25 < X \leq 2\} = 15/16$.

Zadatak 2.43. Neka je X neprekidna slučajna varijabla zadana funkcijom gustoće

$$f_X(x) = \begin{cases} \cos 2x & , \ x \in [-\pi/4, \pi/4\rangle \\ 0 & , \ \text{inače} \end{cases}.$$

a) Odredite funkciju distribucije slučajne varijable X i skicirajte njezin graf.

b) Izračunajte $P\{0 < X \leq \pi/8\}$.

Rješenje:

a) $F_X(x) = \begin{cases} 0 & , \ x < -\pi/4 \\ \frac{1}{2} + \cos(x)\sin(x) & , \ -\pi/4 \leq x < \pi/4 \\ 1 & , \ x \geq \pi/4 \end{cases}$.

b) $P\{0.25 < X \leq 2\} = 15/16$.

Zadatak 2.44. Zadana je funkcija gustoće slučajne varijable X:

$$f_X(x) = \begin{cases} \frac{1}{3} & , \ 0 \leq x < 1 \\ \frac{2}{3} & , \ 1 \leq x < 2 \\ 0 & , \ \text{inače} \end{cases}.$$

Izračunajte funkciju distribucije slučajne varijable X te skicirajte grafove funkcije gustoće i funkcije distribucije.

Rješenje: $F_X(x) = \begin{cases} 0 & , \ x < 0 \\ \frac{1}{3}x & , \ 0 \leq x < 1 \\ \frac{1}{3}(2x-1) & , \ 1 \leq x < 2 \\ 1 & , \ x \geq 2 \end{cases}$.

Zadatak 2.45. Pretpostavimo da je vrijeme trajanja u godinama nekog elektroničkog uređaja modelirano slučajnom varijablom X s funkcijom gustoće

$$f_X(x) = \begin{cases} 2e^{-2x} & , \ x \geq 0 \\ 0 & , \ x < 0 \end{cases}.$$

a) Odredite vjerojatnost da će taj uređaj trajati barem tri godine.

b) Odredite očekivano vrijeme trajanja tog uređaja.

Rješenje:

a) $P\{X \geq 3\} = e^{-6} \approx 0.0025$.

b) $EX = 1/2$.

Zadatak 2.46. Vrijeme trajanja u mjesecima nekog elektroničkog uređaja modelirano je slučajnom varijablom X s funkcijom gustoće

$$f_X(x) = \begin{cases} kxe^{-x^2/9} & , \ x > 0 \\ 0 & , \ x \leq 0 \end{cases}.$$

a) Odredite vrijednost konstante k tako da f_X bude dobro definirana funkcija gustoće.

b) Odredite funkciju distribucije slučajne varijable X.

c) Što je vjerojatnije - da se uređaj pokvari tijekom prva tri mjeseca upotrebe ili da mu vrijeme trajanja bude dulje od mjesec dana? Ako uređaj ima jamstvo u trajanju od mjesec dana, kolika je vjerojatnosti da će uređaj trebati popravak za vrijeme trajanja jamstvenog roka?

Rješenje:

a) $k = 2/9$.

c) $P\{X \leq 3\} = (e-1)/e \approx 0.63$, $P\{X > 1\} = e^{-1/9} \approx 0.89$.

Zadatak 2.47. Neka je X neprekidna slučajna varijabla s funkcijom gustoće

$$f_X(x) = \begin{cases} k\sin\left(\frac{x}{3}\right) & , \ x \in \langle 0, \pi\rangle \\ 0 & , \ x \text{ inače} \end{cases}.$$

a) Izračunajte vrijednost konstante k.

b) Odredite funkciju distribucije slučajne varijable X i izračunajte $P\{\pi/2 \leq X < 2\pi\}$.

c) Izračunajte matematičko očekivanje i varijancu slučajne varijable X.

Rješenje:

a) $k = 2/3$.

b) $F_X(x) = \begin{cases} 0 & , \ x < 0 \\ 2 - 2\cos(x/3) & , \ 0 \leq x < \pi \\ 1 & , \ x \geq \pi \end{cases}$,
$P\{\pi/2 \leq X < 2\pi\} = \sqrt{3} - 1$.

c) $EX = 3\sqrt{3} - \pi$, $\operatorname{Var} X = -2\pi^2 + 12\sqrt{3}\pi - 45$.

Zadatak 2.48. Neka je X neprekidna slučajna varijabla s funkcijom gustoće

$$f_X(x) = \begin{cases} k\cos\left(\frac{x}{3}\right) & , \ x \in \langle 0, \pi/2\rangle \\ 0 & , \ x \text{ inače} \end{cases}.$$

a) Izračunajte vrijednost konstante k.

b) Odredite funkciju distribucije slučajne varijable X i izračunajte $P\{\pi/4 \leq X < \pi\}$.

c) Izračunajte matematičko očekivanje i varijancu slučajne varijable X.

Rješenje:

a) $k = 2/3$.

b) $F_X(x) = \begin{cases} 0 & , \ x < 0 \\ 2\sin(x/3) & , \ 0 \leq x < \pi/2 \\ 1 & , \ x \geq \pi/2 \end{cases}$,
$P\{\pi/4 \leq X < \pi\} = (1 + \sqrt{2} - \sqrt{3})/\sqrt{2}$.

c) $EX = \frac{\pi}{2} + 3(\sqrt{3} - 2)$, $\operatorname{Var} X = 6\pi^2 + 36\sqrt{3} - 81$.

Zadatak 2.49. Svakog radnog dana osoba autobusom odlazi na posao. Autobusi na stanicu dolaze redovito svakih pet minuta. Trenutak dolaska osobe na stanicu smatramo slučajnim trenutkom između dolaska dvaju uzastopnih autobusa. Vrijeme koje protekne od dolaska osobe na stanicu do dolaska prvog sljedećeg autobusa (vrijeme čekanja) modeliramo slučajnom varijablom X za koju je poznato da je vjerojatnost da osoba autobus čeka najviše x vremena proporcionalna tom vremenu čekanja x.

a) Odredite funkciju distribucije i funkciju gustoće slučajne varijable X.

b) Odredite vjerojatnost da osoba autobus čeka

- manje od dvije minute,
- više od tri minute,
- manje od dvije minute ili barem tri minute,
- barem dvije, ali manje od tri minute.

Rješenje: a) $X \sim \mathcal{U}(0,5)$.

Zadatak 2.50. Na stroju koji proizvodi bakrenu žicu povremeno dolazi do smetnji koje uzrokuju nepravilnost na dijelu žice proizvedenom u trenutku smetnje. Duljinu žice (u metrima) između dviju uzastopnih nepravilnosti možemo modelirati slučajnom varijablom s funkcijom gustoće

$$f_X(x) = \begin{cases} k(1+x)^{-3} & , \ x > 0 \\ 0 & , \ x \leq 0 \end{cases}.$$

a) Odredite vrijednost konstante k.

b) Odredite funkciju distribucije slučajne varijable X.

c) Kolika je vjerojatnost da se nepravilnost na žici pojavi između 0.4 i 0.45 metara?

Rješenje:

a) $k = 2$,

b) $F_X(x) = \begin{cases} 1-(1+x)^{-2} & , \ x > 0 \\ 0 & , \ x \leq 0 \end{cases}$,

c) $P\{0.4 \leq X \leq 0.45\} = 0.0346$.

Zadatak 2.51. Unutar kruga radijusa R na slučajan način biramo točku. Nađite funkciju distribucije i funkciju gustoće slučajne varijable koja je definirana kao udaljenost odabrane točke od središta kruga.

Rješenje: $f(x) = \begin{cases} 0 & , \quad x < 0 \\ x^2/R^2 & , \ 0 \leq x < R \\ 1 & , \quad x \geq R \end{cases}.$

Zadatak 2.52. Na slučajan način odabiremo točku T unutar kvadrata stranice 2. Neka je vrijednost neprekidne slučajne varijable X najmanja udaljenost te točke od stranice kvadrata.

a) Odredite funkciju distribucije slučajne varijable X.

b) Odredite matematičko očekivanje i varijancu slučajne varijable X.

Zadatak 2.53. Slučajna varijabla X zadana je funkcijom gustoće:

$$f_X(x) = \begin{cases} x+1 & , \ -1 \leq x < 0 \\ 1-x & , \ 0 \leq x < 1 \\ 0 & , \ \text{inače.} \end{cases}$$

Odredite funkciju distribucije slučajne varijable X te izračunajte njezino matematičko očekivanje i varijancu.

Zadatak 2.54. Odredite vrijednost konstante k tako da funkcija

$$f(x) = \begin{cases} kxe^{-x} & , \ x \in \langle 0, \infty \rangle \\ 0 & , \quad \text{inače} \end{cases}$$

bude funkcija gustoće neprekidne slučajne varijable X te izračunajte njezino matematičko očekivanje.

Zadatak 2.55. Odredite vrijednost konstante k tako da funkcija

$$f(x) = \begin{cases} k\dfrac{x}{a^2}\, e^{-\frac{x^2}{2a^2}} & , \ x \in \langle 0, \infty \rangle \\ 0 & , \quad \text{inače} \end{cases}$$

bude funkcija gustoće neprekidne slučajne varijable X, odredite funkciju distribucije slučajne varijable X, njezino očekivanje i varijancu te izračunajte $P\{X \leq 2\}$.

Zadatak 2.56. Odredite vrijednost konstante k tako da funkcija

$$f(x) = \begin{cases} \dfrac{kx}{1+x} & , \ x \in \langle 0, e-1 \rangle \\ 0 & , \quad \text{inače} \end{cases}$$

bude funkcija gustoće neprekidne slučajne varijable X, odredite funkciju distribucije slučajne varijable X i izračunajte $P\{X \leq 3\}$.

Zadatak 2.57. Odredite vrijednost konstante k tako da funkcija

$$f(x) = \begin{cases} ke^{2x} & , \ x \in \langle -\infty, 0 \rangle \\ 0 & , \quad \text{inače} \end{cases}$$

bude funkcija gustoće neprekidne slučajne varijable X te izračunajte njezinu funkciju distribucije, matematičko očekivanje, varijancu i $P\{0 \leq X < 13579\}$.

Zadatak 2.58. Odredite vrijednost konstante k tako da funkcija

$$f(x) = \begin{cases} ka^{-2x} & , \ x \in \langle 0, +\infty \rangle \\ 0 & , \quad \text{inače} \end{cases},$$

gdje je $a > 0$ i $a \neq 1$, bude funkcija gustoće neprekidne slučajne varijable X, izračunajte njezinu funkciju distribucije, očekivanje i varijancu te $P\{-2468 \leq X < 0\}$.

Zadatak 2.59. Odredite vrijednost konstante k tako da funkcija

$$f_X(x) = \begin{cases} k \ln \sqrt{x} & , \ x \in \langle 1, e \rangle \\ 0 & , \quad \text{inače} \end{cases}$$

bude funkcija gustoće neprekidne slučajne varijable X te odredite njenu funkciju distribucije F_X i vjerojatnost $P\{0 \leq X < e/2\}$.

Zadatak 2.60. Odredite vrijednost konstante k tako da funkcija

$$f(x) = \begin{cases} kxe^{x^2} & , \ x \in [0, e] \\ 0 & , \ \text{inače} \end{cases}$$

bude funkcija gustoće neprekidne slučajne varijable X te odredite vrijednost njezine funkcije distribucije za $x = e - 1$.

Zadatak 2.61. Odredite vrijednost konstante k tako da funkcija

$$f(x) = \begin{cases} kx^2 e^{x^3 - 1} & , \ x \in [0, e] \\ 0 & , \ \text{inače} \end{cases}$$

bude funkcija gustoće neprekidne slučajne varijable X te odredite vrijednost njezine funkcije distribucije za $x = 1$.

Zadatak 2.62. Odredite vrijednost konstante k tako da funkcija

$$f(x) = \begin{cases} ke^{2-x} & , \ x \in [0, +\infty\rangle \\ 0 & , \ \text{inače} \end{cases}$$

bude funkcija gustoće neprekidne slučajne varijable X te izračunajte njezino matematičko očekivanje.

Zadatak 2.63. Odredite vrijednost konstante k tako da funkcija

$$f(x) = \begin{cases} kx \sin x & , \ x \in [0, \pi] \\ 0 & , \ \text{inače} \end{cases}$$

bude funkcija gustoće neprekidne slučajne varijable X te odredite njezino matematičko očekivanje.

Zadatak 2.64. Odredite vrijednost konstante k tako da funkcija

$$f(x) = \begin{cases} k(x-1)^3, & x \in [1, 2] \\ 0, & \text{inače.} \end{cases}$$

bude funkcija gustoće neprekidne slučajne varijable X te odredite njezino matematičko očekivanje, varijancu i $P\{\frac{3}{2} < X \leq \frac{7}{4}\}$.

Zadatak 2.65. Odredite vrijednost konstante k tako da funkcija

$$f(x) = \begin{cases} kx(x-1)^2, & x \in [0, 3] \\ 0, & \text{inače.} \end{cases}$$

bude funkcija gustoće neprekidne slučajne varijable X te odredite njezino matematičko očekivanje, varijancu i $P\{2 < X \leq 4\}$.

Zadatak 2.66. Neka je F_X funkcija distribucije neprekidne slučajne varijable X. Odredite funkciju distribucije slučajne varijable $Y = -X$.

Rješenje: $F_Y(y) = 1 - F_X(-y)$.

Zadatak 2.67. Slučajna varijabla X ima funkciju gustoće $f_X(x)$. Odredite funkcije gustoća sljedećih slučajnih varijabli:

a) $Y = e^X$, b) $Y = e^{-X}$,

c) $Y = \sqrt{X}$, $X > 0$, d) $Y = \mathrm{sh}X = \dfrac{e^X - e^{-X}}{2}$.

Rješenje:

a) $f_Y(y) = \begin{cases} \frac{1}{y} f_X(\ln y) & , \ y > 0 \\ 0 & , \ \text{inače} \end{cases}$,

b) $f_Y(y) = \begin{cases} \frac{1}{y} f_X(-\ln y) & , \ y > 0 \\ 0 & , \ \text{inače} \end{cases}$,

c) $f_Y(y) = \begin{cases} 2y f_X(y^2) & , \ y \geq 0 \\ 0 & , \ \text{inače} \end{cases}$.

Zadatak 2.68. Neka je $f_X(x)$ funkcija gustoće neprekidne slučajne varijable X. Odredite funkcije gustoća slučajnih varijabli $X^3 + 1$, $X^5 + 15$, $X^7 - 7$, $X\sqrt{X}$, $-e^X$, $X - e^{-X}$, e^{X^2}.

Zadatak 2.69. Slučajna varijabla X zadana je funkcijom gustoće

$$f_X(x) = \begin{cases} x^2/3 & , \ -1 < x < 2 \\ 0 & , \quad \text{inače} \end{cases}.$$

Odredite funkciju gustoće slučajne varijable $Y = X^2$.

Rješenje: $f_Y(y) = \begin{cases} \sqrt{y}/3 & , \ 0 < y < 1 \\ \sqrt{y}/6 & , \ 1 \leq y < 4 \\ 0 & , \quad \text{inače} \end{cases}$.

Zadatak 2.70. Slučajna varijabla X ima Cauchyjevu distribuciju s funkcijom gustoće

$$f_X(x) = \frac{1}{\pi(1 + x^2)}, \ x \in \mathbb{R}.$$

Odredite funkciju gustoće i funkciju distribucije slučajne varijable

a) $Y = X^2$,

b) $Z = 1/X$.

Rješenje:

a) $f_Y(y) = \begin{cases} \frac{1}{\pi\sqrt{y}(1+y)} & , y > 0 \\ 0 & , y \leq 0 \end{cases}$,

b) $f_Y(y) = \frac{1}{\pi(1+y^2)}$, $y \in \mathbb{R}$.

Zadatak 2.71. Slučajna varijabla X ima uniformnu distribuciju na intervalu $\langle 0, 1\rangle$, tj. $X \sim \mathcal{U}(0, 1)$. Odredite funkciju gustoće i funkciju distribucije slučajne varijable

a) $Y = aX + b$, $a, b \in \mathbb{R}$, $a \neq 0$,

b) $Z = -\ln X$.

Rješenje:

a) $a > 0: \quad f_Y(y) = \begin{cases} 1/a & , b < y < b - a \\ 0 & , \text{inače} \end{cases}$,

$a < 0: \quad f_Y(y) = \begin{cases} -1/a & , b < y < a + b \\ 0 & , \text{inače} \end{cases}$,

b) $f_Z(z) = \begin{cases} e^{-z} & , z > 0 \\ 0 & , z \leq 0 \end{cases}$.

Zadatak 2.72. Slučajna varijabla X ima uniformnu distribuciju na intervalu $\langle -\pi/2, \pi/2\rangle$. Odredite funkciju distribucije i funkciju gustoće slučajne varijable $Y = \cos X$.

Zadatak 2.73. Slučajna varijabla X ima uniformnu distribuciju na intervalu $\langle 0, 5\rangle$. Odredite funkciju distribucije slučajne varijable $Y = e^{-X}$.

Zadatak 2.74. Slučajna varijabla X ima uniformnu distribuciju na intervalu $\langle 1, 2\rangle$, tj. $X \sim \mathcal{U}(1, 2)$. Odredite funkciju gustoće i funkciju distribucije slučajne varijable $Y = -2X + 1$.

Zadatak 2.75. Slučajna varijabla X ima uniformnu distribuciju na intervalu $\langle 2, 4\rangle$. Odredite funkciju gustoće i funkciju distribucije slučajne varijable $Y = e^X$.

Zadatak 2.76. Slučajna varijabla X ima uniformnu distribuciju na intervalu $\langle 1, 3\rangle$. Odredite funkciju gustoće i funkciju distribucije slučajne varijable $Y = \ln(X)$.

Zadatak 2.77. Slučajna varijabla X ima uniformnu distribuciju na intervalu $\langle -1, 1\rangle$. Odredite funkciju distribucije slučajne varijable $Y = 1/X^2$.

Zadatak 2.78. Slučajna varijabla X ima eksponencijalnu distribuciju s parametrom $\lambda = 7$. Odredite funkciju gustoće slučajne varijable $Y = \ln X$.

Zadatak 2.79. Slučajna varijabla X ima eksponencijalnu distribuciju s parametrom $1/2$. Odredite funkciju gustoće vjerojatnosti i funkciju distribucije vjerojatnosti slučajne varijable $Y = 1/X$.

Zadatak 2.80. Slučajna varijabla X ima eksponencijalnu distribuciju s parametrom $\lambda = 1$. Odredite funkciju distribucije slučajne varijable $Y = (X - 1)^2$.

Poglavlje 3

Slučajni vektor

Ukoliko u jednom istraživanju za dani slučajni pokus pratimo nekoliko različitih slučajnih varijabli, moguće veze među njima nećemo dokučiti ako ih proučavamo samo svaku za sebe. Međutim, vrlo su često veze između pojedinih slučajnih varijabli istog slučajnog pokusa od primarnog interesa.

Primjer 3.1. *Broj komaraca na području grada Osijeka tijekom lipnja ovog ljeta jest slučajna varijabla X. Količina je oborina u svibnju na istom području također slučajna varijabla Y. Teoretski je za očekivati da između tih dviju slučajnih varijabli postoji veza. Da bismo tu vezu opisali, moramo naučiti opisivati uređen par slučajnih varijabli (X, Y).*

Od interesa je, dakle, izraditi matematički model koji omogućuje opisivanje nekoliko slučajnih varijabli danog slučajnog pokusa istovremeno. Radi ilustracije ideje modela proučimo jednostavan primjer slučajnog pokusa koji se sastoji od izvlačenja dvije kuglice iz iste kutije.

Primjer 3.2. *Iz kutije koja sadrži n bijelih i m crnih kuglica izvlačimo dvije kuglice bez vraćanja u kutiju. Neka slučajna varijabla X prima vrijednost 1 ako je u prvom izvlačenju izvučena bijela kuglica, a 0 ako je u prvom izvlačenju izvučena crna kuglica. Analogno, neka slučajna varijabla Y prima vrijednost 1 ako je u drugom izvlačenju izvučena bijela kuglica, a 0 ako je u drugom izvlačenju izvučena crna kuglica.*

Rezultate tog slučajnog pokusa možemo opisati dvjema slučajnim varijablama: X i Y. Ako od njih napravimo uređeni par slučajnih varijabli (X, Y), onda je skup svih mogućih ishoda $\mathcal{R}(X, Y) = \{(1, 1), (1, 0), (0, 1), (0, 0)\}$. Ovdje je prvo mjesto u paru rezervirano za ishode prvog pokusa (odnosno slučajnu varijablu X), a drugo mjesto rezervirano je za ishode drugog pokusa (odnosno slučajnu varijablu Y).

Vjerojatnosna su svojstva tog uređenog para slučajnih varijabli potpuno su određena poznavanjem brojeva:

$$p_{(1,1)} = P(\{X = 1\} \cap \{Y = 1\}) = P\{(X, Y) = (1, 1)\} = \frac{n}{n+m}\,\frac{n-1}{n+m-1},$$

$$p_{(1,0)} = P(\{X = 1\} \cap \{Y = 0\}) = P\{(X,Y) = (1,0)\} = \frac{n}{n+m}\,\frac{m}{n+m-1},$$
$$p_{(0,1)} = P(\{X = 0\} \cap \{Y = 1\}) = P\{(X,Y) = (0,1)\} = \frac{m}{n+m}\,\frac{n}{n+m-1},$$
$$p_{(0,0)} = P(\{X = 0\} \cap \{Y = 0\}) = P\{(X,Y) = (0,0)\} = \frac{m}{n+m}\,\frac{m-1}{n+m-1},$$

koje smo izračunali korištenjem poznate relacije

$$P(A \cap B) = P(A \mid B)P(B),$$

ako je $P(B) > 0$ (vidi poglavlje 1.9).

U ovom poglavlju definirat ćemo **slučajni vektor** te opisati specijalno diskretan i neprekidan slučaj analogno kao diskretnu i neprekidnu slučajnu varijablu.

Primjer 3.2 i dosadašnja saznanja o slučajnim varijablama upućuju na to da je potrebno n-dimenzionalan, $n \in \mathbb{N}$, slučajni vektor definirati kao funkciju s prostora elementarnih događaja Ω u skup $\mathbb{R}^n$. Međutim, s obzirom da vjerojatnost općenito nije definirana na $\mathcal{P}(\Omega)$, ne možemo svaku takvu funkciju zvati slučajnim vektorom. Analogno kao u poglavlju 2.3, n-dimenzionalni slučajni vektor definiramo na sljedeći način.

Definicija 3.1. *Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor pridružen slučajnom pokusu. Funkciju $(X_1, \ldots, X_n)$ koja svakom ishodu pokusa pridružuje uređenu n-torku realnih brojeva $(x_1, \ldots, x_n)$ zovemo* n**-dimenzionalan slučajni vektor** *ako vrijedi:*

$$\{X_1 \leq x_1\} \cap \cdots \cap \{X_n \leq x_n\} \in \mathcal{F}$$

za svaki $x_1 \in \mathbb{R}, \ldots,\ x_n \in \mathbb{R}$.

Zbog jednostavnosti uvest ćemo sljedeću oznaku:

$$\{X_1 \leq x_1\} \cap \cdots \cap \{X_n \leq x_n\} = \{X_1 \leq x_1, \ldots, X_n \leq x_n\}.$$

Slično kao u jednodimenzionalnom slučaju (poglavlje 2.3), vjerojatnosna svojstva svakog slučajnog vektora opisujemo **funkcijom distribucije**. Funkcija distribucije slučajnog vektora dana je sljedećom definicijom.

Definicija 3.2. *Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor pridružen slučajnom pokusu i $(X_1, \ldots, X_n)$ slučajni vektor. Funkciju*

$$\begin{aligned} F &: \mathbb{R}^n \to [0,1], \\ F(x_1, \ldots, x_n) &= P\{X_1 \leq x_1, \ldots, X_n \leq x_n\} \end{aligned} \tag{3.1}$$

zovemo **funkcija distribucije slučajnog vektora** $(X_1, \ldots, X_n)$.

Neka svojstva funkcije distribucije slučajnog vektora izreći ćemo za dvodimenzionalan slučaj.[1]

Svojstva funkcije distribucije dvodimenzionalnoga slučajnog vektora

Neka je zadan dvodimenzionalan slučajni vektor (X, Y) svojom funkcijom distribucije $F(x, y)$. Tada vrijedi:

1. Za realne brojeve x_1, x_2, y, takve da je $x_1 < x_2$, vrijedi $F(x_1, y) \leq F(x_2, y)$.

2. Neka su x, y_1, y_2 realni brojevi. Ako je $y_1 < y_2$, onda je $F(x, y_1) \leq F(x, y_2)$.

3. $$\lim_{x \to -\infty} F(x, y) = F(-\infty, y) = 0,$$
$$\lim_{y \to -\infty} F(x, y) = F(x, -\infty) = 0,$$
$$\lim_{\substack{x \to \infty \\ y \to \infty}} F(x, y) = F(\infty, \infty) = 1.$$

4. Neprekidnost zdesna u svakoj varijabli:
$$\lim_{x \downarrow x_0} F(x, y_0) = \lim_{y \downarrow y_0} F(x_0, y) = F(x_0, y_0).$$

5. Za $x_1 < x_2$ i $y_1 < y_2$ jest
$$F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \geq 0.$$

Zadatak 3.1. Korištenjem svojstava vjerojatnosti dokažite navedena svojstva funkcije distribucije $F(x, y)$.

Kao i kod jednodimenzionalnih slučajnih varijabli razmatrat ćemo diskretan i neprekidan tip slučajnih vektora kod kojih se funkcija distribucije može računati korištenjem tablice distribucije (diskretan slučaj), odnosno funkcije gustoće (neprekidan slučaj). Također ćemo, zbog jednostavnosti zapisa i lakšeg razumijevanja dokaza, uglavnom razmatranja provoditi na dvodimenzionalnom slučaju, dok ćemo za višedimenzionalne slučajeve samo navesti analogne rezultate.

Primjer 3.3. *Igra se temelji na bacanju dvaju pravilno izrađenih novčića. U jednoj partiji igre igrač osvaja one novčiće na kojima se pri bacanju realizirala glava, dok novčiće na kojima se realiziralo pismo ne osvaja. Prema tome, ishod bacanja tih dvaju novčića možemo modelirati dvodimenzionalnim diskretnim slučajnim vektorom (X, Y), gdje je*

[1] Poopćenje za n-dimenzionalan slučajni vektor čitatelj može naći u [26, str. 267].

X slučajna varijabla koja prima vrijednost 1 ako se pri bacanju prvog novčića realizirala glava, a vrijednost 0 ako se realiziralo pismo,

Y slučajna varijabla koja prima vrijednost 1 ako se pri bacanju drugog novčića realizirala glava, a vrijednost 0 ako se realiziralo pismo.

Dakle, (X, Y) je slučajni vektor koji prima vrijednosti iz skupa

$$\mathcal{R}(X, Y) = \{(0,0), (0,1), (1,0), (1,1)\}$$

s vjerojatnostima

$$p(0,0) = P\{X = 0, Y = 0\} = \frac{1}{4}, \qquad p(0,1) = P\{X = 0, Y = 1\} = \frac{1}{4},$$
$$p(1,0) = P\{X = 1, Y = 0\} = \frac{1}{4}, \qquad p(1,1) = P\{X = 1, Y = 1\} = \frac{1}{4}.$$

Ako želimo odrediti funkciju distribucije $F\colon \mathbb{R}^2 \to [0,1]$ tog slučajnog vektora, uočimo:

za $(x, y) \in (\langle -\infty, 0\rangle \times \mathbb{R}) \cup (\mathbb{R} \times \langle -\infty, 0\rangle)$ je $P\{X \le x, Y \le y\} = 0$,

za $(x, y) \in [0,1\rangle \times [0,1\rangle$ je

$$P\{X \le x, Y \le y\} = P\{X = 0, Y = 0\} = \frac{1}{4},$$

za $(x, y) \in [0,1\rangle \times [1,\infty\rangle$ je

$$P\{X \le x, Y \le y\} = P\{X = 0, Y = 0\} + P\{X = 0, Y = 1\} = \frac{1}{2},$$

za $(x, y) \in [1,\infty\rangle \times [0,1\rangle$ je

$$P\{X \le x, Y \le y\} = P\{X = 0, Y = 0\} + P\{X = 1, Y = 0\} = \frac{1}{2},$$

za $(x, y) \in [1,\infty\rangle \times [1,\infty\rangle$ je

$$P\{X \le x, Y \le y\} =$$
$$= P\{X = 0, Y = 0\} + P\{X = 0, Y = 1\} + P\{X = 1, Y = 0\} + P\{X = 1, Y = 1\} = 1.$$

Dakle, funkcija distribucije slučajnog vektora (X, Y) definirana je pravilom

$$F(x,y) = \begin{cases} 0, & (x,y) \in (\langle -\infty, 0\rangle \times \mathbb{R}) \cup (\mathbb{R} \times \langle -\infty, 0\rangle) \\ 1/4, & (x,y) \in [0,1\rangle \times [0,1\rangle \\ 1/2, & (x,y) \in [0,1\rangle \times [1,\infty\rangle \\ 1/2, & (x,y) \in [1,\infty\rangle \times [0,1\rangle \\ 1, & (x,y) \in [1,\infty\rangle \times [1,\infty\rangle. \end{cases}$$

3.1 Diskretan dvodimenzionalan slučajni vektor

Analogno kao kod diskretne slučajne varijable, diskretan slučajni vektor (X, Y) prima vrijednosti iz konačnog ili prebrojivog skupa

$$\mathcal{R}(X, Y) = \{(x_i, y_j) : (i,j) \in I\}, \quad I \subseteq \mathbb{N} \times \mathbb{N},$$

a njegova je funkcija distribucije (tj. vjerojatnosna svojstva) potpuno određena poznavanjem brojeva

$$p(x_i, y_j) = P\{X = x_i, Y = y_j\}, \quad (i,j) \in I \subseteq \mathbb{N} \times \mathbb{N}. \tag{3.2}$$

Naime, za funkciju distribucije vrijedi:

$$F(x,y) = P\{X \le x, Y \le y\} = \sum_{x_i \le x} \sum_{y_j \le y} P\{X = x_i, Y = y_j\} = \sum_{x_i \le x} \sum_{y_j \le y} p(x_i, y_j).$$

Niz brojeva definiran izrazom (3.2) zvat ćemo **distribucija diskretnoga slučajnog vektora** (X,Y). Da bismo pojednostavnili zapis distribucije slučajnog vektora, u nastavku ćemo teksta pisati $I = \mathbb{N} \times \mathbb{N}$, podrazumijevajući pod tim da je skup indeksa najviše $\mathbb{N} \times \mathbb{N}$, a zapravo sadrži samo one indekse koji se pojavljuju u konkretnom slučaju.

Primjer 3.4. *Promatrajmo slučajni pokus koji se sastoji od jednog bacanja pravilno izrađene igraće kocke i jednog bacanja pravilno izrađenog novčića. Skup elementarnih događaja ovog pokusa jest skup*

$$\Omega = \{\{1,P\},\{2,P\},\{3,P\},\{4,P\},\{5,P\},\{6,P\},\\ \{1,G\},\{2,G\},\{3,G\},\{4,G\},\{5,G\},\{6,G\}\}.$$

Neka slučajna varijabla X daje broj koji se okrenuo na kocki, a slučajna varijabla Y daje 1 ako je bacanjem novčića palo pismo i 0 ako je bacanjem novčića pala glava. U tom pokusu bacanje kocke i novčića ne ovise jedno o drugom pa vrijedi:

$$p(1,1) = P\{X = 1, Y = 1\} = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12},$$
$$p(1,0) = P\{X = 1, Y = 0\} = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}.$$

Analogno dobijemo da je

$$p(i,j) = P\{X = i, Y = j\} = \frac{1}{12}, \quad i \in \{1,2,3,4,5,6\}, \; j \in \{0,1\}.$$

Primjer 3.5. *Promatrajmo sljedeću igru: prvo bacamo pravilno izrađenu igraću kockicu. Ako se okrenuo broj 6, ostvarujemo pravo na bacanje pravilno izrađenog novčića, pri čemu osvajamo bod ako padne pismo, a u suprotnom gubimo igru.*

Taj pokus možemo opisati skupom elementarnih događaja

$$\Omega = \{(1,N),(2,N),(3,N),(4,N),(5,N),(6,I_p),(6,I_g)\},$$

gdje N znači da drugu igru ne igramo, I_p da drugu igru igramo i da se u njoj dogodilo pismo, a I_g da drugu igru igramo i da se u njoj nije dogodilo pismo.

Neka je X slučajna varijabla koja prima vrijednost 1 ako se pri bacanju kockice realizirao broj 6, a u suprotnom prima vrijednost 0. Neka je Y slučajna varijabla koja prima vrijednost 1 ako je igrač osvojio bod, a u suprotnom prima vrijednost 0. Osvajanje boda u igri opisano je realizacijom događaja $\{(X,Y) = (1,1)\}$. Distribucija je tog slučajnog vektora sljedeća:

$$p(0,0) = P\{X = 0, Y = 0\} = P(\{Y = 0\} \mid \{X = 0\})\, P\{X = 0\} = 1 \cdot \frac{5}{6},$$
$$p(0,1) = P\{X = 0, Y = 1\} = P(\{Y = 1\} \mid \{X = 0\})\, P\{X = 0\} = 0,$$
$$p(1,0) = P\{X = 1, Y = 0)\} = P(\{Y = 0\} \mid \{X = 1\})\, P\{X = 1\} = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12},$$
$$p(1,1) = P\{X = 1, Y = 1\} = P(\{Y = 1\} \mid \{X = 1\})\, P\{X = 1\} = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}.$$

Uočimo da za niz brojeva $(p(x_i, y_j), i \in \mathbb{N},\ j \in \mathbb{N})$ koji definira distribuciju diskretnog dvodimenzionalnog slučajnog vektora vrijedi:

1. $p(x_i, y_j) \geq 0$ za sve $i \in \mathbb{N},\ j \in \mathbb{N}$,
2. $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p(x_i, y_j) = 1$.

Ukoliko je skup vrijednosti slučajnog vektora (X, Y) konačan, tj. slučajni vektor prima vrijednosti iz skupa $\mathcal{R}(X, Y) = \{(x_i, y_j) : i = 1, 2, \dots, m,\ j = 1, 2, \dots, n\}$, njegova se distribucija pregledno može prikazati tablicom 3.1 koju zovemo **tablica distribucije dvodimenzionalnoga diskretnog slučajnog vektora**.[2]

$X \setminus Y$	y_1	y_2	y_3	$\cdots$	y_n
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	$p(x_1, y_3)$	$\cdots$	$p(x_1, y_n)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	$p(x_2, y_3)$	$\cdots$	$p(x_2, y_n)$
$\vdots$					
x_m	$p(x_m, y_1)$	$p(x_m, y_2)$	$p(x_m, y_3)$	$\cdots$	$p(x_m, y_n)$

Tablica 3.1: Tablica distribucije diskretnog slučajnog vektora (X, Y) sa skupom vrijednosti $\mathcal{R}(X, Y) = \{(x_i, y_j) : i = 1, \dots, m,\ j = 1, \dots, n\}$

Tako je, na primjer, sljedećom tablicom dana distribucija slučajnog vektora (X, Y) iz primjera 3.5:

$X \setminus Y$	0	1
0	$5/6$	0
1	$1/12$	$1/12$

Ukoliko je na diskretnom vjerojatnosnom prostoru $(\Omega, \mathcal{P}(\Omega), P)$ dan slučajni vektor (X, Y), onda su komponente tog slučajnog vektora, tj. funkcije $X : \Omega \to \mathbb{R}$ i $Y : \Omega \to \mathbb{R}$, slučajne varijable na tom istom vjerojatnosnom prostoru. Njihove distribucije nazivamo **marginalnim distribucijama** slučajnog vektora (X, Y). U nastavku ćemo poglavlja pokazati kako iz distribucije slučajnog vektora možemo izračunati njegove marginalne distribucije.

Primjer 3.6. *Iz kutije koja sadrži tri kupona s oznakama* 1*,* 2 *i* 3 *izvlačimo dva kupona bez vraćanja u kutiju. Rezultat izvlačenja modeliramo slučajnim vektorom* (X, Y)*, gdje je* X *slučajna varijabla kojom je modeliran rezultat prvog izvlačenja, a* Y *slučajna varijabla kojom je modeliran rezultat drugog izvlačenja. Distribucija slučajnog vektora* (X, Y) *dana je tablicom*

[2] Takav zapis distribucije slučajnog vektora možemo lako prilagoditi i za slučajni vektor s prebrojivim skupom vrijednosti analogno kao kod tablica distribucije slučajne varijable.

$X \setminus Y$	1	2	3
1	0	1/6	1/6
2	1/6	0	1/6
3	1/6	1/6	0

Slučajne varijable X i Y imaju isti skup vrijednosti: $\mathcal{R}(X) = \mathcal{R}(Y) = \{1, 2, 3\}$. Koristeći činjenicu da familija skupova $\{\{Y = 1\}, \{Y = 2\}, \{Y = 3\}\}$ čini particiju skupa Ω, možemo izračunati $P\{X = i\}$ za $i = 1, 2, 3$ na sljedeći način:

$$P\{X = i\} = P\{X = i, Y = 1\} + P\{X = i, Y = 2\} + P\{X = i, Y = 3\} = \frac{1}{3},$$

i=1,2,3. Dakle, marginalna distribucija komponente X slučajnog vektora (X, Y) dana je tablicom

$$X = \begin{pmatrix} 1 & 2 & 3 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

Sličnim postupkom možemo izračunati i marginalnu distribuciju komponenate Y slučajnog vektora (X, Y):

$$Y = \begin{pmatrix} 1 & 2 & 3 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

Uočimo da su obje marginalne distribucije u tom primjeru jednake, međutim, slučajne varijable X i Y ni u kom slučaju nisu jednake slučajne varijable. Naime, slučajne varijable X i Y definirane na istom vjerojatnosnom prostoru jednake su ako je $X(\omega) = Y(\omega)$ za sve $\omega \in \Omega$. U našem je slučaju čak nemoguće da je $X(\omega) = Y(\omega)$ za bilo koji ω jer bi to značilo da je u oba izvlačenja izvučen isti kupon, što se u spomenutom pokusu ne može dogoditi. Za slučajne varijable koje imaju iste distribucije reći ćemo da su **jednako distribuirane slučajne varijable**, *oznaka $X \stackrel{D}{=} Y$.*

U primjeru 3.6 pokazan je način koji je moguće primijeniti za računanje marginalnih distribucija općenito kod diskretnih dvodimenzionalnih slučajnih vektora. Opišimo sada općenito način kako se iz distribucije diskretnog slučajnog vektora mogu dobiti marginalne distribucije.

Neka je (X, Y) dvodimenzionalan diskretan slučajni vektor sa skupom vrijednosti $\{(x_i, y_j) : i \in \mathbb{N}, j \in \mathbb{N}\}$ i distribucijom

$$p(x_i, y_j) = P\{X = x_i, Y = y_j\}, \quad i \in \mathbb{N}, j \in \mathbb{N}.$$

Uočimo da za sve $i \in \mathbb{N}$ vrijedi:

$$P\{X = x_i\} = P\{X = x_i, Y \in \mathbb{R}\} = \sum_{j \in \mathbb{N}} P\{X = x_i, Y = y_j\} = \sum_{j=1}^{\infty} p(x_i, y_j),$$

a ti brojevi određuju marginalnu distribuciju komponente X slučajnog vektora (X, Y). Analogno možemo izračunati i marginalnu distribuciju komponente Y. Dakle, ako definiramo:

$$\begin{aligned} p_X(x_i) &= \sum_{j=1}^{\infty} p(x_i, y_j), \quad i \in \mathbb{N}, \\ p_Y(y_j) &= \sum_{i=1}^{\infty} p(x_i, y_j), \quad j \in \mathbb{N}, \end{aligned} \tag{3.3}$$

tada skup vrijednosti $\mathcal{R}(X) = \{x_i : i \in \mathbb{N}\}$ i niz $(p_X(x_i),\ i \in \mathbb{N})$ određuju **marginalnu distribuciju komponente** X slučajnog vektora (X, Y), a skup vrijednosti $\mathcal{R}(Y) = \{y_j : j \in \mathbb{N}\}$ i niz $(p_Y(y_j),\ j \in \mathbb{N})$ određuju **marginalnu distribuciju komponente** Y toga slučajnog vektora.

Marginalne se distribucije lako računaju u slučaju konačnog skupa vrijednosti slučajnog vektora koristeći tablični zapis distribucije. Naime, ako je distribucija dana tablicom 3.1, onda sumiranjem vrijednosti u istom stupcu dobijemo vjerojatnost $P\{Y = y_j\} = p_Y(y_j)$, a sumiranjem vrijednosti u istom retku vjerojatnosti $P\{X = x_i\} = p_X(x_i)$, pa tablicu 3.1 možemo proširiti kao što je prikazano tablicom 3.2.

$X \diagdown Y$	y_1	y_2	y_3	$\cdots$	y_n	$p_X(x_i)$
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	$p(x_1, y_3)$	$\cdots$	$p(x_1, y_n)$	$p_X(x_1)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	$p(x_2, y_3)$	$\cdots$	$p(x_2, y_n)$	$p_X(x_2)$
$\vdots$						
x_m	$p(x_m, y_1)$	$p(x_m, y_2)$	$p(x_m, y_3)$	$\cdots$	$p(x_m, y_n)$	$p_X(x_m)$
$p_Y(y_j)$	$p_Y(y_1)$	$p_Y(y_2)$	$p_Y(y_3)$	$\cdots$	$p_Y(y_n)$	1

Tablica 3.2: Proširena tablica distribucije diskretnog slučajnog vektora s konačnim skupom stanja.

Primjer 3.7. *Slučajni pokus sastoji se od bacanja dviju pravilno izrađenih igraćih kockica. Otprije znamo da je skup elementarnih događaja tog slučajnog pokusa*

$$\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}, \quad k(\Omega) = 36.$$

Neka su X i Y diskretne slučajne varijable na $\mathcal{P}(\Omega)$ čije su realizacije određene na sljedeći način:

- *ako se pri bacanju na kockicama okrenu različiti brojevi, X se realizira manjim, a Y većim brojem (npr. ako se na prvoj kockici okrene 3, a na drugoj kockici 6, tada uzimamo da je 3 realizacija slučajne varijable X, a 6 realizacija slučajne varijable Y),*
- *ako se pri bacanju na kockicama okrenu isti brojevi, tada se i X i Y realiziraju tim brojem.*

Očito je

$$\mathcal{R}(X) = \mathcal{R}(Y) = \{1, 2, 3, 4, 5, 6\}.$$

Želimo odrediti marginalne distribucije slučajnog vektora (X, Y).

Slika slučajnog vektora (X, Y) jest skup

$$\mathcal{R}(X, Y) = \{(x, y) : x, y \in \{1, 2, 3, 4, 5, 6\}\}.$$

Vjerojatnosti realizacija slučajnog vektora (X, Y) dane su na sljedeći način:

- *ako je $x > y$, tada je $P\{X = x, Y = y\} = 0$,*
- *ako je $x = y$, tada je $P\{X = x, Y = y\} = 1/36$,*

- *ako je* $x < y$, *tada je* $P\{X = x, Y = y\} = 1/18$.

Marginalne distribucije tog slučajnog vektora računamo na sljedeći način:

- *za fiksni je* $x \in \mathcal{R}(X)$

$$p_X(x) = P\{X = x\} = \sum_{y \in \mathcal{R}(Y)} P\{X = x, Y = y\},$$

- *za fiksni je* $y \in \mathcal{R}(Y)$

$$p_Y(y) = P\{Y = y\} = \sum_{x \in \mathcal{R}(X)} P\{X = x, Y = y\}.$$

Međutim, distribuciju slučajnog vektora (X, Y) *i njegove marginalne distribucije pregledno možemo prikazati sljedećom tablicom (proširenom tablicom distribucije):*

X \ Y	1	2	3	4	5	6	$p_X(x)$
1	1/36	1/18	1/18	1/18	1/18	1/18	11/36
2	0	1/36	1/18	1/18	1/18	1/18	9/36
3	0	0	1/36	1/18	1/18	1/18	7/36
4	0	0	0	1/36	1/18	1/18	5/36
5	0	0	0	0	1/36	1/18	3/36
6	0	0	0	0	0	1/36	1/36
$p_Y(y)$	1/36	3/36	5/36	7/36	9/36	11/36	1

Iz prethodne tablice očitavamo marginalne distribucije:

$$X = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 11/36 & 9/36 & 7/36 & 5/36 & 3/36 & 1/36 \end{pmatrix},$$

$$Y = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/36 & 3/36 & 5/36 & 7/36 & 9/36 & 11/36 \end{pmatrix}.$$

Primjer 3.8. *Na analogan način možemo odrediti marginalne distribucije diskretnog slučajnog vektora koji nema konačnu sliku. Neka je* (X, Y) *diskretan slučajni vektor čija je distribucije dana na sljedeći način:*

$$\mathcal{R}(X, Y) = \mathbb{N}_0 \times \mathbb{N}_0,$$
$$P\{X = x, Y = y\} = \frac{2^x 2^y}{x! y!} e^{-4}, \quad (x, y) \in \mathbb{N}_0 \times \mathbb{N}_0.$$

Odredimo marginalne distribucije tog slučajnog vektora. Slika slučajne varijable X *jest skup* $\mathcal{R}(X) = \mathbb{N}_0$, *a*

$$P\{X = x\} = \sum_{y \in \mathbb{N}_0} P\{X = x, Y = y\} = \frac{2^x}{x!} e^{-4} \sum_{y=0}^{\infty} \frac{2^y}{y!} = \frac{2^x}{x!} e^{-2},$$

tj. $X \sim \mathcal{P}(2)$. *Analognim postupkom slijedi da je* $Y \sim \mathcal{P}(2)$. *Dakle,* X *i* Y *jednako su distribuirane slučajne varijable.*

Primjer 3.9. *Promotrimo igraću kockicu sa svojstvom da se pri jednom njezinom bacanju broj* $i \in \{1, \ldots 6\}$ *okrene s vjerojatnošću* p_i *te pretpostavimo da brojevi* p_i *imaju sljedeća svojstva:*

1. $p_i > 0$ *za sve* $i \in \{1, \ldots 6\}$,

2. $\sum_{i=1}^{6} p_i = 1.$

Uočimo da ako je $p_i = 1/6$ za svaki i, tada se radi o pravilno izrađenoj igraćoj kockici, no općenito ne mora biti tako. Neka je ta kockica bačena n puta zaredom. Realizacija je tog bacanja jedna uređena n-torka elemenata iz skupa $\{1,\ldots,6\}$ ili, preciznije, varijacija s ponavljanjem n-tog razreda skupa $\{1,\ldots,6\}$.

Pretpostavimo da želimo odrediti vjerojatnost da se u tih n bacanja x puta okrenula jedinica i da se y puta okrenula šestica, pri čemu je $x+y \leq n$. U tu svrhu označimo X slučajnu varijablu koja se realizira brojem jedinica, a Y slučajnu varijablu koja se realizira brojem šestica koje su se okrenule u n bacanja te kockice i odredimo vjerojatnost $P\{X=x, Y=y\}$ za $x,y \in \{0,1,\ldots,n\}$ takve da je $x+y \leq n$. Varijacija s ponavljanjem n-tog razreda skupa $\{1,\ldots,6\}$ u kojima su x komponenti jedinice, y komponenti šestice, a preostalih $(n-x-y)$ komponenti dvojke, trojke, četvorke ili petice ima

$$\frac{n!}{x!y!(n-x-y)!}$$

i svaka od njih realizira se s vjerojatnošću

$$p_1^x p_6^y (p_2+p_3+p_4+p_5)^{n-x-y} = p_1^x p_6^y (1-p_1-p_6)^{n-x-y}.$$

Primjenom principa sume slijedi da je

$$P\{X=x, Y=y\} = \frac{n!}{x!y!(n-x-y)!}\, p_1^x p_6^y (1-p_1-p_6)^{n-x-y}. \tag{3.4}$$

Za slučajni vektor (X,Y) s distribucijom 3.4 kažemo da ima multinomnu distribuciju s parametrima n, p_1 i p_6 i pišemo $(X,Y) \sim MULT(n,p_1,p_6)$. Za određivanje distribucije (tablice distribucije) slučajne varijable X iz distribucije 3.4 važno je uočiti da je za realizacije x i y slučajnih varijabli X i Y sa svojstvom $x+y \leq n$ $P\{X=x,Y=y\} > 0$, dok je inače $P\{X=x,Y=y\} = 0$. Sada slijedi da je za $x \in \{0,1,\ldots,n\}$

$$P\{X=x\} = \sum_{y\in\mathcal{R}(\mathcal{Y})} P\{X=x, Y=y\} =$$

$$= \frac{n!}{x!(n-x)!}\, p_1^x \sum_{y=0}^{n-x} \frac{(n-x)!}{y!(n-x-y)!}\, p_6^y (1-p_1-p_6)^{n-x-y} =$$

$$= \binom{n}{x} p_1^x \sum_{y=0}^{n-x} \binom{n-x}{y} p_6^y (1-p_1-p_6)^{n-x-y} = \binom{n}{x} p_1^x (1-p_1)^{n-x},$$

gdje je zadnja suma riješena primjenom binomnog teorema. Dakle, slijedi da je X binomna slučajna varijabla s parametrima $n \in \mathbb{N}$ i $p_1 \in \langle 0,1\rangle$. Analognim postupkom slijedi da je $Y \sim \mathcal{B}(n,p_6)$.

Od velikog su praktičnog interesa slučajne varijable koje nastaju kao kompozicija realne funkcije dviju varijabli i dvodimenzionalnog slučajnog vektora. Preciznije, neka je (X,Y) dvodimenzionalan diskretan slučajni vektor sa skupom vrijednosti $\mathcal{R}(X,Y) = \{(x_i,y_j) : i \in \mathbb{N}, j \in \mathbb{N}\}$ i distribucijom

$$p(x_i,y_j) = P\{X=x_i, Y=y_j\}, \quad i \in \mathbb{N}, \quad j \in \mathbb{N},$$

i neka je g neka funkcija koja uređenom paru realnih brojeva pridružuje realan broj, tj. $g : \mathbb{R}^2 \to \mathbb{R}$. Tada $g(X,Y)$ definira slučajnu varijablu na Ω. Na primjer,

$X+Y$, XY, $(X-\mu)(Y-\nu)$ (μ, ν jesu neki realni brojevi) jednodimenzionalne su slučajne varijable na Ω koje nastaju na takav način. Distribuciju tako nastale slučajne varijable možemo lako izračunati iz tablice distribucije slučajnog vektora (X,Y) i zadane funkcije g. Postupak je ilustriran primjerom 3.10.

Primjer 3.10. *Neka je dan pokus opisan u primjeru 3.6, tj. iz kutije koja sadrži tri kupona s oznakama 1, 2 i 3 izvlačimo dva kupona bez vraćanja u kutiju. Rezultat izvlačenja slučajan je vektor (X,Y), gdje je X rezultat prvog izvlačenja, a Y rezultat drugog izvlačenja. Odredimo distribuciju sume tih dviju slučajnih varijabli: $Z = X+Y$.*

$$X+Y = \begin{pmatrix} 2 & 3 & 4 & 5 & 6 \\ p_2 & p_3 & p_4 & p_5 & p_6 \end{pmatrix},$$

gdje je

$$\begin{aligned}
p_2 &= P\{Z=2\} = P\{X=1, Y=1\} = 0, \\
p_3 &= P\{Z=3\} = P\{X=1, Y=2\} + P\{X=2, Y=1\} = 1/3, \\
p_4 &= P\{Z=4\} = P\{X=1, Y=3\} + P\{X=3, Y=1\} + P\{X=2, Y=2\} = 1/3, \\
p_5 &= P\{Z=5\} = P\{X=3, Y=2\} + P\{X=2, Y=3\} = 1/3, \\
p_6 &= P\{Z=6\} = P\{X=3, Y=3\} = 0.
\end{aligned}$$

Dakle, tablica distribucije slučajne varijable jest

$$X+Y = \begin{pmatrix} 3 & 4 & 5 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

Odredite tablice distribucija slučajnog vektora (X,Y) i slučajne varijable $(X+Y)$ ako se izvlačenje vrši s vraćanjem.

Primjer 3.11. *Iz skupa $\{1,2,\ldots,9\}$ na slučajan način izvlačimo tri broja jedan po jedan (s vraćanjem izvučenih brojeva u skup). Prvi izvučeni broj bit će znamenka stotica, drugi izvučeni broj znamenka desetica, a treći izvučeni broj znamenka jedinica troznamenkastog broja. Izvlačenje znamenke stotica modeliramo slučajnom varijablom Z_1, znamenke desetica slučajnom varijablom Z_2, a znamenke jedinica slučajnom varijablom Z_3. Slučajna je varijabla Z_i, $i \in \{1,2,3\}$, diskretna slučajna varijabla koja se brojem iz skupa $\{1,2,\ldots,9\}$ realizira s vjerojatnošću $1/9$. Dakle, Z_1, Z_2 i Z_3 jesu diskretne uniformne slučajne varijable na skupu $\{1,2,\ldots,9\}$ i zadane su tablicom distribucije*

$$Z_i = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 1/9 & 1/9 & 1/9 & 1/9 & 1/9 & 1/9 & 1/9 & 1/9 & 1/9 \end{pmatrix}, \quad i \in \{1,2,3\}.$$

Uočimo da ima smisla promatrati trodimenzionalan diskretan slučajni vektor (Z_1, Z_2, Z_3) čija je slika $\mathcal{R}(Z_1,Z_2,Z_3) = \{(i,j,k),\ i,j,k \in \{1,2,\ldots,9\}\}$. Jedna realizacija slučajnog vektora (Z_1,Z_2,Z_3) jest varijacija s ponavljanjem trećeg razreda skupa od devet elemenata, stoga skup $\mathcal{R}(Z_1,Z_2,Z_3)$ ima $9^3 = 729$ elemenata. Kako su događaji $\{Z_1 = i, Z_2 = j, Z_3 = k\}$ za sve $i,j,k \in \{1,2,\ldots,9\}$ jednako vjerojatni, slijedi da je

$$P\{Z_1 = i, Z_2 = j, Z_3 = k\} = \frac{1}{729}, \quad (i,j,k) \in \mathcal{R}(Z_1,Z_2,Z_3).$$

Time je u potpunosti zadana distribucija trodimenzionalnog diskretnog slučajnog vektora (Z_1,Z_2,Z_3).

Nadalje, označimo s X slučajnu varijablu koja se realizira najvećom znamenkom, a s Y slučajnu varijablu koja se realizira najmanjom znamenkom u tako sastavljenom troznamenkastom broju:

$$X = \max\{Z_1,Z_2,Z_3\}, \quad Y = \min\{Z_1,Z_2,Z_3\}.$$

Budući da se izvlačenje brojeva iz skupa $\{1, 2, \ldots, 9\}$ vrši s vraćanjem, slijedi da se i X i Y realiziraju brojevima iz skupa $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, tj. $\mathcal{R}(X) = \mathcal{R}(Y)$. Distribucija slučajne varijable X dana je tablicom

$$X = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ \frac{1}{729} & \frac{7}{729} & \frac{19}{729} & \frac{36}{729} & \frac{62}{729} & \frac{91}{729} & \frac{127}{729} & \frac{169}{729} & \frac{217}{729} \end{pmatrix},$$

a distribucija slučajne varijable Y tablicom

$$Y = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ \frac{217}{729} & \frac{169}{729} & \frac{127}{729} & \frac{91}{729} & \frac{62}{729} & \frac{36}{729} & \frac{19}{729} & \frac{7}{729} & \frac{1}{729} \end{pmatrix}.$$

Uočimo da događaji $\{\omega \in \Omega : X(\omega) = x\} = \{X = x\}$, $x \in \mathcal{R}(X)$, i $\{\omega \in \Omega : Y(\omega) = y\} = \{Y = y\}$, $y \in \mathcal{R}(Y)$, nisu nezavisni. Na primjer, u slučaju troznamenkastog broja 111 i slučajna varijabla X i slučajna varijabla Y realiziraju se s 1. Slijedi da se u tom slučaju presjek događaja $\{X = 1\}$ i $\{Y = 1\}$ realizira s vjerojatnošću

$$P\{X = 1, Y = 1\} = \frac{1}{729}.$$

S druge strane, iz distribucija slučajnih varijabli X i Y znamo da je

$$P\{X = 1\} = \frac{1}{729}, \quad P\{Y = 1\} = \frac{217}{729}.$$

Dakle, ne vrijedi da je

$$P\{X = x, Y = y\} = P\{X = x\}P\{Y = y\}, \quad \forall (x, y) \in \mathcal{R}(X) \times \mathcal{R}(Y),$$

pa događaji $\{X = x\}$ i $\{Y = y\}$ nisu nezavisni, što otežava određivanje distribucije slučajnog vektora (X, Y).

Pretpostavimo da nas zanima razlika najveće i namanje znamenke takvog troznamenkastog broja. Ta je razlika modelirana slučajnom varijablom $(X-Y)$ koja prima vrijednosti iz skupa $\mathcal{R}(X-Y) = \{0, 1, \ldots, 8\}$. Da bismo odredili vjerojatnosti pojedinih realizacija slučajne varijable $(X - Y)$, koristimo distribuciju slučajnog vektora (X, Y):

$$P\{X - Y = n\} = \sum_{k=n+1}^{9} P\{X = k, Y = k - n\}, \quad n \in \mathcal{R}(X - Y).$$

Na primjer, iz gornje formule slijedi da je

$$P\{X - Y = 0\} = \sum_{k=1}^{9} P\{X = k, Y = k\} = \frac{9}{729}.$$

Važno je napomenuti da se očekivanje slučajne varijable $g(X, Y)$, nastale na temelju kompozicije realne funkcije g i diskretnog slučajnog vektora, može izračunati koristeći tablicu distribucije slučajnog vektora (X, Y) i oblik funkcije g bez računanja tablice distribucije slučajne varijable $g(X, Y)$ (naravno, ukoliko očekivanje postoji!). Ta tvrdnja može se lako provjeriti koristeći rezultate teorije ponovljenih redova (vidi poglavlje 5.3 u Dodatku) tj. činjenice da se kod apsolutno konvergentnih redova može mijenjati redoslijed sumacije bez utjecaja na konačan rezultat. Precizna formulacija te tvrdnje dana je teoremom 3.1.

Teorem 3.1. *Neka je $(\Omega, \mathcal{P}(\Omega), P)$ diskretan vjerojatnosni prostor, (X, Y) slučajni vektor na njemu dan distribucijom 3.2 i $g : \mathcal{R}(X, Y) \to \mathbb{R}$ zadana funkcija. Tada redovi*

$$Eg(X, Y) = \sum_{\omega \in \Omega} g(X, Y)(\omega)$$

i

$$\sum_i \sum_j g(x_i, y_j) p(x_i, y_j)$$

istovremeno ili apsolutno konvergiraju ili apsolutno divergiraju. U slučaju su im apsolutne konvergencije sume jednake.

Zadatak 3.2. Dokažite taj teorem!

Primjer 3.12. *Standardan svežanj od 52 karte sastoji se od 26 crvenih karata (13 hercova i 13 karo karata) i 26 crnih karata (13 pikova i 13 trefova). Iz takvog se svežnja na slučajan način jedna za drugom (bez vraćanja) izvlače dvije karte. Označimo s X slučajnu varijablu koja se realizira brojem izvučenih pikova, a s Y slučajnu varijablu koja se realizira brojem izvučenih trefova. Zaključujemo da je distribucija slučajnog vektora (X, Y) sljedeća:*

$$\mathcal{R}(X, Y) = \{(x, y) : x \in \{0, 1, 2\}, y \in \{0, 1, 2\}, x + y \leq 2\},$$

$$P\{X = x, Y = y\} = \frac{\binom{13}{x}\binom{13}{y}\binom{26}{2-x-y}}{\binom{52}{2}}.$$

Distribuciju slučajnog vektora (X, Y) i njegove marginalne distribucije pregledno prikazujemo sljedećom tablicom:

$X \setminus Y$	0	1	2	$p_X(x)$
0	25/102	13/51	1/17	19/34
1	13/51	13/102	0	13/34
2	1/17	0	0	1/17
$p_Y(y)$	19/34	13/34	1/17	1

Tablica 3.3: Tablica distribucije slučajnog vektora (X, Y) iz primjera 3.12.

Iz proširene tablice distribucije slučajnog vektora (X, Y) vidimo da su X i Y jednako distribuirane slučajne varijable.
Broj izvučenih crnih karata modeliramo slučajnom varijablom $(X + Y)$ čija je slika $\mathcal{R}(X + Y) = \{0, 1, 2\}$. Ako želimo izračunati očekivani broj izvučenih crnih karata, tj. očekivanje slučajne varijable $(X + Y)$, to možemo učiniti pomoću tablice distribucije slučajnog vektora (X, Y), dakle bez računanja distribucije slučajne varijable $(X + Y)$:

$$E(X + Y) = \sum_{x=0}^{2} \sum_{y=0}^{2} (x + y) P\{X = x, Y = y\} =$$

$$= 1 \cdot \frac{13}{51} + 2 \cdot \frac{1}{17} + 1 \cdot \frac{13}{51} + 2 \cdot \frac{13}{102} + 2 \cdot \frac{1}{17} = 1.$$

Radi provjere dobivenog rezultata odredimo ipak distribuciju slučajne varijable $(X+Y)$ i izračunajmo njezino očekivanje. Pomoću tablice distribucije slučajnog vektora (X,Y) računamo:

$$P\{X+Y=0\} = P\{X=0, Y=0\} = \frac{25}{102},$$

$$P\{X+Y=1\} = P\{X=1, Y=0\} + P\{X=0, Y=1\} = \frac{26}{51},$$

$$P\{X+Y=2\} = P\{X=2, Y=0\} + P\{X=1, Y=1\} + P\{X=0, Y=2\} = \frac{25}{102}.$$

Dakle, distribucija slučajne varijable $(X+Y)$ dana je tablicom

$$X+Y = \begin{pmatrix} 0 & 1 & 2 \\ 25/102 & 26/51 & 25/102 \end{pmatrix},$$

a njezino je očekivanje broj

$$E(X+Y) = 0 \cdot \frac{25}{102} + 1 \cdot \frac{26}{51} + 2 \cdot \frac{25}{102} = 1,$$

čime smo se uvjerili u točnost prethodnog rezultata.

3.1.1 Uvjetne distribucije

Ako želimo proučavati veze među komponentama slučajnog vektora, sama tablica distribucije nije praktična, a marginalne distribucije su nedovoljne. Bolji uvid dobije se analizom uvjetnih vjerojatnosti oblika $P(\{X \in A\}|\{Y=y\})$, gdje je A podskup od $\mathcal{R}(X)$ ili $P(\{Y \in A\}|\{X=x\})$, gdje je A podskup od $\mathcal{R}(Y)$. U tu svrhu definirat ćemo uvjetne distribucije slučajnog vektora. Radi jednostavnijeg zapisa koristit ćemo sljedeću oznaku:

$$P(\{X \in A\}|\{Y=y\}) = P\{X \in A|Y=y\}.$$

Primjer 3.13. *Tablicom distribucije 3.4 opisan je slučajni vektor (T,Z) kojim je modeliran nivo uspjeha pri polaganju teorijskog dijela (T) i zadataka (Z) iz jednog ispita matematičkog sadržaja.*

$T \setminus Z$	1	2	3	
1	$\frac{66}{100}$	$\frac{2}{100}$	$\frac{2}{100}$	$\frac{70}{100}$
2	$\frac{5}{100}$	$\frac{5}{100}$	$\frac{2}{100}$	$\frac{12}{100}$
3	$\frac{2}{100}$	$\frac{4}{100}$	$\frac{12}{100}$	$\frac{18}{100}$
	$\frac{73}{100}$	$\frac{11}{100}$	$\frac{16}{100}$	1

Tablica 3.4: Tablica distribucije slučajnog vektora iz primjera 3.13.

Koristeći pravila za računanje uvjetnih vjerojatnosti, lako možemo izračunati:

$$P\{T=1|Z=1\} = \frac{P\{T=1, Z=1\}}{P\{Z=1\}} = \frac{66}{73},$$

$$P\{T=2|Z=1\} = \frac{P\{T=2, Z=1\}}{P\{Z=1\}} = \frac{5}{73},$$

$$P\{T=3|Z=1\} = \frac{P\{T=3, Z=1\}}{P\{Z=1\}} = \frac{2}{73}.$$

Uočimo da je suma dobivenih triju brojeva jednaka 1 te da oni daju uvjetne vjerojatnosti da se postigne nivo uspjeha "i" iz teorije, $i \in \{1,2,3\}$, uz uvjet da je na zadacima postignut nivo 1. Na taj način određena je **uvjetna distribucija** *od T, uz uvjet da je $Z=1$, koju možemo prikazati i tablično:*

$$T_{|Z=1} = \begin{pmatrix} 1 & 2 & 3 \\ \frac{66}{73} & \frac{5}{73} & \frac{2}{73} \end{pmatrix}.$$

Analogno, možemo izračunati i prikazati uvjetnu distribuciju od T, uz uvjet da je $Z=2$, te uvjetnu distribuciju od T, uz uvjet da je $Z=3$, i komentirati sličnost, odnosno različitost tih distribucija:

$$T_{|Z=2} = \begin{pmatrix} 1 & 2 & 3 \\ \frac{2}{11} & \frac{5}{11} & \frac{4}{11} \end{pmatrix}, \qquad T_{|Z=3} = \begin{pmatrix} 1 & 2 & 3 \\ \frac{2}{16} & \frac{2}{16} & \frac{12}{16} \end{pmatrix}.$$

Iz navedenih uvjetnih distribucija može se vidjeti da je, uz uvjet $Z=1$, puno veća vjerojatnost da se na teoriji postigne nivo 1 nego da se postigne neki viši nivo uspjeha. Nasuprot tomu, uz uvjet da je $Z=3$ veća je vjerojatnost da se i na teoriji postigne nivo uspjeha 3 nego manji nivoi. Te činjenice daju naslutiti da se uvjetna distribucija od T mijenja s različitim vrijednostima od Z, tj. da između nivoa uspjeha postignutog na zadacima i teoriji postoji neka veza.

Opisujući numeričke karakteristike dobivenih uvjetnih distribucija, ta činjenica postaje još slikovitija. Tako za očekivanja vrijedi:

$$E(T|Z=1) = 1.12, \quad E(T|Z=2) = 2.18, \quad E(T|Z=3) = 2.63,$$

pa možemo reći da očekivanje nivoa uspjeha iz teorije raste s porastom nivoa uspjeha iz zadataka.

U općenitom slučaju pretpostavimo da je tablicom 3.5 dana distribucija dvodimenzionalnog slučajnog vektora (X,Y).

$X \setminus Y$	y_1	y_2	y_3	$\cdots$	p_X
x_1	$p(x_1,y_1)$	$p(x_1,y_2)$	$p(x_1 y_3)$	$\cdots$	$p_X(x_1)$
x_2	$p(x_2,y_1)$	$p(x_2,y_2)$	$p(x_2,y_3)$	$\cdots$	$p_X(x_2)$
$\vdots$					
p_Y	$p_Y(y_1)$	$p_Y(y_2)$	$p_Y(y_3)$	$\cdots$	1

Tablica 3.5: Proširena tablica distribucije dvodimenzionalnog slučajnog vektora.

Tada je tablicom

$$\begin{pmatrix} x_1 & x_2 & \cdots \\ p_{X|Y=y_j}(x_1) & p_{X|Y=y_j}(x_2) & \cdots \end{pmatrix}, \tag{3.5}$$

$$p_{X|Y=y_j}(x_i) = \frac{p(x_i,y_j)}{p_Y(y_j)}, \ i \in \{1,2,\ldots\}$$

dana **uvjetna distribucija komponente** X, uz uvjet da je $Y=y_j$.

Zadatak 3.3. Dokažite da je tablicom 3.5 dobro definirana tablica distribucije, tj. da vrijedi

$$p_{X|Y=y_j}(x_i) \geq 0, \ \forall i, \ \sum_i p_{X|Y=y_j}(x_i) = 1.$$

Na analogan način možemo definirati i komentirati uvjetne distribucije komponente Y, uz uvjet da je $X = x_i$ za svaki i.

Primjer 3.14. *Neka je (X, Y) slučajni vektor koji se realizira uređenim parom brojeva u kojemu je prva komponenta broj pikova, a druga komponenta broj trefova izvučenih pri slučajnom odabiru (bez vraćanja) dviju karata iz standardnog svežnja od 52 karte. Distribucija slučajnog vektora (X, Y) dana je tablicom 3.3 u primjeru 3.12. Odredimo sada uvjetne distribucije i uvjetna očekivanja komponente X, uz uvjet $Y = y$ za sve $y \in \{0, 1, 2\}$, te komponente Y, uz uvjet $X = x$ za sve $x \in \{0, 1, 2\}$. Usmjerimo se prvo na uvjetnu distribuciju od X, uz uvjet da je $Y = y$.*

Za $y = 0$ je

$$P\{X = 0|Y = 0\} = \frac{P\{X = 0, Y = 0\}}{P\{Y = 0\}} = \frac{25}{57},$$
$$P\{X = 1|Y = 0\} = \frac{P\{X = 1, Y = 0\}}{P\{Y = 0\}} = \frac{26}{57},$$
$$P\{X = 2|Y = 0\} = \frac{P\{X = 2, Y = 0\}}{P\{Y = 0\}} = \frac{2}{19}.$$

Za $y = 1$ je $P\{X = 0|Y = 1\} = \frac{2}{3}$, $P\{X = 1|Y = 1\} = \frac{1}{3}$ i $P\{X = 2|Y = 1\} = 0$.

Za $y = 2$ je $P\{X = 0|Y = 2\} = 1$, $P\{X = 1|Y = 2\} = 0$ i $P\{X = 2|Y = 2\} = 0$.

Slijedi da su uvjetne distribucije redom dane tablicama

$$X_{|Y=0} = \begin{pmatrix} 0 & 1 & 2 \\ 25/57 & 26/57 & 2/19 \end{pmatrix}, \qquad X_{|Y=1} = \begin{pmatrix} 0 & 1 \\ 2/3 & 1/3 \end{pmatrix},$$

dok je vjerojatnost da se X realizira nulom, uz uvjet da je $Y = 2$, jednaka jedan. Očekivanja su tih uvjetnih distribucija sljedeća:

$$E(X|Y = 0) = \frac{38}{57}, \qquad E(X|Y = 1) = \frac{1}{3}, \qquad E(X|Y = 2) = 1.$$

Analognim postupkom dobivamo uvjetne distribucije i uvjetna očekivanja komponente Y, uz uvjet $X = x$ za sve $x \in \{0, 1, 2\}$. Uočimo da je u ovom primjeru uvjetna distribucija od X, uz uvjet $Y = i$, jednaka uvjetnoj distribuciji od Y, uz uvjet da je $X = i$, za sve $i \in \{0, 1, 2\}$.

Primjer 3.15. *U primjeru 3.9 pokazali smo da su marginalne distribucije slučajnog vektora $(X, Y) \sim MULT(n, p_1, p_6)$ binomne, tj. da je $X \sim \mathcal{B}(n, p_1)$, a $Y \sim \mathcal{B}(n, p_6)$. Odredimo sada uvjetne distribucije tih komponenti tog slučajnog vektora. Za $x \in \mathcal{R}(X)$ je*

$$P\{Y = y|X = x\} = \frac{P\{X = x, Y = y\}}{P\{X = x\}} = \binom{n-x}{y} \left(\frac{p_6}{1-p_1}\right)^y \left(1 - \frac{p_6}{1-p_1}\right)^{n-x-y}.$$

Slijedi da je uvjetna distribucija od Y, uz uvjet $X = x$, binomna s parametrima $(n-x)$ i $p_6/(1-p_1)$. Analogno slijedi da je uvjetna distribucija od X, uz uvjet $Y = y$, binomna s parametrima $(n - y)$ i $p_1/(1 - p_6)$. Budući da je matematičko očekivanje binomne slučajne varijable jednako umnošku njezinih parametara, slijedi da je

$$E(Y|X = x) = \frac{p_6(n - x)}{1 - p_1}, \quad x \in \mathcal{R}(X),$$
$$E(X|Y = y) = \frac{p_1(n - y)}{1 - p_6}, \quad y \in \mathcal{R}(Y).$$

3.1.2 Nezavisnost

Ukoliko za dani diskretan slučajni vektor (X, Y) uvjetne distribucije komponente Y, uz uvjet $X = x_i$, ostaju stalno iste pri promjeni x_i, $i \in \mathbb{N}$, to daje naslutiti da su komponente tog slučajnog vektora nezavisne. Međutim, nezavisnost slučajnih varijabli definirat ćemo koristeći definiciju nezavisnosti događaja. Također ćemo pokazati da je navedena intuitivna činjenica o nezavisnosti komponenti danog slučajnog vektora u suglasnosti s definicijom nezavisnosti slučajnih varijabli.

Za događaje A i B danog vjerojatnosnog prostora $(\Omega, \mathcal{F}, P)$ rekli smo da su nezavisni ako vrijedi $P(A \cap B) = P(A)P(B)$ (vidi poglavlje 1.9). Ako pretpostavimo da takav tip jednakosti vrijedi za sve događaje inducirane diskretnim slučajnim varijablama X i Y na vjerojatnosnom prostoru $(\Omega, \mathcal{P}(\Omega), P)$, tj. ako je

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}$$

za sve $A \subseteq \mathcal{R}(X)$, $B \subseteq \mathcal{R}(Y)$, onda ćemo reći da su slučajne varijable X i Y nezavisne.

Definicija 3.3. *Za diskretne slučajne varijable X i Y definirane na istom diskretnom vjerojatnosno prostoru $(\Omega, \mathcal{P}(\Omega), P)$ kažemo da su* **nezavisne** *ako za sve skupove $A \subseteq \mathbb{R}$ i $B \subseteq \mathbb{R}$ vrijedi*

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}.$$

Teorem 3.2. *Neka je (X, Y) diskretan slučajni vektor na vjerojatnosnom prostoru $(\Omega, \mathcal{P}(\Omega), P)$ čija je distribucija dana tablicom 3.5, tj.*

$X \setminus Y$	y_1	y_2	y_3	$\cdots$	p_X
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	$p(x_1 y_3)$	$\cdots$	$p_X(x_1)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	$p(x_2, y_3)$	$\cdots$	$p_X(x_2)$
$\vdots$					
p_Y	$p_Y(y_1)$	$p_Y(y_2)$	$p_Y(y_3)$	$\cdots$	*1*

Slučajne varijable X i Y jesu nezavisne onda i samo onda ako vrijedi

$$p(x_i, y_j) = p_X(x_i)p_Y(y_j)$$

za sve $x_i \in \mathcal{R}(X)$, $y_j \in \mathcal{R}(Y)$.

Dokaz. Pretpostavimo da su X i Y nezavisne slučajne varijable. Za proizvoljne $x_i \in \mathcal{R}(X)$, $y_j \in \mathcal{R}(Y)$ je $\{x_i\} \subseteq \mathbb{R}$, $\{y_j\} \subseteq \mathbb{R}$, pa je

$$p(x_i, y_j) = P\{X = x_i, Y = y_j\} = P\{X = x_i\}P\{Y = y_j\} = p_X(x_i)p_Y(y_j),$$

tj. vrijedi tvrdnja teorema.

Obratno, pretpostavimo da je $p(x_i, y_j) = P\{X = x_i\}P\{Y = y_j\}$ za sve $x_i \in \mathcal{R}(X)$, $y_j \in \mathcal{R}(Y)$ i neka su $A \subseteq \mathbb{R}$, $B \subseteq \mathbb{R}$ proizvolji skupovi. Tada vrijedi:

$$\begin{aligned} P\{X \in A, Y \in B\} &= \sum_{\{X(\omega)\in A, Y(\omega)\in B\}} P\{\omega\} = \\ &= \sum_{x_i\in A}\sum_{y_j\in B} p(x_i, y_j) = \sum_{x_i\in A}\sum_{y_j\in B} P\{X = x_i\}P\{Y = y_j\} = \\ &= \sum_{x_i\in A} P\{X = x_i\} \sum_{y_j\in B} P\{Y = y_j\} = \\ &= P\{X \in A\}P\{Y \in B\}. \end{aligned}$$

Primjer 3.16. *Iz kutije koja sadrži tri kupona s oznakama 1, 2 i 3 izvlačimo dva kupona tako da prvi izvučeni kupon vratimo u kutiju prije izvlačenja drugog kupona (izvlačenje s vraćanjem). Rezultat izvlačenja modeliran je slučajnim vektorom (X, Y), gdje je X rezultat prvog izvlačenja, a Y razultat drugog izvlačenja. Distribucija tog slučajnog vektora dana je sljedećom tablicom:*

$X \setminus Y$	1	2	3	p_X
1	1/9	1/9	1/9	1/3
2	1/9	1/9	1/9	1/3
3	1/9	1/9	1/9	1/3
p_Y	1/3	1/3	1/3	1

Koristeći teorem 3.2 vidimo da su X i Y nezavisne. Također možemo uočiti da su uvjetne distribucije komponente X, uz dano $Y = i$, jednake za sve $i \in \{1, 2, 3\}$ i da su jednake marginalnim distribucijama komponente X. Analogna tvrdnja vrijedi i za uvjetne distribucije komponente Y, uz dano $X = i$, $i \in \{1, 2, 3\}$.

Promotrimo sada izvlačenje dvaju kupona, ali bez vraćanja prvog izvučenog kupona u kutiju. Rezultat izvlačenja modeliran je slučajnim vektorom (U, V), gdje je U rezultat prvog izvlačenja, a V razultat drugog izvlačenja. Distribucija slučajnog vektora (U, V) dana je tablicom

$U \setminus V$	1	2	3	p_X
1	0	1/6	1/6	1/3
2	1/6	0	1/6	1/3
3	1/6	1/6	0	1/3
p_Y	1/3	1/3	1/3	1

Koristeći teorem 3.2 vidimo da U i V nisu nezavisne. Također uočimo da se marginalne distribucije slučajnih vektora (X, Y) i (U, V) podudaraju.

Općenito su kod nezavisnih komponenti diskretnog slučajnog vektora (X, Y) uvjetne distribucije od X, uz dano $Y = y_j$, iste za sve $y_j \in \mathcal{R}(Y)$ i jednake marginalnoj distribuciji komponente X. Slično vrijedi i za uvjetne distribucije komponente Y, uz dano $X = x_i$, $x_i \in \mathcal{R}(X)$.

Zaista, neka je (X, Y) slučajni vektor s tablicom distribucije 3.1 i nezavisnim komponentama X i Y. Tada je

$$p_{X|Y=y_j}(x_i) = \frac{p(x_i, y_j)}{p_Y(y_j)} = \frac{p_X(x_i)p_Y(y_j)}{p_Y(y_j)} = p_X(x_i), \quad i \in \{1, 2, \ldots\}.$$

Sljedeći teorem daje vrlo važno svojstvo nezavisnih slučajnih varijabli.

Teorem 3.3. *Neka su X i Y nezavisne slučajne varijable takve da postoji $E(X)$ i $E(Y)$. Tada postoji $E(X \cdot Y)$ i vrijedi:*

$$E(XY) = E(X)E(Y).$$

Dokaz. Koristeći poznate rezultate o sumi redova[3] kao i rezultate teorije ponovljenih redova (vidi poglavlje 5.3 u Dodatku), znamo da konvergencija redova

$$\sum_i |x_i| p_X(x_i), \quad \sum_j |y_j| p_Y(y_j)$$

povlači konvergenciju reda

$$\sum_i \sum_j |x_i y_j| p_X(x_i) p_Y(y_j).$$

Osim toga, vrijedi:

$$\begin{aligned} \sum_i \sum_j x_i y_j p(x_i, y_j) &= \sum_i \sum_j x_i y_j p_X(x_i) p_Y(y_j) = \\ &= \sum_i x_i p_X(x_i) \sum_j y_j p_Y(y_j) = \\ &= E(X)\, E(Y). \end{aligned}$$

Koristeći teorem 3.1 možemo zaključiti da $E(XY)$ postoji i da vrijedi tvrdnja teorema.

[3] Vidi npr. [13].

3.2 Neprekidan dvodimenzionalan slučajni vektor

U neprekidnom slučaju dvodimenzionalan slučajni vektor prima vrijednosti u skupu $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ ali, za razliku od diskretnog slučajnog vektora, njegov skup vrijednosti sadrži neko područje iz $\mathbb{R}^2$.

Definicija 3.4. *Reći ćemo da je (X, Y) dvodimenzionalan neprekidan slučajni vektor ako postoji nenegativna realna funkcija $f : \mathbb{R}^2 \to \mathbb{R}$, takva da se funkcija distribucije od (X, Y) može napisati u obliku*

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\, du\, dv. \tag{3.6}$$

Takvu funkciju f nazivamo **funkcija gustoće** *neprekidnog slučajnog vektora.*

Vidimo da su, zbog jednakosti (3.6), vjerojatnosna svojstva neprekidnog slučajnog vektora u potpunosti određena njegovom funkcijom gustoće.

Primjer 3.17. *Slučajni vektor uniformne distribucije na krugu $\{(x, y) : x^2 + y^2 < r^2\}$ dan je funkcijom gustoće*

$$f(x,y) = \begin{cases} \frac{1}{r^2\pi} & , \; x^2+y^2<r^2 \\ 0 & , \; \text{inače} \end{cases}.$$

Primjer 3.18. *Dvodimenzionalan normalan slučajni vektor dan je funkcijom gustoće $f : \mathbb{R}^2 \to \mathbb{R}$ koja je definirana pravilom*

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}, \tag{3.7}$$

gdje su $\mu_1 \in \mathbb{R}, \mu_2 \in \mathbb{R}, \sigma_1 > 0, \sigma_2 > 0, \rho \in \langle -1, 1\rangle$. U tom slučaju koristimo oznaku $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. [4]

Ako je (X, Y) neprekidan slučajni vektor s funkcijom gustoće $f(x, y)$ onda će njegove marginalne distribucije također biti neprekidnog tipa s gustoćama koje možemo izračunati iz $f(x, y)$.

Zaista, neka je (X, Y) dvodimenzionalan slučajni vektor s gustoćom $f(x, y)$. Vrijedi:

$$P\{X \le x\} = P\{X \le x, Y < \infty\} = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u,v)\, du\, dv =$$

$$= \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} f(u,v)\, dv \right) du.$$

[4] Vidi Sarapa [26, str. 274].

Definiramo li

$$f_X(x) = \int_{-\infty}^{\infty} f(x,v)\,dv\,,$$

onda je f_X realna funkcija realne varijable koja ima potrebna svojstva da bi bila gustoća slučajne varijable X, tj. ona je nenegativna i za svaki $x \in \mathbb{R}$ vrijedi:

$$P\{X \le x\} = \int_{-\infty}^{x} f_X(u)\,du.$$

Funkciju f_X zovemo **marginalna funkcija gustoće** komponente X slučajnog vektora (X, Y). Analogno je

$$f_Y(y) = \int_{-\infty}^{\infty} f(u,y)\,du\,,$$

marginalna funkcija gustoće komponente Y slučajnog vektora (X, Y).

Primjer 3.19. *Odredimo marginalne distribucije dvodimenzionalnog normalnog slučajnog vektora $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ zadanog funkcijom gustoće (3.7). Funkciju gustoće slučajne varijable X računamo na sljedeći način:*

$$f_X(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho)^2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho)^2}\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}-2\rho\frac{x-\mu_1}{\sigma_1}\right)}\,dy =$$

$$\left| \begin{array}{c} \frac{y-\mu_2}{\sigma_2} - 2\rho\frac{x-\mu_1}{\sigma_1} = z \;\Rightarrow\; \frac{y-\mu_2}{\sigma_2} = z + 2\rho\frac{x-\mu_1}{\sigma_1} \\ dy = \sigma_2\,dz \end{array} \right|$$

$$= \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}\; e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \frac{1}{2(1-\rho^2)}\rho^2\left(\frac{x-\mu_1}{\sigma_1}\right)^2} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2(1-\rho^2)}}\,dz =$$

$$\left| \frac{z}{\sqrt{1-\rho^2}} = t,\; dz = \sqrt{1-\rho^2}\,dt \right|$$

$$= \frac{1}{2\pi\sigma_1}\; e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\,dt = \frac{1}{2\pi\sigma_1}\; e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\,\sqrt{2\pi} = \frac{1}{\sigma_1\sqrt{2\pi}}\; e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}.$$

Uočimo da je

$$f_X(x) = \frac{1}{\sigma_1\sqrt{2\pi}}\, e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, \quad x \in \mathbb{R} \tag{3.8}$$

funkcija gustoće normalne distribucije s parametrima $\mu_1 \in \mathbb{R}$ i σ_1^2, $\sigma_1 > 0$, tj. $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$. Analognim postupkom slijedi da je $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Dakle, marginalne distribucije dvodimenzionalnog normalnog slučajnog vektora su normalne.

Neka je (X, Y) neprekidan slučajni vektor s funkcijom gustoće f i neka je g neka funkcija koja uređenom paru realnih brojeva pridružuje realan broj, tj. $g : \mathbb{R}^2 \to \mathbb{R}$. Može se dokazati (vidi npr. [26]) da u mnogo zanimljivih slučajeva (npr. ako je g neprekidna) $g(X, Y)$ definira slučajnu varijablu na Ω.

Može se također pokazati da za računanje očekivanja tako nastalih slučajnih varijabli nije potrebno tražiti funkciju distribucije od $g(X,Y)$, već se može iskoristiti funkcija gustoće od (X,Y) i oblik funkcije g (ako očekivanje uopće postoji) na sljedeći način:

$$Eg(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\,dx\,dy. \tag{3.9}$$

Na primjer, komponiranjem projekcije uređenog para (x,y) na prvu koordinatnu os

$$\pi_1(x,y) = x$$

i slučajnog vektora (X,Y) dobijemo komponentu X, pa očekivanje EX (ako postoji) možemo izračunati pomoću $E(\pi_1(X,Y))$, tj.

$$E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,f(x,y)\,dx\,dy = \int_{-\infty}^{\infty} xf_X(x)\,dx, \tag{3.10}$$

gdje je f funkcija gustoće slučajnog vektora (X,Y), a f_X marginalna gustoća komponente X.

Analogno je

$$E(Y) = \int_{-\infty}^{\infty} yf_Y(y)\,dy.$$

Primjer 3.20. *Neka je neprekidan dvodimenzionalan slučajni vektor (X,Y) zadan funkcijom gustoće*

$$f(x,y) = \begin{cases} \lambda_1\lambda_2 e^{-\lambda_1 x-\lambda_2 y} & , \ (x,y)\in\langle 0,\infty\rangle\times\langle 0,\infty\rangle \\ 0 & , \ \textit{inače} \end{cases},$$

gdje su λ_1 i λ_2 pozitivni realni brojevi. Slučajna varijabla XY nastaje kompozicijom slučajnog vektora $(X,Y)\colon \Omega\to\mathbb{R}^2$ i funkcije $g\colon\mathbb{R}^2\to\mathbb{R}$ definirane pravilom $g(x,y)=xy$. Matematičko očekivanje slučajne varijable XY računamo na sljedeći način:

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\,dx\,dy = \lambda_1\lambda_2\int_0^{\infty}\int_0^{\infty} xye^{-\lambda_1 x-\lambda_1 y}\,dx\,dy = \frac{\lambda_1\lambda_2}{\lambda_1^2\lambda_2^2} = \frac{1}{\lambda_1\lambda_2}.$$

Odredimo i marginalne distribucije te očekivanja marginalnih distribucija slučajnog vektora (X,Y). Funkciju gustoće slučajne varijable X računamo na sljedeći način:

$$f_X(x) = \lambda_1\lambda_2\int_0^{\infty} e^{-\lambda_1 x-\lambda_2 y}\,dy = \ldots = \lambda_1 e^{-\lambda_1 x}, \quad x>0.$$

Dakle, X ima eksponencijalnu distribuciju s parametrom $\lambda_1>0$. Slijedi da je $EX=1/\lambda_1$. Analognim postupkom slijedi da Y ima eksponencijalnu distribuciju s parametrom $\lambda_2>0$, pa je $EY=1/\lambda_2$.

3.2.1 Uvjetne gustoće. Nezavisnost

U ovom ćemo poglavlju ćemo bez dokaza navesti nekoliko rezultata korisnih za statističke primjene po analogiji s diskretnim slučajem. Tu se pojavljuje funkcija gustoće slučajnog vektora kao nositelj vjerojatnosnih svojstava umjesto tablice distribucije diskretnoga slučajnog vektora.

Neka je (X, Y) neprekidan slučajni vektor s gustoćom $f(x, y)$ i y takav realan broj da je $f_Y(y) > 0$. Funkciju

$$f_{X|Y=y}(x) = \frac{f(x, y)}{f_Y(y)} \tag{3.11}$$

zovemo **uvjetna gustoća** od X uz uvjet $Y = y$. Analogno je

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)} \tag{3.12}$$

uvjetna gustoća od Y uz uvjet $X = x$. Ovdje su f_Y i f_X pripadne marginalne gustoće.

Veličine definirane formulom

$$E(X \mid Y = y) = \int_{-\infty}^{\infty} x\, f_{X|Y=y}(x)\, dx, \tag{3.13}$$

odnosno

$$E(Y \mid X = x) = \int_{-\infty}^{\infty} y\, f_{Y|X=x}(y)\, dy \tag{3.14}$$

zovemo **uvjetno očekivanje** komponente X, uz uvjet $Y = y$, odnosno uvjetno očekivanje komponente Y, uz uvjet $X = x$.

Primjer 3.21. *Odredimo uvjetne distribucije dvodimenzionalnoga normalnog slučajnog vektora $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ zadanog funkcijom gustoće (3.7). Odredimo prvo uvjetnu distribuciju komponente X, uz uvjet $Y = y$, $y \in \mathbb{R}$. Funkciju gustoće uvjetne distrbiucije od X, uz uvjet $Y = y$, računamo pomoću funkcije gustoće $f(x, y)$ slučajnog vektora (X, Y) i funkcije gustoće $f_Y(y)$ komponenete $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$:*

$$f_{X|Y=y}(x) = \frac{f(x, y)}{f_Y(y)} = \frac{1}{\sigma_1\sqrt{2\pi(1-\rho^2)}}\, e^{-\frac{\left(x-\left(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)\right)\right)^2}{2\sigma_1^2(1-\rho^2)}}, \quad x \in \mathbb{R}.$$

Uočimo da je $f_{X|Y=y}(x)$ funkcija gustoće normalne distribucije s parametrima

$$\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2) \quad i \quad \sigma_1^2(1 - \rho^2).$$

Iz tog rezultata odmah slijedi da su uvjetno očekivanje i varijanca slučajne varijable X, uz uvjet $Y = y$, dani sljedećim izrazima:

$$E(X|Y=y) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2), \qquad Var(X|Y=y) = \sigma_1^2(1-\rho^2).$$

Analogan rezultat vrijedi za uvjetnu distribuciju komponente Y, uz uvjet $X = x$:

$$Y_{|X=x} \sim \mathcal{N}\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1), \sigma_2^2(1-\rho^2)\right),$$

$$E(Y|X=x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1), \qquad Var(Y|X=x) = \sigma_2^2(1-\rho^2).$$

Dakle, i uvjetne distribucije normalnog slučajnog vektora su normalne.

U poglavlju o nezavisnosti komponenata dvodimenzionalnog slučajnog vektora vidjeli smo da se nezavisnost može karakterizirati činjenicom da se distribucija slučajnog vektora može dobiti kao produkt marginalnih distribucija. Problemom nezavisnosti komponenti neprekidnog slučajnog vektora nećemo se baviti detaljno. Navodimo samo da je u tom slučaju nezavisnost ekvivalentna činjenici da se funkcija gustoće slučajnog vektora može prikazati kao produkt marginalnih gustoća kao i činjenici da se funkcija distribucije slučajnog vektora može prikazati kao produkt marginalnih funkcija distribucije:

Neka je (X, Y) neprekidan slučajni vektor s funkcijom gustoće f i funkcijom distribucije F. Neka su f_X (F_X) odnosno f_Y (F_Y) njegove marginalne gustoće (funkcije distribucije). Reći ćemo da su slučajne varijable X i Y nezavisne ako za svaki $(x, y) \in \mathbb{R}^2$ vrijedi

$$F(x, y) = F_X(x)\, F_Y(y).$$

Tada je i

$$f(x, y) = f_X(x)\, f_Y(y)$$

za svaki $(x, y) \in \mathbb{R}^2$.
Obratno, ako prethodna jednakost za gustoće vrijedi za sve $(x, y) \in \mathbb{R}^2$ osim eventualno na skupu površine nula, tada su X i Y nezavisne slučajne varijable.

Sada je lako vidjeti da se i ovdje, u slučaju nezavisnih komponenti, uvjetne gustoće podudaraju s odgovarajućim marginalnim gustoćama.

Primjer 3.22. *Promotrimo ponovno slučajni vektor zadan funkcijom gustoće*

$$f(x, y) = \begin{cases} \lambda_1\lambda_2 e^{-\lambda_1 x - \lambda_2 y} & , (x, y) \in \langle 0, \infty\rangle \times \langle 0, \infty\rangle \\ 0 & , \quad \text{inače} \end{cases},$$

gdje su λ_1 i λ_2 pozitivni realni brojevi. U primjeru 3.20 pokazali smo da su njegove marginalne distribucije eksponencijalne, tj. da je X eksponencijalna slučajna varijabla s parametrom $\lambda_1 > 0$, a Y eksponencijalna slučajna varijabla s parametrom $\lambda_2 > 0$:

$$f_X(x) = \begin{cases} \lambda_1 e^{-\lambda_1 x} & , \ x > 0 \\ 0 & , \ x \leq 0 \end{cases}, \qquad f_Y(y) = \begin{cases} \lambda_2 e^{-\lambda_2 y} & , \ y > 0 \\ 0 & , \ y \leq 0 \end{cases}.$$

Uočimo da je u tom slučaju

$$f(x,y) = f_X(x) f_Y(y)$$

pa su slučajne varijable X i Y nezavisne. Zbog nezavisnosti slučajnih varijabli X i Y vrijede sljedeće tvrdnje.

Marginalna distribucija komponente X i uvjetna distribucija komponente X, uz uvjet $Y = y$ (za bilo koji $y > 0$), jednake su - eksponencijalne s parametrom λ_1.

Marginalna distribucija komponente Y i uvjetna distribucija komponente Y, uz uvjet $X = x$ (za bilo koji $x > 0$), jednake su - eksponencijalne s parametrom λ_2.

Primjer 3.23. *Na temelju primjera 3.19 u kojemu su određene marginalne distribucije dvodimenzionalnoga normalnog slučajnog vektora (X, Y) s funkcijom gustoće (3.7) i karakterizacije nezavisnosti slučajnih varijabli, slijedi da komponente $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ i $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ normalnog slučajnog vektora (X, Y) općenito nisu nezavisne. Međutim, za $\rho = 0$ funkcija gustoće (3.7) dana je izrazom*

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right)}, \quad (x,y) \in \mathbb{R}^2.$$

Uočimo da je u tom slučaju $f(x,y) = f_X(x) f_Y(y)$, tj. ako je $\rho = 0$, komponente normalnoga slučajnog vektora (X, Y) nezavisne su.

Svojstvo nezavisnih slučajnih varijabli, da se očekivanje produkta može izračunati kao produkt očekivanja, vrijedi i ovdje. Naime imamo:

Teorem 3.4. *Neka je (X, Y) neprekidan slučajni vektor takav da su X i Y nezavisne slučajne varijable za koje postoji EX i EY. Tada postoji $E(X\,Y)$ i vrijedi:*[5]

$$E(X\,Y) = E(X)\,E(Y).$$

Primjer 3.24. *U primjeru 3.22 pokazali smo da su komponente X i Y slučajnog vektora (X, Y) nezavisne i eksponencijalno distribuirane s parametrima $\lambda_1 > 0$ i $\lambda_2 > 0$, redom. Budući da je $EX = 1/\lambda_1$, a $EY = 1/\lambda_2$, prema teoremu 3.4 slijedi da je*

$$E(XY) = EX\,EY = \frac{1}{\lambda_1 \lambda_2}.$$

Sjetimo se da smo $E(XY)$ u primjeru 3.20 izračunali pomoću funkcije gustoće slučajnog vektora (X, Y).

[5] Taj teorem nećemo dokazivati, a zainteresirani čitatelj dokaz može naći u [26, teorem 11.5]

3.3 Kovarijanca i koeficijent korelacije

U ovom poglavlju bavimo se definiranjem osnovnih numeričkih katakteristika slučajnog vektora - momentima. Rezultati se odnose na diskretan i neprekidan dvodimenzionalan slučajni vektor za kojega smo opisali način računanja očekivanja slučajne varijable $g(X, Y)$, gdje je g realna funkcija dviju varijabli.

Momenti slučajnog vektora poslužit će za opisivanje nekih svojstava slučajnog vektora slično kao što momenti slučajne varijable opisuju slučajnu varijablu.

Definicija 3.5. *Neka je (X, Y) diskretan ili neprekidan dvodimenzionalan slučajni vektor. Očekivanje*

$$E(X^k Y^l), \quad k, l \in \mathbb{N}_0$$

slučajne varijable $X^k Y^l$ (ako postoji) nazivamo **ishodišni moment reda** (k, l) *slučajnog vektora (X, Y) i pišemo*

$$\mu_{kl} = E(X^k Y^l). \tag{3.15}$$

Očekivanje $E\left((X - EX)^k (Y - EY)^l\right)$ (ako postoji) nazivamo **centralni moment reda** (k, l) *slučajnog vektora (X, Y) i pišemo*

$$m_{kl} = E\left((X - EX)^k (Y - EY)^l\right). \tag{3.16}$$

Uočimo da momenti reda $(0, 1)$, $(0, 2)$, $(1, 0)$, $(2, 0)$ daju ustvari očekivanje i varijancu komponenata slučajnog vektora. Naime, vrijedi:

$$\begin{aligned}
&\mu_{10} = EX, \\
&\mu_{01} = EY, \\
&m_{20} = E\,(X - EX)^2 = \operatorname{Var} X, \\
&m_{02} = E\,(Y - EY)^2 = \operatorname{Var} Y.
\end{aligned}$$

Od posebne je važnosti za izučavanje slučajnih vektora centralni moment reda $(1, 1)$ kojega nazivamo **korelacijski moment** ili **kovarijanca** dvodimenzionalnoga slučajnog vektora. Dakle, kovarijanca dvodimenzionalnog slučajnog vektora (X, Y) definirana je izrazom

$$\operatorname{Cov}(X, Y) = E\left((X - EX)(Y - EY)\right). \tag{3.17}$$

Definicijski oblik kovarijance može se lako, korištenjem linearnosti očekivanja, transformirati u oblik koji je nekada prikladniji za računanje:

$$\operatorname{Cov}(X, Y) = E\,(XY - X\,EY) - Y\,EX + EX\,EY) = E(XY) - EX\,EY.$$

Primjer 3.25. *Prisjetimo se primjera 3.16 koji je govorio o izvlačnju dvaju kupona iz kutije koja sadrži ukupno tri kupona označena brojevima* 1, 2 *i* 3. *U tom smo primjeru promatrali izvlačenje dvaju od triju kupona na sljedeća dva načina:*

prvi izvučeni kupon vraća se u kutiju prije izvlačenja drugog kupona; S X smo označili slučajnu varijablu koja se realizira brojem kojim je označen prvi izvučeni kupon, a s Y slučajnu varijablu koja se realizira brojem kojim je označen drugi izvučeni kupon,

prvi izvučeni kupon ne vraća se u kutiju prije izvlačenja drugog kupona; S U smo označili slučajnu varijablu koja se realizira brojem kojim je označen prvi izvučeni kupon, a s V slučajnu varijablu koja se realizira brojem kojim je označen drugi izvučeni kupon.

Sjetimo se da su slučajne varijable X i Y nezavisne, dok slučajne varijable U i V nisu nezavisne. Distribucije slučajnih vektora (X, Y) i (U, V) dane su tablicama u primjeru 3.16. X, Y, U i V sve su jednako distribuirane s matematičkim očekivanjem

$$1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3} = 2.$$

Odredimo sada $\mathrm{Cov}\,(X, Y)$ *i* $\mathrm{Cov}\,(U, V)$*:*

- *budući da su X i Y nezavisne, znamo da je $E(XY) = EX\,EY$ pa slijedi da je* $\mathrm{Cov}\,(X, Y) = E(XY) - EX\,EY = 0$.
- *kako U i V nisu nezavisne, problem izračunavanja njihove kovarijance svodi se na problem izračunavanja očekivanja $E(UV)$:*

$$\begin{aligned} E(UV) &= \sum_{i=1}^{2}\sum_{j=1}^{2} ijP\{X = i, Y = j\} = \\ &= 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \\ &= \frac{11}{3}. \end{aligned}$$

Sada slijedi da je

$$\mathrm{Cov}\,(U, V) = E(UV) - EU\,EV = \frac{11}{3} - 4 = -\frac{1}{3}.$$

Primjer 3.26. *Odredimo kovarijancu komponenti dvodimenzionalnoga multinomnog slučajnog vektora $(X, Y) \sim MULT(n, p, q)$ čija je distribucija dana u primjeru 3.9. Budući da su marginalne distribicije toga slučajnog vektora binome, tj. $X \sim \mathcal{B}(n, p)$ i $Y \sim \mathcal{B}(n, q)$, zaključujemo da postoje $x \in \mathcal{R}(X)$ i $y \in \mathcal{R}(Y)$ takvi da je $P\{X = x, Y = y\} \neq P\{X = x\}P\{Y = y\}$, pa slučajne varijable X i Y nisu nezavisne. Također znamo da je $EX = np$ i $EY = nq$. Da bismo izračunali kovarijancu slučajnih varijabli X i Y, preostaje još izračunati $E(XY)$:*

$$E(XY) = \sum_{x=0}^{n}\sum_{y=0}^{n} xyP\{X = x, Y = y\} = \sum_{x=0}^{n}\sum_{y=0}^{n-x} xyP\{X = x, Y = y\} = n(n-1)pq.$$

Sada slijedi da je

$$\mathrm{Cov}\,(X, Y) = E(XY) - EX\,EY = n(n-1)pq - n^2pq = -npq.$$

Važnost tog momenta za opisivanje slučajnog vektora proizlazi iz činjenice da se kovarijanca, kao numerička karakteristika slučajnog vektora, može povezati s pojmom nezavisnosti slučajnih varijabli. Naime, vrijedi sljedeći teorem.

Teorem 3.5. *Neka je* (X, Y) *neprekidan ili diskretan dvodimenzionalan slučajni vektor za koji postoje* EX *i* EY*. Ako su slučajne varijable* X *i* Y *nezavisne, onda je* $\text{Cov}\,(X, Y) = 0$.

Dokaz. Neka su ispunjeni uvjeti teorema. Zbog nezavisnosti je

$$E(XY) = EX\,EY.$$

Dakle, vrijedi:

$$\text{Cov}\,(X, Y) = E(XY) - EX\,EY = 0.$$

Posljedica je tog terema sljedeća: **ako dvodimenzionalan slučajni vektor ima kovarijancu različitu od nule, onda su njegove komponente nužno zavisne**.

Valja napomenuti da **općenito ne vrijedi obrat tog teorema**, tj. ako je kovarijanca od (X, Y) jednaka nuli, to ne znači nužno da su slučajne varijable X i Y nezavisne, što je ilustrirano sljedećim primjerom.

Primjer 3.27. *Neka je* U *diskretna slučajna vrijabla definirana tablicom distribucije*

$$U = \begin{pmatrix} -\pi/2 & 0 & \pi/2 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

i neka je (X, Y) *slučajni vektor definiran kao* $X = \sin U$, $Y = \cos U$*. Distribucija je tog slučajnog vektora zadana sljedećom tablicom:*

$X \setminus Y$	0	1	p_X
-1	$1/3$	0	$1/3$
0	0	$1/3$	$1/3$
1	$1/3$	0	$1/3$
p_Y	$2/3$	$1/3$	1

Kovarijanca je tog slučajnog vektora 0*, a komponente očito nisu nezavisne (npr.* $p(1, 1) \neq p_X(1)p_Y(1)$*).*

Definicija 3.6. *Neka je* (X, Y) *slučajni vektor za koji je* $\text{Cov}\,(X, Y) = 0$*. Tada kažemo da su njegove komponente* X *i* Y **nekorelirane**.

Iz definicije kovarijance vidimo da važnu ulogu u njezinom iznosu ima odstupanje pojedine komponente slučajnog vektora od njezinog očekivanja (devijacija). Međutim, devijaciju smo slučajne varijable već opisali varijancom, odnosno standardnom devijacijom slučajne varijable. Da bismo smanjili utjecaj devijacije na numeričku karakteristiku kojom želimo dobiti nove informacije o slučajnom vektoru, od velikog je interesa proučavanje kovarijance standardiziranog oblika slučajnog vektora

(X,Y). Naime, ukoliko X i Y imaju varijance $\sigma_X^2 \neq 0$ i $\sigma_Y^2 \neq 0$, možemo ih standardizirati (vidi poglavlje 2.7). Ako označimo $\mu_X = EX$, $\mu_Y = EY$, postupak standardizacije daje vektor

$$(X_s, Y_s) = \left(\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}\right)$$

čija je kovarijanca

$$\mathrm{Cov}\,(X_s, Y_s) = \frac{\mathrm{Cov}\,(X,Y)}{\sigma_X \sigma_Y}.$$

Tako nastaje **koeficijent korelacije** slučajnog vektora (X,Y) koji je definiran izrazom

$$\rho_{X,Y} = \frac{\mathrm{Cov}\,(X,Y)}{\sigma_X \sigma_Y}, \tag{3.18}$$

gdje su σ_X i σ_Y oznake za standardnu devijaciju odgovarajuće komponente slučajnog vektora.

Primjer 3.28. *Problem određivanja koeficijenta korelacije normalnog slučajnog vektora $(X,Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ (funkcija gustoće definirana je pravilom (3.7)), svodi se na izračunavanje njegove kovarijance. Za to nam treba $E(XY)$:*

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\,dxdy = \rho\sigma_1\sigma_2 + \mu_1\mu_2.$$

Sada slijedi da je

$$\mathrm{Cov}(X,Y) = E(XY) - EX\,EY = \rho\sigma_1\sigma_2 + \mu_1\mu 2 - \mu_1\mu_2 = \rho\sigma_1\sigma_2.$$

Prema tome, koeficijent korelacije komponenti X i Y dvodimenzionalnoga normalnog slučajnog vektora jest

$$\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}\,X \mathrm{Var}\,Y}} = \frac{\rho\sigma_1\sigma_2}{\sigma_1\sigma 2} = \rho.$$

Dakle, parametar ρ u distribuciji normalnoga slučajnog vektora (X,Y) jest zapravo koeficijent korelacije njegovih komponenti. Sjetimo se da smo u primjeru 3.23 pokazali da su za $\rho = 0$ komponente normalnoga slučajnog vektora nezavisne. To znači da u tom slučaju nekoreliranost slučajnih varijabli X i Y povlači njihovu nezavisnost, tj. u slučaju kad je $\rho = 0$ komponente dvodimenzionalnoga normalnog slučajnog vektora nezavisne su onda i samo onda ako su nekorelirane.

Sljedeći teorem daje dovoljne uvjete za postojanje kovarijance slučajnog vektora i jednu korisnu nejednakost.

Teorem 3.6. *Neka je (X,Y) slučajni vektor za koji je $0 < E\left(X^2\right) < \infty$ i $0 < E\left(Y^2\right) < \infty$. Tada postoji kovarijanca i vrijede nejednakosti:*

$$|\rho_{X,Y}| \leq 1,$$

$$\left(E(XY)\right)^2 \leq E\left(X^2\right)\, E\left(Y^2\right).$$

Dokaz. Pretpostavimo da je (X, Y) slučajni vektor takav da su X i Y slučajne varijable s $EX = EY = 0$ i $\operatorname{Var} X = \operatorname{Var} Y = 1$. Ako postoji kovarijanca tog slučajnog vektora, onda je $\operatorname{Cov}(X, Y) = E(XY)$ s obzirom da je $EX\,EY = 0$. Također je, za takve slučajne varijable, $\operatorname{Var} X = E\left(X^2\right)$ i $\operatorname{Var} Y = E\left(Y^2\right)$. Dakle, kovarijanca postoji ako postoji $E|XY|$.

Korištenjem nejednakosti $2|xy| \leq x^2 + y^2$ i monotonosti očekivanja vidimo da vrijedi:

$$E|XY| \leq \frac{1}{2}\left(E\left(X^2\right) + E\left(Y^2\right)\right) < \infty.$$

Dakle, kovarijanca postoji. Osim toga,

$$|E(XY)| \leq E|XY| \leq \frac{1}{2}\left(E\left(X^2\right) + E\left(Y^2\right)\right) = \frac{1}{2}(1+1) = 1.$$

S obzirom da je, za taj slučajni vektor, kovarijanca ujedno i koeficijent korelacije, vrijedi prva nejednakost teorema. Uočimo također da su za takve slučajne varijable prva i druga nejednakost u iskazu teorema ekvivalentne.

Pretpostavimo da je (X, Y) bilo koji slučajni vektor kao u iskazu teorema. Tada možemo provesti postupak standardizacije, a standardizirani oblik (X_s, Y_s) zadovoljava uvjete prvog dijela dokaza. Dakle, postoji kovarijanca od (X_s, Y_s) i vrijedi $\left(E(X_sY_s)\right)^2 \leq E\left(X_s^2\right)\, E\left(Y_s^2\right) = 1$. Osim toga vrijedi da je $X = \sigma_X X_s + \mu_X$, $Y = \sigma_Y Y_s + \mu_Y$, gdje su $\sigma_X, \mu_X, \sigma_Y, \mu_Y$ uobičajene oznake za standardne devijacije, odnosno očekivanja odgovarajućih slučajnih varijabli.

S obzirom da je

$$\begin{aligned} E|XY| &= E|(\sigma_X X_s + \mu_X)(\sigma_Y Y_s + \mu_Y)| \\ &\leq \sigma_X \sigma_Y E|X_sY_s| + \sigma_X|\mu_Y|E|X_s| + \sigma_Y|\mu_X|E|Y_s| + |\mu_X\mu_Y| \end{aligned}$$

vidimo da kovarijanca postoji. Koeficijent korelacije varijabli X i Y isti je kao koeficijent korelacije njihovih standardiziranih oblika, pa je, po prvom dijelu dokaza, po apsolutnoj vrijednosti manji ili jednak 1.

Za dokaz druge nejednakosti uočimo da je

$$\begin{aligned} \left(E(XY)\right)^2 &= \sigma_X^2\sigma_Y^2\left(E(X_sY_s)\right)^2 + 2\sigma_X\sigma_y\mu_X\mu_Y E(X_sY_s) + \mu_X^2\mu_Y^2 \\ &\leq \sigma_X^2\sigma_Y^2 + 2\sigma_X\sigma_y|\mu_X\mu_Y| + \mu_X^2\mu_Y^2 \\ &\leq (\sigma_X^2 + \mu_X^2)(\sigma_Y^2 + \mu_Y^2) \\ &= E\left(X^2\right)\, E\left(Y^2\right). \end{aligned}$$

Da je koeficijent korelacije korisna numerička karakteristika za opisivanje veze među komponentama slučajnog vektora, slijedi i iz sljedećeg teorema koji pokazuje da se linearna veza među komponentama može uočiti na osnovu vrijednosti koeficijenta korelacije.

Teorem 3.7. *Neka je (X, Y) slučajni vektor za koji je $0 < \sigma_X < \infty$ i $0 < \sigma_Y < \infty$. Veza je među komponentama linearna, tj. postoje realni brojevi a $(a \neq 0)$ i b takvi da je*

$$Y = aX + b$$

onda i samo onda ako je $|\rho_{X,Y}| = 1$. Pritom je koeficijent korelacije 1 ako je $a > 0$, odnosno -1 ako je $a < 0$.

Dokaz. Neka je slučajni vektor (X, Y) takav da je $Y = aX + b$, $(a \neq 0)$. Po pretpostavkama teorema postoji kovarijanca pa vrijedi:

$$E(X - EX)(Y - EY) =$$

$$= E\left((X - EX)(aX + b) - E(aX + b)\right) = E\left((X - EX)\, a(X - EX)\right) =$$

$$= aE(X - EX)^2.$$

Dakle,

$$\mathrm{Cov}\,(X, Y) = a\sigma_X^2.$$

No, $\mathrm{Var}\, Y = a^2 \mathrm{Var}\, X$, pa je

$$\rho_{X,Y} = \frac{\mathrm{Cov}\,(X, Y)}{\sigma_Y \sigma_X} = \frac{a\sigma_X^2}{\sigma_X |a| \sigma_X} = \frac{a}{|a|}.$$

Ako je $a > 0$, onda je $\rho_{X,Y} = 1$. Ako je $a < 0$, $\rho_{X,Y} = -1$.

Dokažimo obratnu tvrdnju. Pretpostavimo da je $\rho_{X,Y} = 1$. Definirajmo slučajnu varijablu

$$Z = \frac{1}{\sigma_X} X - \frac{1}{\sigma_Y} Y.$$

Tada je $\mathrm{Var}\, Z = 0$. Zaista,

$$\mathrm{Var}\, Z = E\left(\frac{1}{\sigma_X} X - \frac{1}{\sigma_Y} Y\right)^2 - \left(E(\frac{1}{\sigma_X} X - \frac{1}{\sigma_Y} Y)\right)^2 = 1 - 2\rho_{X,Y} + 1 = 0.$$

Dakle, Z je konstanta, označimo $Z = c$. Vrijedi:

$$c = \frac{1}{\sigma_X} X - \frac{1}{\sigma_Y} Y,$$

$$Y = \frac{\sigma_Y}{\sigma_X} X - \sigma_Y c.$$

Za $\rho_{X,Y} = -1$ se tvrdnja može dokazati na isti način analiziranjem slučajne varijable

$$Z_1 = \frac{1}{\sigma_X} X + \frac{1}{\sigma_Y} Y.$$

Slično kao kod vektora realnih brojeva, često je korisno i slučajni vektor zapisivati u matričnom obliku, tj. slučajni vektor (X, Y) zapisujemo kao $[X, Y]^\tau$. Tu oznaku koristit ćemo u nastavku za definiranje očekivanja slučajnog vektora.

Ako za slučajni vektor $\mathbf{Z} = [X, Y]^\tau$ postoje EX i EY, zapisivat ćemo ih u vektorskom obliku kao $[EX, EY]^\tau$ i zvati **očekivanje** slučajnog vektora $\mathbf{Z} = [X, Y]^\tau$. Varijance i kovarijancu također zapisujemo matrično. Naime, označimo li $E\mathbf{Z} = [EX, EY]^\tau$, onda je

$$E(\mathbf{Z} - E\mathbf{Z})(\mathbf{Z} - E\mathbf{Z})^\tau = \begin{bmatrix} E(X - EX)^2 & E(X - EX)(Y - EZ) \\ E(X - EX)(Y - EZ) & E(Y - EY)^2 \end{bmatrix}$$

$$= \begin{bmatrix} \operatorname{Var} X & \operatorname{Cov}(X, Y) \\ \operatorname{Cov}(X, Y) & \operatorname{Var} Y \end{bmatrix}.$$

Takvu matricu $\mathbf{Cov}\,(X, Y) = E(\mathbf{Z} - E\mathbf{Z})(\mathbf{Z} - E\mathbf{Z})^\tau$ zovemo **matrica kovarijanci** slučajnog vektora (X, Y).
Uočimo:

- matrica kovarijanci jest simetrična,
- matrica kovarijanci jest pozitivno semidefinitna.[6]

Neka je $\mathbf{Z_s}$ standardizirana varijanta slučajnog vektora $\mathbf{Z}$, tj.

$$\mathbf{Z}_s = \left[\frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y} \right]^\tau.$$

Njegovu matricu kovarijanci zovemo **korelacijska matrica** slučajnog vektora $\mathbf{Z} = [X, Y]^\tau$ i označavamo $\mathbf{Corr}\,(X, Y)$. Očigledno vrijedi

$$\mathbf{Corr}\,(X, Y) = \begin{bmatrix} 1 & \rho_{X,Y} \\ \rho_{X,Y} & 1 \end{bmatrix}.$$

[6] Dokažite ove tvrdnje.

Primjer 3.29. *Neka su* $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ *i* $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ *nezavisne normalne slučajne varijable. Njih možemo shvatiti kao komponente dvodimenzionalnoga normalnog slučajnog vektora s očekivanjem* $[\mu_1, \mu_2]^\tau$*, matricom kovarijanci*

$$\mathbf{Cov}(X, Y) = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

i korelacijskom matricom

$$\mathbf{Corr}(X, Y) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Očekivanje slučajnog vektora, matrica kovarijanci i korelacijska matrica mogu se na analogan način definirati za n-dimenzionalan slučajni vektor $\mathbf{Z} = [X_1, \ldots, X_n]^\tau$.

3.4 Općenito o nezavisnosti slučajnih varijabli

Iz definicije nezavisnosti diskretnih slučajnih varijabli (definicija 3.3), može se vidjeti da je jednakost

$$p(x_i, y_j) = p_X(x_i)\, p_Y(y_j), \quad \text{za sve } x_i \in \mathcal{R}(X) \text{ i } y_j \in \mathcal{R}(Y),$$

ekvivalentna jednakosti

$$F(x, y) = F_X(x) F_Y(y), \quad \text{za sve } (x, y) \in \mathbb{R}^2, \tag{3.19}$$

gdje je F funkcija distribucije diskretnoga slučajnog vektora (X, Y), F_X marginalna funkcija distribucije komponente X, a F_Y marginalna funkcija distribucije komponente Y.

Budući da su vjerojatnosna svojstva slučajne varijable općenito definirana funkcijom distribucije, nezavisnost ćemo slučajnih varijabli općenito opisati upravo korištenjem jednakosti (3.19).

Neka su X i Y slučajne varijable s funkcijama distribucije F_X i F_Y, redom, te neka je F funkcija distribucije slučajnog vektora (X, Y). Reći ćemo da su X i Y **nezavisne** slučajne varijable ako vrijedi

$$F(x, y) = F_X(x)\, F_Y(y) \quad \text{za sve } (x, y) \in \mathbb{R}^2. \tag{3.20}$$

Takav je opis nezavisnih slučajnih varijabli lako poopćiti i na n-dimenzionalan slučajni vektor.

Neka su $X_1, \dots, X_n$ slučajne varijable na nekom vjerojatnosnom prostoru $(\Omega, \mathcal{F}, P)$ s pripadnim funkcijama distribucije $F_1, \dots, F_n$ i neka je F funkcija distribucije slučajnog vektora $(X_1, \dots, X_n)$, tj.

$$F(x_1, \dots, x_n) = P\{X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n\}, \quad (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Reći ćemo da su slučajne varijable $X_1, \dots, X_n$ **nezavisne** ako vrijedi[7]

$$F(x_1, \dots, x_n) = F_1(x_1) \cdots F_n(x_n). \tag{3.21}$$

Posljedica je nezavisnosti skupa slučajnih varijabli $\{X_1, \dots, X_n\}$ da i svi podskupovi tog skupa sadrže međusobno nezavisne slučajne varijable. (Npr. X_i i X_j, $i, j = 1, \dots, n$, $i \neq j$ jesu također nezavisne.)

Vezano uz pojam nezavisnosti važno je istaknuti da za nezavisne slučajne varijable koje imaju varijancu vrijedi sljedeći teorem.

Teorem 3.8. *Neka su $X_1, \dots, X_n$ nezavisne slučajne varijable takve da postoji varijanca $\operatorname{Var} X_k$, $k = 1, \dots, n$ i neka su $a_1, \dots, a_n$ realni brojevi. Tada vrijedi:*

$$\operatorname{Var}\left(\sum_{k=1}^{n} a_k X_k\right) = \sum_{k=1}^{n} a_k^2 \operatorname{Var} X_k. \tag{3.22}$$

Dokaz. Promotrimo prvo $\operatorname{Var}\left(\sum\limits_{k=1}^{n} a_k X_k\right)$:

$$\operatorname{Var}\left(\sum_{k=1}^{n} a_k X_k\right) = E\left(\sum_{k=1}^{n} a_k X_k - E\left(\sum_{k=1}^{n} a_k X_k\right)\right)^2 =$$

$$= E\left(\sum_{k=1}^{n} (a_k X_k - a_k E X_k)\right)^2 = E\left(\sum_{k,j=1}^{n} (a_k X_k - a_k E X_k)(a_j X_j - a_j E X_j)\right).$$

Za $k \neq j$ jest zbog nezavisnosti

$$E(X_k - EX_k)(X_j - EX_j) = 0,$$

pa zbog linearnosti očekivanja vrijedi:

$$\operatorname{Var}\left(\sum_{k=1}^{n} a_k X_k\right) = E\left(\sum_{k=1}^{n} (a_k X_k - a_k E X_k)^2\right) =$$

$$= \sum_{k=1}^{n} a_k^2 E(X_k - EX_k)^2 = \sum_{k=1}^{n} a_k^2 \operatorname{Var} X_k.$$

---

[7] Općenitu definiciju nezavisnosti slučajnih varijabli $X_1, \dots, X_n$ zainteresiran čitatelj može pronaći u [26].

Primjer 3.30. *Neka su* $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, $\ldots$, $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$ *nezavisne normalne slučajne varijable. Njih možemo shvatiti kao komponente n-dimenzionalnoga normalnog slučajnog vektora s očekivanjem* $[\mu_1, \ldots, \mu_n]^\tau$, *matricom kovarijanci*

$$\mathbf{Cov}(X_1, \ldots, X_n) = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \sigma_n^2 \end{bmatrix}$$

i jediničnom korelacijskom matricom reda n.
Definirajmo slučajnu varijablu

$$S_n = \sum_{k=1}^{n} X_k = X_1 + \ldots + X_n$$

i izračunajmo njezino matematičko očekivanje i varijancu. Zbog linearnosti očekivanja slijedi da je

$$ES_n = E\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} EX_k = \sum_{k=1}^{n} \mu_k.$$

Varijancu slučajne varijable S_n, *zbog nezavisnosti slučajnih varijabli* $X_1, \ldots, X_n$, *računamo pomoću teorema 3.8. Dakle,*

$$\mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathrm{Var}\, X_k = \sum_{k=1}^{n} \sigma_k^2.$$

Uočimo da ukoliko su slučajne varijable $X_1, \ldots, X_n$ *i jednako distribuirane, tj. ako je* $X_k \sim \mathcal{N}(\mu, \sigma^2)$ *za svaki* $k \in \{1, \ldots, n\}$, *tada je* $ES_n = n\mu$, *a* $\mathrm{Var}\, S_n = n\sigma^2$.

3.5 Neki rezultati o nizovima slučajnih varijabli

U poglavlju 1.3 o statističkom pristupu modeliranju vjerojatnosti naveli smo svojstvo statističke stabilnosti relativnih frekvencija koje kaže da se, kod **nezavisnog** ponavljanja istog pokusa puno puta, relativna frekvencija pojavljivanja odabranog događaja stabilizira u okolini nekog broja.

U poglavlju u kojemu je definirana normalna slučajna varijabla opisan je pokus nazvan "Galtonova daska" koji ilustrira činjenicu da normalna distribucija dobro modelira realizacije pokusa čiji su ishodi posljedica sume mnogo međusobno nezavisnih i jednako distribuiranih utjecaja.

U ovom poglavlju navodimo dva rezultata koji se odnose na granično ponašanje nizova slučajnih varijabli i dokazuju istinitost navedenih svojstava.

3.5.1 Slabi zakon velikih brojeva i Bernoullijeva shema

Korištenjem slabog zakona velikih brojeva možemo objasniti statističku stabilnost relativnih frekvencija u uvjetima navedenim u poglavlju 1.3.

Definicija 3.7. *Kažemo da niz slučajnih varijabli $X_1, X_2, \dots$ zadovoljava* **slabi zakon velikih brojeva** *ako za svaki $\varepsilon > 0$ vrijedi:*

$$\lim_{n\to\infty} P\left\{\frac{1}{n}|S_n - ES_n| \geq \varepsilon\right\} = 0,$$

gdje je

$$S_n = \sum_{k=1}^{n} X_k.$$

Taj zakon možemo interpretirati na sljedeći način:
Ako niz slučajnih varijabli zadovoljava slabi zakon velikih brojeva, onda vjerojatnost da se realizacija prosjeka slučajnih varijabli $X_1, \dots, X_n$ razlikuje od očekivanja prosjeka za više od proizvoljno izabranog malog broja teži nuli s povećanjem broja slučajnih varijabli u izračunu prosjeka.
Postoje mnogi rezultati koji daju dovoljne uvjete na niz slučajnih varijabli da bi on zadovoljavao slabi zakon velikih brojeva. Ovdje ćemo dokazati samo jedan koji će biti dovoljan za objašnjenje statističke stabilnosti relativnih frekvencija odabranog događaja kod nezavisnog ponavljanja istog pokusa.

Teorem 3.9. *Neka je $(X_n, n \in \mathbb{N})$ niz slučajnih varijabli takav da su, za svaki $n \in \mathbb{N}$, slučajne varijable $X_1, \dots, X_n$ nezavisne i neka su varijance tih slučajnih varijabli uniformno ograničene, tj. postoji $M > 0$ takav da je*

$$\max_{k\in\mathbb{N}} \operatorname{Var} X_k \leq M.$$

Tada za svaki $\varepsilon > 0$ vrijedi

$$\lim_{n\to\infty} P\left\{\frac{1}{n}|S_n - ES_n| \geq \varepsilon\right\} = 0.$$

Dokaz. *U dokazu koristimo Čebiševljevu nejednakost, tj. za svaku je slučajnu varijablu Y koja ima varijancu i za svaki $\varepsilon > 0$*

$$P\{|Y - EY| \geq \varepsilon\} \leq \frac{\operatorname{Var} Y}{\varepsilon^2}.$$

Neka je n dan prirodan broj i $Y_n = \dfrac{S_n}{n}$. Zbog nezavisnosti je

$$\operatorname{Var} Y_n = \frac{1}{n^2} \operatorname{Var} S_n$$

pa vrijedi:

$$P\left\{\frac{1}{n}|S_n - ES_n| \geq \varepsilon\right\} = P\{|Y_n - EY_n| \geq \varepsilon\} \leq$$
$$\leq \frac{\operatorname{Var} S_n}{n\varepsilon^2} \leq \frac{M}{n\varepsilon^2}.$$

Dakle, slijedi da je

$$0 \leq P\left\{\frac{1}{n}|S_n - ES_n| \geq \varepsilon\right\} \leq \frac{M}{n\varepsilon^2}.$$

Stoga je

$$\lim_{n\to\infty} P\left\{\frac{1}{n}|S_n - ES_n| \geq \varepsilon\right\} = 0.$$

Primjer 3.31. *Neka je $(X_n,\, n \in \mathbb{N})$ niz nezavisnih slučajnih varijabli čije su distribucije zadane tablicama*

$$X_n = \begin{pmatrix} -2^n & 0 & 2^n \\ \frac{1}{2^{2n+1}} & 1 - \frac{1}{2^{2n}} & \frac{1}{2^{2n+1}} \end{pmatrix}, \quad n \in \mathbb{N}.$$

Da bismo provjerili zadovoljava li taj niz slabi zakon velikih brojeva, izračunajmo očekivanja i varijance slučajnih varijabli X_n i $S_n = \sum\limits_{i=1}^{n} X_i$:

$$\begin{aligned}
EX_n &= -2^n \cdot \frac{1}{2^{2n+1}} + 2^n \cdot \frac{1}{2^{2n+1}} = 0, \quad \forall n \in \mathbb{N},\\
\operatorname{Var} X_n &= EX_n^2 = 2^{2n} \cdot \frac{1}{2^{2n+1}} + 2^{2n} \cdot \frac{1}{2^{2n+1}} = 1, \quad \forall n \in \mathbb{N},\\
ES_n &= E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} EX_i = 0, \quad \forall n \in \mathbb{N},\\
\operatorname{Var} S_n &= \operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var} X_i = \sum_{i=1}^{n} 1 = n.
\end{aligned}$$

Budući da je $(X_n,\, n \in \mathbb{N})$ niz nezavisnih slučajnih varijabli čije su varijance jednake jedan, slijedi da niz $(X_n,\, n \in \mathbb{N})$ zadovoljava slabi zakon velikih brojeva 3.9, tj. da je za svaki $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\frac{1}{n}\,|S_n| \geq \varepsilon\right\} = 0.$$

Riječima rečeno, za svaki se $\varepsilon > 0$ vjerojatnost odstupanja prosjeka niza nezavisnih slučajnih varijabli iz primjera 3.31 od nule za ε ili više može učiniti proizvoljno malom odabirom dovoljno velikog prirodnog broja n.

Slabi zakon velikih brojeva specijalno vrijedi i za niz nezavisnih i jednako distribuiranih slučajnih varijabli koje imaju varijancu, a koji je opisan sljedećom definicijom.

Definicija 3.8. *Kažemo da je niz slučanih varijabli $(X_n, n \in \mathbb{N})$* **niz nezavisnih i jednako distribuiranih** *slučajnih varijabli ako vrijedi:*

- *sve slučajne varijable $X_1, X_2, \ldots$ imaju istu distribuciju i*
- *za sve su $n \in \mathbb{N}$ slučajne varijable $X_1, \ldots, X_n$ nezavisne.*

Za takav niz koristimo kraticu "n. j. d. niz".

Pretpostavimo da je $(X_n, n \in \mathbb{N})$ n. j. d. niz i neka je varijanca $\mathrm{Var}\, X_1 = \sigma^2$, a $EX_1 = \mu$. Uočimo da svaka slučajna vrijabla iz tog niza ima isto očekivanje i varijancu kao X_1 zbog jednake distribuiranosti.

Neka je $\overline{X}_n$ aritmetička sredina (prosjek) slučajnih varijabli $X_1, \ldots X_n$, tj.

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

i $\left(\overline{X}_n,\, n \in \mathbb{N}\right)$ pripadni niz aritmetičkih sredina. Koristeći svojstva očekivanja i varijance vidimo da vrijedi:

$$E\overline{X}_n = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} EX_i = \mu,$$

$$\mathrm{Var}\, \overline{X}_n = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}\, X_i = \frac{\sigma^2}{n}.$$

Osim toga,

$$\overline{X}_n = \frac{1}{n} S_n$$

pa se na taj niz aritmetičkih sredina može primijeniti slabi zakon velikih brojeva, tj. za svaki je $\varepsilon > 0$

$$\lim_{n\to\infty} P\{|\overline{X}_n - \mu| \geq \varepsilon\} = 0.$$

Taj rezultat možemo tumačiti na sljedeći način:

Vjerojatnost da se aritmetička sredina n. j. d. niza slučajnih varijabli razlikuje od njihovog očekivanja za ε ili više možemo učiniti proizvoljno malom birajući n dovoljno veliko.

Primjer 3.32. *Neka je $(X_n,\, n \in \mathbb{N})$ niz nezavisnih standardnih normalnih slučajnih varijabli, tj.*

$$X_n \sim \mathcal{N}(0,1), \quad \forall n \in \mathbb{N}.$$

Sada znamo da je za svaki prirodan broj n $EX_n = 0$ i $\mathrm{Var}\, X_n = 1$, pa stoga taj niz slučajnih varijabli zadovoljava slabi zakon velikih brojeva, tj. za svaki je $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{|\overline{X}_n| \geq \varepsilon\right\} = 0.$$

Riječima rečeno, za svaki se $\varepsilon > 0$ vjerojatnost da aritmetička sredina niza nezavisnih standardnih normalnih slučajnih varijabli odstupa od nule za barem ε može učiniti proizvoljno malom odabirom dovoljno velikog prirodnog broja n.

Budući da je za jednako distribuirane slučajne varijable $X_1, \dots, X_n$ $\operatorname{Var} \overline{X}_n = \sigma^2/n$, iz Čebiševljeve nejednakosti slijedi da je

$$P\{|\overline{X}_n - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2 n}.$$

Uzmimo npr. $\varepsilon = 0.01$. Ako želimo postići vjerojatnost manju od 0.05 da se aritmetička sredina niza $(X_n,\, n \in \mathbb{N})$ razlikuje od očekivanja $\varepsilon = 0.01$ ili više, tada n ocjenjujemo na sljedeći način:

$$n \geq \frac{\sigma^2}{\varepsilon^2 P\left\{\left|\overline{X}_n\right| \geq \varepsilon\right\}} = 200000.$$

Statistička definicija vjerojatnosti

Korištenjem slabog zakona velikih brojeva primijenjenog na n. j. d. niz Bernoullijevih slučajnih varijabli možemo se uvjeriti da je opravdan statistički pristup modeliranju vjerojatnosti, pod pretpostavkom da modeliramo vjerojatnost pojavljivanja događaja na temalju nezavisnih ponavljanja istog pokusa (poglavlje 1.3).

Taj problem opisat ćemo koristeći tzv. Bernoullijevu shemu.

Bernoullijeva shema jest slučajni pokus koji se sastoji od n međusobno nezavisnih ponavljanja uvijek istog Bernoullijevog pokusa opisanog distribucijom

$$X = \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}, \quad 0 \leq p \leq 1.$$

Dakle, $(X_1, \dots X_n)$ je n. j. d. vektor s distribucijom svake komponente jednakom distribuciji od X.

Primjer 3.33. *Model Bernoullijeve sheme može se primijeniti kod nezavisnog bacanja istog novčića n puta zaredom, pri čemu je p vjerojatnost da se okrene npr. glava.*

Skup je svih mogućih realizacija tog slučajnog vektora

$$\mathcal{R}(X_1, \dots, X_n) = \{(x_1, \dots, x_n) : x_i \in \{0, 1\}\},$$

a distribucija je zadana s

$$\begin{aligned} p(x_1, \dots, x_n) &= P\{X_1 = x_1, \dots, X_n = x_n\} \\ &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}, \;\; (x_1, \dots, x_n) \in \mathcal{R}(X_1, \dots, X_n). \end{aligned}$$

Tako je, npr.

$$p(1, 1, 1, 0, 0, 0, \dots, 0) = p^3(1-p)^{n-3},$$

$$p(0, 1, 1, 0, 0, 0, \dots, 0) = p^2(1-p)^{n-2}.$$

U statističkoj definiciji vjerojatnosti koristili smo upravo takav model, tj. isti pokus $(\Omega, \mathcal{F}, P)$ ponavljali smo n puta nezavisno i definirali vjerojatnost događaja $A \in \mathcal{F}$ kao broj oko kojega se gomilaju relativne frekvencije događaja A s povećavanjem broja ponavljanja pokusa, tj.

$$P(A) \approx \frac{f_A}{n},$$

gdje je f_A frekvencija događaja A.
Neka je X slučajna varijabla u tom pokusu definirana na sljedeći način:

$$X(\omega) = \begin{cases} 1 \, , \, \omega \in A \\ 0 \, , \, \omega \notin A \end{cases}.$$

Tada je

$$X = \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}, \quad 0 \leq p \leq 1,$$

gdje je $p = P\{X = 1\}$ vjerojatnost da se dogodio događaj A.

Nezavisnim ponavljanjem našeg pokusa nastaje n. j. d. niz $(X_1, \ldots, X_n)$ Bernoullijevih slučajnih varijabli na koji možemo primijeniti slabi zakon velikih brojeva. Pri tome je:

$$EX = EX_i = p, \quad i \in \{1, \ldots n\},$$

$$f_A = X_1 + X_2 + \ldots + X_n = \sum_{i=1}^{n} X_i,$$

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{f_A}{n},$$

pa je

$$\lim_{n \to \infty} P\left\{ \left| \frac{f_A}{n} - p \right| \geq \varepsilon \right\} = 0$$

za sve $\varepsilon > 0$, što opravdava statistički pristup modeliranju vjerojatnosti pod navedenim uvjetima.

Primjer 3.34. *Isti model može se primijeniti i kod nezavisnog bacanja jedne pravilno izrađene igraće kockice n puta zaredom. Označimo sa 1 realizaciju događaja "okrenula se šestica", a s 0 realizaciju događaja "nije se okrenula šestica". Ishod svakog od n bacanja te kockice modeliramo Bernoullijevom slučajnom varijablom*

$$X = \begin{pmatrix} 0 & 1 \\ 5/6 & 1/6 \end{pmatrix},$$

čiju distribuciju možemo, osim gornjom tablicom, zapisati i na sljedeći način:

$$\mathcal{R}(X) = \{0, 1\},$$

$$P\{X=x\}=\left(\frac{1}{6}\right)^x\left(\frac{5}{6}\right)^{1-x}=\frac{5^{1-x}}{6},\quad x\in\mathcal{R}(X).$$

Dakle, X je Bernoullijeva slučajna varijabla s parametrom $p=1/6$.
Ishode n nezavisnih bacanja te kockice, pri čemu nas zanimaju događaji "okrenula se šestica" (1) i "nije se okrenula šestica" (0), modelirat ćemo nezavisnim slučajnim varijablama $X_1,\ldots,X_n$ koje su jednako distribuirane kao slučajna varijabla X, tj. n-dimenzionalnim slučajnim vektorom $(X_1,\ldots,X_n)$. Zbog nezavisnosti i jednake distribuiranosti slučajnih varijabli $X_1,\ldots,X_n$, slijedi da je distribucija slučajnog vektora $(X_1,\ldots,X_n)$ dana izrazom

$$P\{X_1=x_1,\ldots,X_n=x_n\}=\prod_{i=1}^n\frac{5^{1-x_i}}{6},\quad x_i\in\mathcal{R}(X_i),\quad i\in\{1,\ldots,n\}.$$

Tako je, primjerice, vjerojatnost da se u n nezavisnih bacanja jedne pravilno izrađene igraće kockice svaki put (tj. svih n puta) okrene šestica

$$P\{X_1=1,\ldots,X_n=1\}=\frac{1}{6^n}.$$

3.5.2 Centralni granični teorem

Centralni granični teoremi govore o uvjetima pod kojima niz funkcija distribucije standardiziranih suma slučajnih varijabli konvergira prema funkciji distribucije standardne normalne slučajne varijable.

Definicija 3.9. *Neka je $(X_n,\, n\in\mathbb{N})$ niz slučajnih varijabli s pripadnim nizom funkcija distribucije $(F_n,\, n\in\mathbb{N})$. Ako postoji funkcija distribucije F takva da $F_n(x)\to F(x)$ za svaki x u kojemu je F neprekidna, tada kažemo da niz $(X_n,\, n\in\mathbb{N})$ konvergira po distribuciji prema slučajnoj varijabli X čija je funkcija distribucije F i pišemo:*

$$X_n\xrightarrow{D}X.$$

Konvergencija po distribuciji zapravo znači da, za dovoljno velike n, $P\{X_n\le x\}$ možemo aproksimirati s $F(x)$.

Primjer 3.35. *Neka je $(X_n,\, n\in\mathbb{N})$ niz nezavisnih slučajnih varijabli uniformno distribuiranih na intervalu $\langle 0,a\rangle$, $a>0$, tj. za svaki $i\in\mathbb{N}$ jest*

$$F_{X_i}(x)=\begin{cases}0 & ,\ x\in\langle-\infty,0\rangle\\ x/a & ,\ x\in[0,a\rangle\\ 1 & ,\ x\in[a,\infty\rangle\end{cases}.$$

Ako je $Y_n=\max\{X_1,\ldots,X_n\}$, tada je zbog nezavisnosti i jednake distribuiranosti slučajnih varijabli iz niza $(X_n,\, n\in\mathbb{N})$ niz $(Y_n,\, n\in\mathbb{N})$ niz slučajnih varijabli s funkcijama distribucije

$$F_{Y_n}(y)=P\{Y_n\le y\}=P\{\max\{X_1,\ldots,X_n\}\le y\}=P\{X_1\le y,\ldots,X_n\le y\}=$$
$$=P\{X_1\le y\}\cdot\ldots\cdot P\{X_n\le y\}=\left(F_{X_1}(y)\right)^n.$$

Definirajmo sada novi niz slučajnih varijabli $(Z_n,\, n\in\mathbb{N})$, gdje je

$$Z_n=n(a-Y_n),\quad a>0.$$

Budući da za svaki $n \in \mathbb{N}$ znamo funkciju distribucije slučajne varijable Y_n, slijedi da je funkcija distribucije slučajne varijable Z_n dana izrazom

$$F_{Z_n}(z) = P\{Z_n \leq z\} = P\{n(a - Y_n) \leq z\} = 1 - P\left\{Y_n < a - \frac{z}{n}\right\} =$$

$$= \begin{cases} 0 & , z \in \langle -\infty, 0\rangle \\ 1 - \left(1 - \frac{z}{an}\right)^n & , z \in [0, an\rangle \\ 1 & , z \in [an, \infty\rangle \end{cases}.$$

Budući da je

$$\lim_{n\to\infty} \left(1 - \left(1 - \frac{z}{an}\right)^n\right) = 1 - e^{-\frac{z}{a}}, \quad z \geq 0,$$

slijedi da je

$$\lim_{n\to\infty} F_{Z_n}(z) = \begin{cases} 0 & , z \in \langle -\infty, 0\rangle \\ 1 - e^{-\frac{z}{a}} & , z \in [0, \infty\rangle \end{cases}.$$

Dakle, niz slučajnih varijabli $(Z_n,\, n \in \mathbb{N})$ po distribuciji konvergira prema eksponencijalnoj slučajnoj varijabli s parametrom $1/a$.

Primjer 3.36. *Neka je $(X_n,\, n \in \mathbb{N})$ niz slučajnih varijabli s funkcijama distribucija*

$$F_{X_n}(x) = \begin{cases} 0 & , x \in \langle -\infty, 0\rangle \\ x^n & , x \in [0, 1\rangle \\ 1 & , x \in [1, \infty\rangle \end{cases}.$$

Budući da je za $x \in [0, 1\rangle$

$$\lim_{n\to\infty} x^n = 0,$$

slijedi da je

$$\lim_{n\to\infty} F_{X_n}(x) = \begin{cases} 0 & , x \in \langle -\infty, 1\rangle \\ 1 & , x \in [1, \infty\rangle \end{cases},$$

tj. niz slučajnih varijabli $(X_n,\, n \in \mathbb{N})$ konvergira po distribuciji prema degeneriranoj slučajnoj varijabli X za koju je $P\{X = 1\} = 1$. Kažemo da niz slučajnih varijabli $(X_n,\, n \in \mathbb{N})$ konvergira po distribuciji prema jedinici.

Sljedeći teorem govori o konvergenciji po distribuciji niza standardiziranih parcijalnih suma niza nezavisnih jednako distribuiranih slučajnih varijabli prema standardnoj normalnoj slučajnoj varijabli.

Teorem 3.10 (Lévy)**.** *Neka je $(X_n, n \in \mathbb{N})$ n. j. d. niz slučajnih varijabli s očekivanjem μ i varijancom σ^2 i neka je S_n njegova n-ta parcijalna suma. Tada*[8]

$$\frac{S_n - ES_n}{\sqrt{\operatorname{Var} S_n}} \xrightarrow{D} Z \sim \mathcal{N}(0, 1).$$

Primjer 3.37. *Iznimno, ako je n. j. d. niz $(X_n,\, n \in \mathbb{N})$ niz Bernoullijevih slučajnih varijabli*

$$X_1 = \begin{pmatrix} 0 & 1 \\ 1 - p & p \end{pmatrix}, \quad p \in \langle 0, 1\rangle,$$

(vidi Bernoullijenu shemu), tada je $S_n \sim \mathcal{B}(n, p)$, $ES_n = np$, $\operatorname{Var} S_n = np(1 - p)$, pa možemo zaključiti da

$$\frac{S_n - ES_n}{\sqrt{\operatorname{Var} S_n}} = \frac{S_n - np}{\sqrt{np(1 - p)}} \xrightarrow{D} Z \sim \mathcal{N}(0, 1).$$

[8] Za dokaz pogledati [26].

Dakle, za velike se n vjerojatnosti po binomnoj distribuciji mogu približno računati koristeći normalnu distribuciju.

Primjer 3.38. *Neka je $X \sim \mathcal{B}(500, 0.4)$. Tada znamo da je*

$$P\{X \leq 175\} = \sum_{i=0}^{175} \binom{500}{i} 0.4^i 0.6^{500-i}.$$

Budući da je izračunavanje te sume komplicirano, a n velik, za izračun vjerojatnosti $P\{X \leq 175\}$ možemo koristiti aproksimaciju binomne distribucije normalnom distribucijom. Uočimo da je $np = 200$ i $\sqrt{np(1-p)} = 10.95$, pa slijedi da je

$$P\{X \leq 175\} = P\left\{\frac{X-200}{10.95} \leq \frac{175-200}{10.95}\right\} = P\left\{\frac{X-200}{10.95} \leq -2.28\right\} \approx P\{Z \leq -2.28\},$$

gdje je Z standardna normalna slučajna varijabla. Prethodnu vjerojatnost možemo izračunati korištenjem prikladnog matematičkog softvera:

$$P\{X \leq 175\} \approx 0.0113.$$

Analognim postupkom možemo aproksimirati i mnoge druge vjerojatnosti - tako je, npr. $P\{175 < X \leq 225\} \approx 0.98$.

Primjena na aritmetičku sredinu

Neka je $(X_n,\, n \in \mathbb{N})$ n. j. d. niz slučajnih varijabli s očekivanjm μ i varijancom σ^2 i

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

S obzirom da je

$$\overline{X}_n = \frac{1}{n} S_n, \quad E\overline{X}_n = \mu, \quad \operatorname{Var}\overline{X}_n = \frac{\sigma^2}{n},$$

očito je

$$\frac{S_n - ES_n}{\sqrt{\operatorname{Var} S_n}} = \sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma}.$$

Dakle, prema centralnom graničnom teoremu 3.5.2 slijedi da

$$\sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma} \xrightarrow{D} Z \sim \mathcal{N}(0,1),$$

pa kažemo da se $\sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma}$ asimptotski ponaša kao slučajna varijabla s distribucijom $\mathcal{N}(0,1)$, odnosno da $\overline{X}_n$ asimptotski ima $\mathcal{N}(\mu, \frac{\sigma^2}{n})$ distribuciju.

Primjer 3.39. *Neka je $(X_n,\, n \in \mathbb{N})$ niz nezavisnih eksponencijalnih slučajnih varijabli s parametrom $\lambda = 1$. Tada iz tablice 2.6.5 znamo da je $\mu = EX_n = 1/\lambda = 1$ i $\sigma^2 = \operatorname{Var}X_n = 1/\lambda^2 = 1$ za svaki $n \in \mathbb{N}$. Prema teoremu 3.5.2 slijedi da za veliki n slučajna varijabla*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma}$$

ima približno standardnu normalnu distribuciju ($Z \sim \mathcal{N}(0,1)$). Sada možemo aproksimirati npr. sljedeće vjerojatnosti:

$$P\{S_{100} \geq 110\} = P\left\{\frac{S_{100} - 100}{10} \geq 1\right\} \approx P\{Z \geq 1\} = \frac{1}{\sqrt{2\pi}} \int_1^{\infty} e^{-\frac{z^2}{2}}\, dz \approx 0.1587,$$

$$P\{1.1 < \overline{X}_{100} < 1.2\} = P\left\{1 < 10(\overline{X}_{100} - 1) < 2\right\} \approx$$
$$\approx P\{1 < Z < 2\} = \frac{1}{\sqrt{2\pi}} \int_1^{2} e^{-\frac{z^2}{2}}\, dz \approx 0.136.$$

Primjer 3.40. *Neka je $(X_n,\, n \in \mathbb{N})$ niz nezavisnih jednako distribuiranih slučajnih varijabli s očekivanjem $\mu = EX_n = 20$ i varijancom $\sigma^2 = \mathrm{Var} X_n = 4$. Pretpostavimo da nas zanima kolika je vjerojatnost da se aritmetička sredina $\overline{X}_n$ slučajnih varijabli $X_1, \dots X_n$ realizira brojem iz intervala $\langle 19.9, 20.1\rangle$. Budući da distribucija slučajnih varijabli iz tog niza nije zadana, traženu vjerojatnost ne možemo točno izračunati. No taj niz zadovoljava uvjete teorema 3.5.2 pa znamo da*

$$\sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma} = \frac{\overline{X}_n - 20}{\frac{2}{\sqrt{n}}} \xrightarrow{D} Z \sim \mathcal{N}(0,1).$$

Traženu vjerojatnost možemo aproksimirati na sljedeći način:

$$P\{19.9 < \overline{X}_n < 20.1\} = P\left\{-\frac{\sqrt{n}}{20} < \frac{\overline{X}_n - 20}{\frac{2}{\sqrt{n}}} < \frac{\sqrt{n}}{20}\right\} \approx$$

$$\approx P\left\{-\frac{\sqrt{n}}{20} < Z < \frac{\sqrt{n}}{20}\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{n}/20}^{\sqrt{n}/20} e^{-z^2/2}\, dz.$$

3.6 Zadaci

Zadatak 3.4. Dokažite svojstva 1 do 5 funkcije distribucije dvodimenzionalnog slučajnog vektora.

Zadatak 3.5. Dokažite da za slučajne varijable X_1 i X_2 koje imaju končnu varijancu i koje su nezavisne vrijedi jednakost

$$\mathrm{Var}\,(aX_1 + bX_2 - c) = a^2 \mathrm{Var}\, X_1 + b^2 \mathrm{Var}\, X_2,$$

gdje su a, b i c proizvoljni realni brojevi.

Zadatak 3.6. Neka je (X, Y) slučajni vektor s kovarijancom $\mathrm{Cov}(X, Y)$. Dokažite da za proizvoljne realne brojeve a_1, a_2, b_1, b_2 vrijedi

$$\mathrm{Cov}(a_1 X + b_1, a_2 Y + b_2) = a_1 a_2\, \mathrm{Cov}(X, Y).$$

Zadatak 3.7. Neka je X slučajna varijabla takva da je $EX = 6$, $\mathrm{Var}(X) = 1$ i neka je $Y = -2X + 5$. Odredite EY, $\mathrm{Var}\,Y$ i $\rho_{X,Y}$.
Rješenje: $EY = -7$, $\mathrm{Var}\,Y = 4$, $\rho_{X,Y} = 109/2$.

Zadatak 3.8. Neka su X i Y slučajne varijable za koje je $EX = 2$, $EY = -1$, $\mathrm{Var}\,X = 9$, $\mathrm{Var}\,Y = 4$, $\rho_{X,Y} = 0.5$. Neka je $Z = 2XY + 1$. Odredite EZ.
Rješenje: $EZ = 2\left(\rho(X,Y)\sqrt{\mathrm{Var}\,X\,\mathrm{Var}\,Y} + EX\,EY\right) + 1 = 3$.

Zadatak 3.9. Neka su X_1 i X_2 slučajne varijable s binomnim razdiobama: $X_1 \sim \mathcal{B}(n_1, p_1)$ i $X_2 \sim \mathcal{B}(n_2, p_2)$. Odredite $E(X_1X_2)$ ako znamo da je $E(X_1 + X_2)^2 = 0.5$.
Rješenje: $E(X_1X_2) = \frac{1}{4} - \frac{1}{2}\left(n_1p_1\left(1 + p_1(n_1-1)\right) + n_2p_2\left(1 + p_2(n_2-1)\right)\right)$.

Zadatak 3.10. Dvodimenzionalan slučajni vektor (X, Y) ima distribuciju danu tablicom

X/Y	-2	-1	0	1	2
1	1/3	0	1/15	0	1/10
2	2/15	2/15	0	1/30	0
3	1/30	1/30	1/15	1/15	0

a) Odredite marginalne distribucije.
Rješenje:

hspace*1cm $X = \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 1/2 & 1/6 & 2/15 & 1/10 & 1/10 \end{pmatrix}$,

$$Y = \begin{pmatrix} 1 & 2 & 3 \\ 1/2 & 3/10 & 1/5 \end{pmatrix}$$

b) Odredite $P\{X < 2, Y \geq 2\}$ i $P\{X > -1, Y > 3\}$.
Rješenje: $P\{X < 2, Y \geq 2\} = 1/2$, $P\{X > -1, Y > 3\} = 0$.

c) Odredite $P\{|X| < 1\}$.
Rješenje: $P\{|X| < 1\} = P\{X = 0\} = 2/15$.

d) Odredite EX, EY i $E(3X + 5Y)$.
Rješenje: $EX = -13/15$, $EY = 17/10$, $E(3X + 5Y) = 59/10$.

Zadatak 3.11. Neka je $\mathbb{X} = (X, Y)$ diskretan slučajni vektor s distribucijom zadanom na sljedeći način:

$$P(X = x, Y = y) = \begin{cases} c(x+y) & , \ (x,y) \in \{0,1,2\} \times \{0,1,2\} \\ 0 & , \quad \text{inače} \end{cases}.$$

Izračunajte vrijednost konstante c.
Rješenje: $c = 1/18$.

Zadatak 3.12. Za slučajni vektor iz zadatka 3.10 odredite distribuciju od X, uz uvjet $Y = 3$, i distribuciju od Y, uz uvjet $X = -2$, te izračunajte uvjetna očekivanja $E(X|Y = 3)$ i $E(Y|X = -2)$.
Rješenje:

$$X|_{Y=3} = \begin{pmatrix} -2 & -1 & 0 & 1 \\ 1/6 & 1/6 & 1/3 & 1/3 \end{pmatrix}, \qquad E(X|Y = 3) = -\frac{1}{6};$$

$$Y|_{X=-2} = \begin{pmatrix} 1 & 2 & 3 \\ 2/3 & 1/15 & 1/15 \end{pmatrix}, \qquad E(Y|X = -2) = \frac{21}{15}.$$

Zadatak 3.13. Provjerite jesu li komponente slučajnog vektora iz zadatka 3.10 nezavisne te izračunajte njihovu kovarijancu i koeficijent korelacije.
Rješenje: *slučajne varijable X i Y nisu nezavisne,*

$$\mathrm{Cov}(X, Y) = -\frac{208}{75}, \quad \rho_{X,Y} = -\frac{416}{\sqrt{26291}}.$$

Zadatak 3.14. Dan je dvodimenzionalan slučajni vektor (X, Y) tablicom distribucije

X/Y	-2	-1	0	1	2	3
0	0.05	0.05	0.1	0	0.05	0.05
1	0.1	0.05	0.05	0.1	0	0.05
2	0.03	0.12	0.07	0.06	0.03	0.04

a) Odredite marginalne distribucije tog slučajnog vektora.
Rješenje:

$$X = \begin{pmatrix} -2 & -1 & 0 & 1 & 2 & 3 \\ 0.18 & 0.22 & 0.22 & 0.16 & 0.08 & 0.14 \end{pmatrix},$$

$$Y = \begin{pmatrix} 0 & 1 & 2 \\ 0.3 & 0.35 & 0.35 \end{pmatrix}.$$

b) Odredite EX, EY, $\mathrm{Var}\, X$ i $\mathrm{Var}\, Y$.
Rješenje: $EX = 0.16$, $EY = 1.05$, $\mathrm{Var}\, X \approx 2.65$, $\mathrm{Var}\, Y \approx 0.65$.

c) Provjerite jesu li komponente tog slučajnog vektora nezavisne.
Rješenje: slučajne varijable X i Y nisu nezavisne.

Zadatak 3.15. Slučajni pokus sastoji se od dvaju uzastopnih bacanja pravilno izrađenog novčića. Neka je (X, Y) slučajni vektor gdje slučajna varijabla X označava broj glava, a slučajna varijabla Y broj pisama koja se realiziraju u tim dvama bacanjima. Odredite distribuciju i marginalne distribucije slučajnog vektora (X, Y), uvjetnu distribuciju slučajne varijable X, uz uvjet $\{Y = 1\}$. te izračunajte koeficijent korelacije $\rho_{X,Y}$.

Rješenje:
Tablica distribucije:

X/Y	0	1	2
0	0	0	$\frac{1}{4}$
1	0	$\frac{1}{2}$	0
2	$\frac{1}{4}$	0	0

Marginalne distribucije:

$$X = \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}, \qquad Y = \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}.$$

Uvjetna distribucija: $P\{X = 1|Y = 1\} = 1$.
Koeficijent korelacije: $\rho_{X,Y} = -1$.

Zadatak 3.16. Neka je (X, Y) slučajni vektor kod kojeg slučajna varijabla X predstavlja broj izlazaka trgovačkog putnika na teren, a slučajna varijabla Y broj prodanih proizvoda. Distribucija slučajnog vektora zadana je sljedećom tablicom:

Y/X	0	1	2	3	4	5	6
1	$\frac{1}{32}$	$\frac{1}{16}$	$\frac{1}{32}$	0	0	0	0
2	0	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$	0	0
3	0	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	0
4	0	$\frac{1}{128}$	$\frac{1}{64}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{128}$

a) Odredite marginalne distribucije tog slučajnog vektora.
Rješenje:

$$X = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{32} & \frac{33}{128} & \frac{15}{64} & \frac{3}{16} & \frac{3}{16} & \frac{3}{32} & \frac{1}{128} \end{pmatrix}; \qquad Y = \begin{pmatrix} 1 & 2 & 3 & 4 \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{2} & \frac{1}{8} \end{pmatrix}.$$

b) Izračunajte EX, EY i $\rho_{X,Y}$.
Rješenje: $EX = 327/128$, $EY = 21/8$, $\rho_{X,Y} = 0.579$.

c) Odredite vjerojatnost da je trgovački putnik prodao četiri knjige, ako je dva puta izašao na teren.
Rješenje: $P\{Y = 4|X = 2\} = 1/15$.

Zadatak 3.17. Promotrimo slučajni pokus koji se sastoji od nezavisnog bacanja dvaju novčića tri puta zaredom (misli se da su realizacije na novčićima pri jednom bacanju međusobno nezavisne). Novčić A pravilno je izrađen, odnosno $P_A(G) = P_A(P) = 0.5$, ali nočić B nije i vrijedi: $P_B(G) = 0.25$, $P_B(P) = 0.75$. Neka je (X, Y) slučajni vektor u kojem X predstavlja broj glava realiziranih bacanjem novčića A, a Y broj glava realiziranih bacanjem novčića B. Odredite distribuciju i marginalne distribucije slučajnog vektora (X, Y), uvjetnu distribuciju slučajne varijable X, uz uvjet $\{Y = 2\}$, te izračunajte koeficijent korelacije $\rho_{X,Y}$.

Rješenje: $X \sim \mathcal{B}(3, 0.5)$, $X \sim \mathcal{B}(3, 0.25)$

X/Y	0	1	2	3
0	$\frac{27}{512}$	$\frac{27}{512}$	$\frac{9}{512}$	$\frac{1}{512}$
1	$\frac{81}{512}$	$\frac{81}{512}$	$\frac{27}{512}$	$\frac{3}{512}$
2	$\frac{81}{512}$	$\frac{81}{512}$	$\frac{27}{512}$	$\frac{3}{512}$
3	$\frac{27}{512}$	$\frac{27}{512}$	$\frac{9}{512}$	$\frac{1}{512}$

Zadatak 3.18. Slučajni vektor (X, Y) zadan je na sljedeći način:

$$P\{X = x_i, Y = y_j\} = \begin{cases} k(2x_i + y_j) & , \ x_i = 1, 2; \ y_j = 1, 2 \\ 0 & , \text{ inače} \end{cases}.$$

Odredite vrijednost konstante k, marginalne distribucije slučajnog vektora (X, Y) te provjerite jesu li slučajne varijable X i Y nezavisne.

Zadatak 3.19. Za slučajni vektor iz prethodnog zadatka odredite sljedeće uvjetne vjerojatnosti:

a) $P\{Y = y_j \,|\, X = x_i\}$, $P\{X = x_i \,|\, Y = y_j\}$,

b) $P\{Y = y_2 \,|\, X = x_2\}$, $P\{X = 2 \,|\, Y = 2\}$.

Zadatak 3.20. Funkcija gustoće slučajnog vektora (X, Y) dana je izrazom

$$f(x,y) = \begin{cases} 1/4 \ , & (x,y) \in [-1,1] \times [-1,1] \\ 0 \ , & \text{inače.} \end{cases}$$

a) Odredite marginalne gustoće $f_X(x)$ i $f_Y(y)$.
Rješenje: $f_X(x) = \begin{cases} 1/2 \ , & x \in [-1,1] \\ 0 \ , & \text{inače,} \end{cases} \qquad X \stackrel{D}{=} Y.$

b) Jesu li komponente tog slučajnog vektora nezavisne slučajne varijable?
Rješenje: X i Y nezavisne su.

c) Odredite $P\{0 \le X \le \frac{1}{2}, 0 \le Y \le \frac{1}{2}\}$.
Rješenje: $P\{0 \le X \le \frac{1}{2}, 0 \le Y \le \frac{1}{2}\} = 1/16$.

Zadatak 3.21. Slučajni vektor $\mathbb{X} = (X, Y)$ zadan je funkcijom gustoće

$$f_{\mathbb{X}}(x,y) = \begin{cases} k \ , & \text{za } x^2 + y^2 \le 4 \text{ i } y > 0 \\ 0 \ , & \text{za ostale } (x,y) \in \mathbb{R}^2. \end{cases}$$

a) Odredite vrijednost konstante k.
Rješenje: $k = 1/2\pi$.

b) Odredite marginalne funkcije gustoća slučajnog vektora $\mathbb{X}$.
Rješenje:

$$f_X(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4 - x^2} \ , & x \in \langle -2, 2\rangle \\ 0 \ , & \text{inače,} \end{cases} \qquad f_Y(y) = \begin{cases} \frac{1}{\pi}\sqrt{4 - y^2} \ , & y \in [0, 2\rangle \\ 0 \ , & \text{inače,} \end{cases}$$

c) Pokažite da slučajne varijable X i Y nisu nezavisne, ali jesu nekorelirane.

Zadatak 3.22. Funkcija gustoće dvodimenzionalnoga slučajnog vektora (X, Y) dana je izrazom

$$f(x,y) = \begin{cases} \dfrac{1}{\pi ab} \ , & \frac{x^2}{a^2} + \frac{y^2}{b^2} \le 1 \\ 0 \ , & \text{inače,} \end{cases}$$

a) Odredite marginalne gustoće $f_X(x)$ i $f_Y(y)$.

b) Jesu li komponente tog slučajnog vektora nezavisne slučajne varijable?

c) Odredite kovarijancu slučajnih varijabli X i Y.

Zadatak 3.23. Slučajni vektor (X, Y) zadan je funkcijom gustoće

$$f(x,y) = \begin{cases} x+y \ , \ (x,y) \in \langle 0,1\rangle \times \langle 0,1\rangle \\ 0 \quad , \ \text{inače.} \end{cases} .$$

Odredite funkciju distribucije i funkciju gustoće slučajne varijable $Z = X + Y$.
Rješenje:

$$f_Z(z) = \begin{cases} z \ , \ z \in \langle 0,1\rangle \\ 0 \ , \ \text{inače.} \end{cases} \qquad F_Z(z) = \begin{cases} 0 \quad , \ z \in \langle -\infty, 0\rangle \\ z^2/2 \ , \ z \in [0,1\rangle \\ 1 \quad , \ z \in [1,\infty\rangle . \end{cases} .$$

Zadatak 3.24. Neka je X slučajna varijabla kojom je modelirana koncentracija peludi u zraku u jutarnjem mjerenju, a Y slučajna varijabla kojom je modelirana koncentracija peludi u zraku u večernjem mjerenju. Pretpostavimo da je zajednička funkcija gustoće slučajnih varijabli X i Y (tj. funkcija gustoće slučajnog vektora $\mathbb{X} = (X, Y)$) definirana na sljedeći način:

$$f_{\mathbb{X}}(x,y) = \begin{cases} 4xy \ , \ (x,y) \in \langle 0,1\rangle \times \langle 0,1\rangle \\ 0 \quad , \qquad \text{inače.} \end{cases}$$

a) Odredite funkciju distribucije slučajnog vektora $\mathbb{X}$.
Rješenje:

$$F_{\mathbb{X}}(x,y) = \begin{cases} 1 \quad , \ (x,y) \in [1,\infty\rangle \times [1,\infty\rangle \\ x^2 \quad , \ (x,y) \in \langle 0,1\rangle \times [1,\infty\rangle \\ y^2 \quad , \ (x,y) \in [1,\infty\rangle \times \langle 0,1\rangle \\ x^2y^2 \ , \ (x,y) \in \langle 0,1\rangle \times \langle 0,1\rangle \\ 0 \quad , \qquad \text{inače.} \end{cases}$$

b) Odredite marginalne funkcije gustoća slučajnog vektora $\mathbb{X}$.
Rješenje: $X \stackrel{D}{=} Y$, $f_X(x) = \begin{cases} 2x \ , \ x \in \langle 0,1\rangle \\ 0 \quad , \ \text{inače.} \end{cases}$

c) Odredite vjerojatnost da je prosječna koncentracija peludi u zraku (temeljena samo na jutarnjem i večernjem mjerenju) manja od 0.5.
Rješenje: $P\{(X+Y)/2 < 0.5\} = 1/6$.

Zadatak 3.25. Posjed na potpuno ravnom terenu ima oblik pravokutnog trokuta s južnom granicom duljine dva kilometra i istočnom granicom duljine jedan kilometar. Koordinate točke na koju pada sjemenka javora nošena vjetrom modelirane su slučajnim vektorom $\mathbb{X} = (X, Y)$. Poznato je da ako sjemenka padne unutar posjeda, slučajni vektor $\mathbb{X}$ ima uniformnu distribuciju nad posjedom.

a) Odredite funkciju gustoće slučajnog vektora $\mathbb{X}$ te njegove marginalne funkcije gustoća. Jesu li slučajne varijable X i Y nezavisne ili ne? Obrazložite svoj odgovor.
Rješenje:

$$f_{\mathbb{X}}(x,y) = \begin{cases} 1 \ , \ (x,y) \in \{(x,y) \in \langle 0,2\rangle \times \langle 0,1\rangle : y < 0.5x\} \\ 0 \ , \qquad \text{inače.} \end{cases}$$

$$f_X(x) = \begin{cases} x/2 \ , \ x \in \langle 0,2\rangle \\ 0 \quad , \ \text{inače,} \end{cases} \qquad f_Y(y) = \begin{cases} 2(1-y) \ , \ x \in \langle 0,1\rangle \\ 0 \quad , \ \text{inače.} \end{cases}$$

Slučajne varijable X i Y nisu nezavisne.

b) Odredite uvjetne funkcije gustoća $f_{X|Y=y}(x)$ i $f_{Y|X=x}(y)$ slučajnog vektora (X, Y).
Rješenje:

$$f_{X|Y=y}(x) = \begin{cases} 1/2(1-y) & , x \in \langle 2y, 2\rangle \\ 0 & , \text{ inače,} \end{cases} \qquad f_{Y|X=x}(y) = \begin{cases} 2/x & , y \in \langle 0, x/2\rangle \\ 0 & , \text{ inače.} \end{cases}$$

c) Odredite vjerojatnost $P\{0.1 < Y \leq 0.7|X = 0.5\}$.
Rješenje: $P\{0.1 < Y \leq 0.7|X = 0.5\} = 0.6$.

Zadatak 3.26. Neprekidan slučajni vektor $\mathbb{X} = (X, Y)$ zadan je funkcijom gustoće

$$f_{\mathbb{X}}(x, y) = \begin{cases} -\cos x \sin y & , (x, y) \in [0, \pi/2] \times [-\pi/2, 0] \\ 0 & , \text{ inače.} \end{cases}$$

a) Odredite funkciju distribucije slučajnog vektora $\mathbb{X}$.
Rješenje:

$$F_{\mathbb{X}}(x, y) = \begin{cases} 1 & , (x, y) \in \langle \pi/2, \infty\rangle \times \langle 0, \infty\rangle \\ \sin(x)\cos(y) & , (x, y) \in [0, \pi/2] \times [-\pi/2, 0] \\ \cos(y) & , (x, y) \in \langle \pi/2, \infty\rangle \times [-\pi/2, 0] \\ \sin(x) & , (x, y) \in [0, \pi/2] \times \langle 0, \infty\rangle \\ 0 & , \text{ inače.} \end{cases}$$

b) Odredite marginalne funkcije distribucija slučajnog vektora $\mathbb{X}$.
Rješenje:

$$f_X(x) = \begin{cases} \cos(x) & , x \in [0, \pi/2] \\ 0 & , \text{ inače,} \end{cases} \qquad f_Y(y) = \begin{cases} -\sin(y) & , x \in [-\pi/2, 0] \\ 0 & , \text{ inače.} \end{cases}$$

c) Ispitajte nezavisnost slučajnih varijabli X i Y te odredite korelacijsku matricu slučajnog vektora $\mathbb{X}$.
Rješenje: X i Y nezavisne su slučajne varijable; korelacijska matrica jest jedinična matrica drugog reda.

d) Odredite uvjetne funkcije gustoća i uvjetne funkcije distribucija slučajnog vektora $\mathbb{X}$.
Rješenje: $f_{X|Y=y}(x) = f_X(x)$, $f_{Y|X=x}(y) = f_Y(y)$.

Zadatak 3.27. Funkcija distribucije slučajnog vektora $\mathbb{X} = (X, Y)$ jest

$$F_{\mathbb{X}}(x, y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-(x+y)} & , (x, y) \in \langle 0, \infty\rangle \times \langle 0, \infty\rangle \\ 0 & , \text{ za ostale } (x, y) \in \mathbb{R}^2. \end{cases}$$

a) Odredite marginalne funkcije distribucija toga slučajnog vektora te provjerite jesu li njegove komponente nezavisne slučajne varijable.
Rješenje: $X, Y \sim \mathcal{E}(1)$.

b) Odredite funkciju gustoće i marginalne funkcije gustoća slučajnog vektora $\mathbb{X}$.
Rješenje: $f(x, y) = \begin{cases} e^{-(x+y)} & , (x, y) \in \langle 0, \infty\rangle \times \langle 0, \infty\rangle \\ 0 & , \text{ inače.} \end{cases}$

c) Odredite uvjetne funkcije gustoća $f_{X|Y=y}(x)$ i $f_{Y|X=x}(y)$.
Rješenje: zbog nezavisnosti je slučajnih varijabli X i Y $f_{X|Y=y}(x) = f_X(x)$, $f_{Y|X=x}(y) = f_Y(y)$.

d) izračunajte $P\{0 < X \le 1, 0 < Y \le 1\}$.
Rješenje: $P\{0 < X \le 1, 0 < Y \le 1\} = 1 - \frac{2}{e} + \frac{1}{e^2}$.

Zadatak 3.28. Zadana je funkcija gustoće dvodimenzionalnog slučajnog vektora (X, Y):

$$f(x,y) = \begin{cases} e^{-(x+y)} & , (x,y) \in \langle 0,\infty\rangle \times \langle 0,\infty\rangle \\ 0 & , \text{inače.} \end{cases}$$

a) Jesu li komponente toga slučajnog vektora nezavisne slučajne varijable?
Rješenje: X i Y nezavisne su eksponencijalne slučajne varijable s parametrom $\lambda = 1$.

b) Odredite distribuciju slučajnog vektora $(X, 2Y)$.
Rješenje: Y je eksponencijalna slučajna varijabla s parametrom $\lambda = 1/2$, pa zbog nezavisnosti X i $2Y$ slijedi da je

$$f_{(X,2Y)}(x,y) = \begin{cases} \frac{1}{2}\, e^{-(x+\frac{1}{2}\, y)} & , (x,y) \in \langle 0,\infty\rangle \times \langle 0,\infty\rangle \\ 0 & , \text{inače.} \end{cases}$$

Zadatak 3.29. Neka je $\mathbb{X} = (X, Y)$ neprekidan slučajni vektor zadan funkcijom gustoće

$$f_{\mathbb{X}}(x,y) = \begin{cases} \dfrac{1}{8}(6-x-y) & , (x,y) \in \langle 0,2\rangle \times \langle 2,4\rangle \\ 0 & , \text{za ostale } (x,y) \in \mathbb{R}^2. \end{cases}$$

a) Ispitajte nezavisnost komponenti X i Y toga slučajnog vektora.

b) Odredite uvjetne funkcije gustoća $f_{X|Y=y}(x)$ i $f_{Y|X=x}(y)$.

c) Odredite matricu kovarijanci i korelacijsku matricu toga slučajnog vektora.

Zadatak 3.30. Provjerite vrijedi li slabi zakon velikih brojeva za sljedeće nizove nezavisnih slučajnih varijabli $(X_n,\, n \in \mathbb{N})$:

a) $X_n = \begin{pmatrix} -\sqrt{n} & 0 & \sqrt{n} \\ \dfrac{1}{n} & 1 - \dfrac{2}{n} & \dfrac{1}{n} \end{pmatrix}, \quad n \in \mathbb{N},$

b) $X_n = \begin{pmatrix} -3^{-n} & 3^{-n} \\ \dfrac{1}{2} & \dfrac{1}{2} \end{pmatrix}, \quad n \in \mathbb{N}.$

Rješenje: slabi zakon velikih brojeva vrijedi za oba niza slučajnih varijabli.

Zadatak 3.31. Neka su $X_1, \ldots, X_{100}$ nezavisne eksponencijalne slučajne varijable s parametrom $\lambda = 1$ te neka je

$$Y = X_1 + \ldots + X_{100} = \sum_{k=1}^{100} X_k.$$

Korištenjem centralnoga graničnog teorema aproksimirajte sljedeće vjerojatnosti:

a) $P\{Y \ge 110\}$,

b) $P\{1.1 < \overline{X}_{100} < 1.2\}$, gdje je $\overline{X}_{100} = \frac{1}{100}\, Y$.

Rješenje:

a) $P\{Y \geq 110\} \approx 0.1587$,

b) $P\{1.1 < \overline{X}_{100} < 1.2\} \approx 0.136$.

Zadatak 3.32. Zadane su nezavisne jednako distribuirane slučajne varijable $X_1, \ldots, X_n$ s matematičkim očekivanjem μ i varijancom σ^2. Izračunajte

$$E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\right).$$

Poglavlje 4

Statistika

Korištenje je riječi **statistika** u svakodnevnom životu najčešće povezano s brojčanim vrijednostima kojima pokušavamo opisati bitne karakteristike nekog skupa podataka.

Statistika, kao znanstvena disciplina, bavi se razvojem metoda prikupljanja, opisivanja i analiziranja podataka te primjenom tih metoda u procesu donošenja zaključaka na temelju prikupljenih podataka.

Statističko istraživanje fokusirano je na skup **objekata**, tj. **jedinki** (ljudi, životinja, biljaka, stvari, država, gradova, poduzeća itd.) i skup odabranih veličina koje se na njima promatraju. Veličine koje se na jedinkama promatraju zovemo **varijablama**. Sve jedinke koje se žele obuhvatiti istraživanjem, tj. o kojima se želi zaključivati, čine **populaciju**. Podaci se prikupljaju da bi se zaključivalo o varijablama na populaciji.

Prije provođenja istraživanja populacija mora biti precizno opisana. Na primjer, ako je varijabla od interesa broj riječi koji je osoba u stanju pročitati u minuti, onda valja precizirati čine li populaciju djeca predškolskog uzrasta u Hrvatskoj ili djeca prvog razreda osnovne škole u Hrvatskoj ili djeca prvog razreda neke specifične škole u Hrvatskoj (npr. O.Š. Tin Ujević u Osijeku) ili djeca prvog razreda osnovne škole u Hrvatskoj koja idu po posebnom programu itd. U svakom slučaju, populacija mora biti precizirana.

Iz definirane populacije uzima se u statističko istraživanje **uzorak**. Podaci moraju biti prikupljeni na uzorku koji je reprezentativan za opisanu populaciju. Reprezentativan uzorak mora odražavati populaciju tj. u njemu trebaju biti zastupljene sve tipične karakteristike populacije. Najčešći je način odabira jedinki iz populacije u reprezentativan uzorak tzv. **slučajni uzorak**, tj. takav izbor u kojemu svaka jedinka ima jednaku šansu biti izabrana u uzorak.

Izbor jedinki u reprezentativan uzorak iz populacije može se postići pomoću generatora slučajnih brojeva. Da bi ilustrirali tu proceduru, označimo s M broj jedinki u populaciji i numerirajmo jedinke brojevima od 1 do M. Neka je n broj jedinki koje želimo uzeti u uzorak. Potrebno je generirati n međusobno različitih realizacija diskretne uniformne slučajne varijable na skupu $\{1, \ldots, M\}$ i uzeti u uzorak jedinke koje za oznaku imaju dobivene prirodne brojeve. Treba naglasiti da tu proceduru nije uvijek jednostavno provesti u primjeni.

Zadatak 4.1. U studentskoj referadi pribavite popis studenata koji slušaju UVIS ove godine. Numerirajte studente i provedite proceduru izbora 40 studenata u slučajni uzorak. Od svakog studenta odabranog u uzorak prikupite podatke o spolu, broju do sada položenih ispita, broju do sada ostvarenih ECTS bodova, prosječnoj ocjeni svih do sada položenih ispita, ocjeni iz predmeta Diferencijalni račun i Integralni račun. Formirajte bazu podataka u nekom programskom paketu.

U okviru statističke teorije valja se pozabaviti sljedećim metodama:

- metodama prikupljanja podataka,
- metodama opisivanja skupa podataka (deskriptivna statistika),
- metodama statističkog zaključivanja.

Metode prikupljanja podataka ovdje nećemo posebno proučavati. Ipak, valja napomenuti da se podaci mogu prikupiti npr. na osnovu javnih izvora (knjige, časopisi, novine, web), dizajniranog eksperimenta, anketa, promatranjem itd.

4.1 Deskriptivna statistika

U statističkim istraživanjima razlikujemo nekoliko osnovnih tipova varijabli koje se međusobno razlikuju po svojstvima vrijednosti koje mogu poprimiti.

Kvalitativne varijable

Karakteristika je kvalitativnih varijabli da njihove vrijednosti nisu, po svojim svojstvima korištenim u istraživanju, realni brojevi. Tipičan je primjer takve varijable spol osobe. Vrijednosti kvalitativne varijable uobičajeno nazivamo kategorijama. Kategorije kvalitativnih varijabli mogu biti definirane u skladu s potrebama statističkog istraživanja.

Primjer 4.1. *Sljedeće su varijable kvalitativnog tipa:*

- *radna mjesta u školi (spremačica, domar, tajnik, nastavnik, pedagog, ravnatelj),*
- *boja očiju (plava, smeđa, zelena),*
- *krvne grupe (A, B, AB, 0),*
- *spol (m ili ž).*

Numeričke varijable

Numeričke varijable prirodno primaju vrijednosti iz skupa **realnih brojeva**. Tipični su primjeri numeričkih varijabli masa i visina osobe. Međutim, treba naglasiti da se i kategorije kvalitativnih varijabli mogu izražavati brojevima što ih ne čini numeričkim varijablama. Primjerice, spol osobe jedna je kvalitativna varijabla. Kategoriju "ženski spol" možemo označiti s "1", a kategoriju "muški spol" s "2", što može biti korisno prilikom unošenja podataka u bazu. Time smo kategorijama kvalitativne varijable pridružili numeričke vrijednosti, ali samu varijablu nismo učinili numeričkom po njezinim svojstvima.

Primjer 4.2. *Sljedeće su varijable numeričkog tipa:*

- *postotak prolaznosti na pojedinim ispitima tijekom jedne akademske godine,*
- *broj bodova na državnoj maturi iz matematike,*
- *broj ulovljenih komaraca u klopku,*
- *temperatura mora,*
- *koncentracija soli u morskoj vodi.*

Među numeričkim varijablama razlikujemo diskretne i kontinuirane varijable.

Diskretne numeričke varijable mogu poprimiti samo konačno ili prebrojivo mnogo vrijednosti.

Primjer 4.3. *Sljedeće su numeričke varijable diskretne:*

- *broj bodova na državnoj maturi iz matematike,*
- *broj ulovljenih komaraca u klopku,*
- *broj dana u godini s temperaturom zraka većom od $35^o C$.*

Skup je mogućih vrijednosti kontinuiranih numeričkih varijabli cijeli skup realnih brojeva ili neki interval.

Primjer 4.4. *Sljedeće su numeričke varijable kontinuirane:*

- *postotak prolaznosti na pojedinim ispitima tijekom jedne akademske godine,*
- *temperatura mora,*
- *vodostaj neke rijeke.*

Primjer 4.5. (djelatnici.xls)
Baza podataka djelatnici.xls *sadrži podatke o uzorcima djelatnika dviju konkurentskih tvornica - tvornice A i tvornice B. U tablici s imenom "tvornica A" zabilježene su vrijednosti sljedećih varijabli za djelatnike* **tvornice** *A:*

spol - *kvalitativna varijabla koja sadrži informaciju o spolu (*M *- muški spol,* Z *- ženski spol),*

odjel - *kvalitativna varijabla sadrži naziv odjela u kojemu je djelatnik zaposlen (*TR *- transport,* P- *pakiranje,* IS *- isporuka),*

obrazovanje - *kvalitativna varijabla koja sadrži stručnu spremu djelatnika (*SSS *- srednja stručna sprema,* VŠSS *- viša stručna sprema,* VSS *- visoka stručna sprema),*

dob - *kontinuirana numerička varijabla koja sadrži starost djelatnika u godinama,*

visina - *kontinuirana numerička varijabla koja sadrži visinu djelatnika u centimetrima,*

rukovodstvo - *diskretna numerička varijabla koja sadrži broj godina rada koje je djelatnik proveo na nekoj od rukovodećih pozicija u toj tvornici,*

placa_prije - *kontinuirana numerička varijabla koja sadrži iznos godišnje plaće djelatnika prije reorganizacije poslovnog sustava,*

placa_poslije - *kontinuirana numerička varijabla koja sadrži iznos godišnje plaće djelatnika nakon reorganizacije poslovnog sustava.*

U tablici s imenom "tvornica B", u varijabli placa_konkurencija*, zabilježeni su iznosi godišnje plaće za svakog djelatnika iz uzorka iz* **tvornice** *B.*

U svrhu prikaza podataka i nekih statističkih analiza, vrijednosti se numeričke varijable također mogu svrstati u **kategorije**. Za razliku od kategorija kvalitativnih varijabli, među kategorijama se numeričke varijable uvijek može prepoznati prirodan poredak.

Primjer 4.6. (auto-centar.xls)
Svrha je ovog primjera prikazati mogućnost kategorizacije numeričke varijable. Taj se postupak najčešće provodi stvaranjem nove varijable čije su vrijednosti kategorije kojih je (znatno) manje nego svih mogućih vrijednosti odgovarajuće numeričke varijable. Baza podataka auto-centar.xls *sastoji se od sljedećih varijabli:*

automobili - *diskretna numerička varijabla koja sadrži podatke o broju prodanih automobila u jednom danu za sto promatranih dana. Budući da broj prodanih automobila u jednom danu može biti vrlo mali (npr. samo nekoliko osobnih automobila), ali i vrlo velik (npr. narudžbe automobila za vozni park nekog poduzeća), zaključujemo da diskretna numerička varijabla* automobili *može poprimiti velik broj različitih vrijednosti iz skupa prirodnih brojeva. Zato je u nekim situacijama korisno kategorizirati vrijednosti te varijable prema točno određenom kriteriju. Na primjer, kategorizacija broja prodanih automobila u jednome danu može se napraviti kao što je prikazano varijablom* kategorija.

kategorija - *kvalitativna varijabla koja podatke iz varijable* automobili *svrstava u pet kategorija prema kriteriju prikazanom u tablici 4.1.*

broj prodanih automobila	kategorija
0 - 9	E
10 i 11	D
12 i 13	C
14 i 15	B
16 i više	A

Tablica 4.1: Primjer kategorizacije numeričke varijable automobili

Ordinalne varijable

Karakteristika je ordinalnih varijabli da su one, po svom karakteru, kvalitativne, ali među kategorijama se može uspostaviti prirodan poredak. Tipičan su primjeri takvih varijabli stručna sprema osobe i ocjena u školi.

Primjer 4.7. (matematika.sta)
Baza podataka matematika.sta *sadrži podatke prikupljene anketiranjem studenata nakon održanih predavanja, vježbi, kolokvija te usmenog ispita iz jednog matematičkog kolegija. Prikupljeni podaci organizirani su na sljedeći način:*

prosjek - *varijabla koja sadrži podatke o prosječnoj ocjeni studiranja za 49 anketiranih studenata,*

polozeno - *varijabla koja studente svrstava u dvije kategorije s obzirom na to jesu li položili ispit iz promatranog kolegija prema kriteriju prikazanom u tablici 4.2.*

položen/nepoložen ispit	kategorija
položen ispit	1
nepoložen ispit	0

Tablica 4.2: Kategorizacija studenata prema položenosti ispita.

predavanja, vjezbe - *dvije varijable koje prisutnost studenata na predavanjima/vježbama (p/v) svrstavaju u tri kategorije na način prikazan u tablici 4.3.*

prisutnost studenta na p/v	kategorija
student s p/v nije nikada izostao	1
student je s p/v izostao samo jednom	2
student je s p/v izostao barem dva puta	3

Tablica 4.3: Kategorizacija studenata prema broju izostanaka s predavanja/vježbi.

tezina kolegija, materijali - *dvije varijable koje sadrže subjektivne ocjene (u standardnoj skali od 1 do 5) studenata o težini kolegija i dostatnosti dostupnih materijala za pripremanje ispita iz promatranog kolegija.*

Uočimo da se varijabla prosjek *može promatrati kao neprekidna numerička varijabla, varijabla* položeno *jest kvalitativna, dok se varijable* predavanja, vjezbe, tezina kolegija *i* materijali *mogu svrstati u ordinalne varijable.*

4.1.1 Metode opisivanja kvalitativnih varijabli

Vrijednosti kvalitativne varijable jesu kategorije. Mjere kojima opisujemo zastupljenost jedne kategorije u uzorku jesu **frekvencija** kategorije i **relativna frekvencija** kategorije.

Frekvencija kategorije broj je izmjerenih vrijednosti varijable koje pripadaju danoj kategoriji.

Relativna frekvencija kategorije broj je izmjerenih vrijednosti varijable koje pripadaju danoj kategoriji podijeljen ukupnim brojem izmjerenih vrijednosti za ispitivanu varijablu.

Pretpostavimo da varijabla može primiti vrijednosti k različitih kategorija, a da se u podacima nalazi n izmjerenih vrijednosti za tu varijablu. Frekvenciju i-te kategorije označit ćemo f_i, a relatvnu frekvenciju dobijemo kao

$$\frac{f_i}{n}.$$

Relativna frekvencija kategorije mjera je zastupljenosti koja daje informaciju o udjelu kategorije u uzorku poznate veličine i često se izražava kao postotak. Frekvencije i relativne frekvencije pojedinih kategorija prikazujemo tablično i grafički.

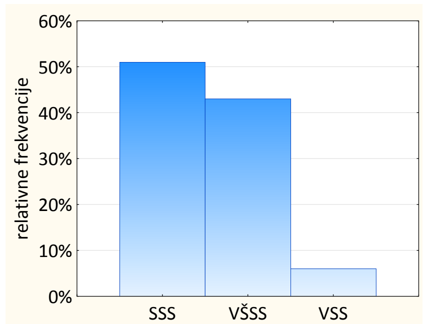
Frekvencije i relativne frekvencije kategorija kvalitativnih varijabli grafički prikazujemo pomoću **stupčastog dijagrama frekvencija** i **stupčastog dijagrama relativnih frekvencija**. U istu svrhu može se koristiti i **kružni dijagram** frekvencija i relativnih frekvencija . Popularan je naziv za isti grafički prikaz "pita".

Primjer 4.8. (djelatnici.xls)
Baza podataka djelatnici.xls *opisana je u primjeru 4.5. Promotrimo kvalitativnu varijablu* obrazovanje *čije su vrijednosti svrstane u tri kategorije:* SSS - *srednja stručna sprema,* VŠSS - *viša stručna sprema,* VSS - *visoka stručna sprema. Zastupljenost tih kategorija u promatranom uzorku od* 100 *djelatnika opisana je tablicom frekvencija i relativnih frekvencija 4.4.*

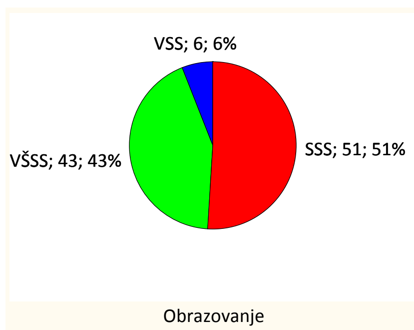
kategorija	frekvencija	relativna frekvencija
SSS	51	$51/100 = 0.51$
VŠSS	43	$43/100 = 0.43$
VSS	6	$6/100 = 0.06$

Tablica 4.4: Tablica frekvencija i relativnih frekvencija svih kategorija varijable obrazovanje.

Grafički prikazi relativnih frekvencija dani su u obliku stupčastog dijagrama na slici 4.1, a prikazi frekvencija i relativnih frekvencija u obliku kružnog dijagrama na slici 4.2.



Slika 4.1: Stupčasti dijagram relativnih frekvencija svih kategorija varijable obrazovanje.



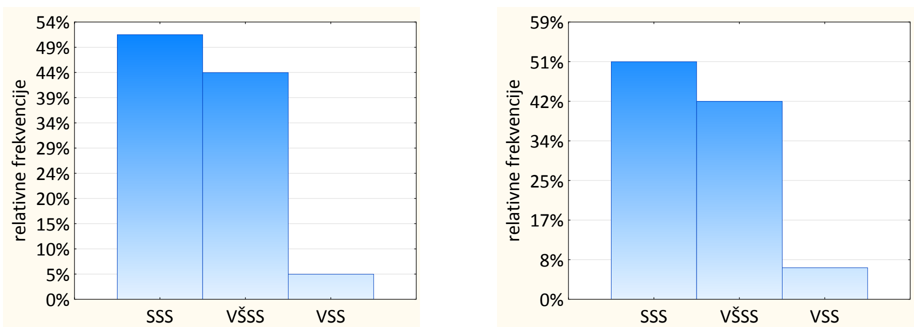
Slika 4.2: Kružni dijagram frekvencija i relativnih frekvencija svih kategorija varijable obrazovanje.

Primjer 4.9. (djelatnici.xls)
Često se u praksi pokazuje korisnim poznavanje zastupljenosti kategorija jedne varijable za svaku od kategorija neke druge kvalitativne varijable proučavane na istom uzorku. U ovom ćemo primjeru tablično i grafički prikazati frekvencije i relativne frekvencije svih kategorija varijable obrazovanje *posebno za ispitanike ženskog spola, a posebno za ispitanike muškog spola iz promatranog uzorka djelatnika.*

Tablice tako kategoriziranih frekvencija i relativnih frekvencija varijable obrazovanje *prikazane su u tablici 4.5. Stupčasti dijagrami relativnih frekvencija svih kategorija varijable* obrazovanje *za kategorije* Z *i* M *varijable* spol *prikazani su na slici 4.3, a kružni dijagrami frekvencija i relativnih frekvencija na slici 4.4.*

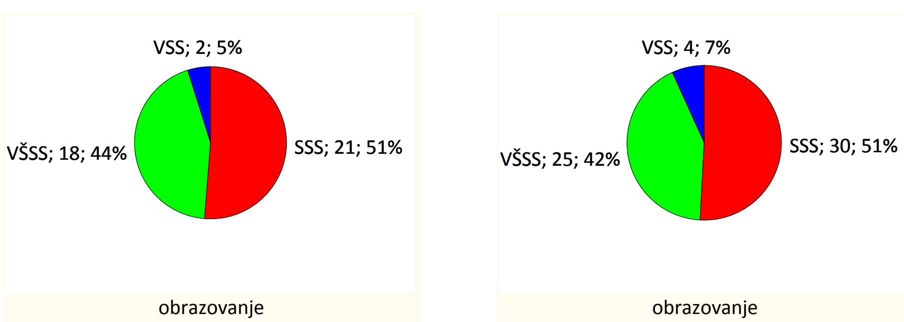
kategorija	spol=Z		spol=M	
	frekvencija	relativna frekvencija	frekvencija	relativna frekvencija
SSS	21	$21/41 = 0.5122$	30	$30/59 = 0.5085$
VŠSS	18	$18/41 = 0.4390$	25	$25/59 = 0.4237$
VSS	2	$2/41 = 0.0488$	4	$4/59 = 0.0678$

Tablica 4.5: Tablica frekvencija i relativnih frekvencija svih kategorija varijable obrazovanje posebno za svaku kategoriju varijable spol.



(a) spol=Z (b) spol=M

Slika 4.3: Stupčasti dijagrami relativnih frekvencija svih kategorija varijable obrazovanje posebno za svaku kategoriju varijable spol.



(a) spol=Z (b) spol=M

Slika 4.4: Kružni dijagram frekvencija i relativnih frekvencija svih kategorija varijable obrazovanje posebno za svaku kategoriju varijable spol.

4.1.2 Metode opisivanja numeričkih varijabli

Numeričke varijable po svojoj prirodi mogu biti diskretne i neprekidne. U oba slučaja, a posebno kod neprekidnih varijabli, može se dogoditi da u prikupljenim podacima postoji mnogo međusobno različitih vrijednosti. U takvim slučajevima tablični i grafički prikazi uvedeni za kvalitativne varijable mogu biti nedovoljno informativni.

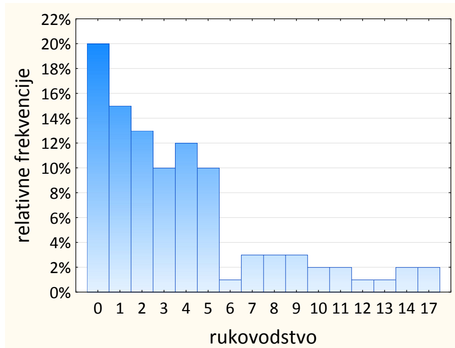
Ako su numeričke varijable diskretne s malo mogućih vrijednosti, za opis podataka možemo koristiti iste metode kao pri opisivanju kvalitativnih podataka, tj. frekvencije i relativne frekvencije, te ih grafički prikazivati stupčastim dijagramima i kružnim dijagramima.

Primjer 4.10. (djelatnici.xls)
Broj zabilježenih vrijednosti diskretnih numeričkih varijabli znatno utječe na preglednost tabličnih i grafičkih prikaza frekvencija i relativnih frekvencija. Analizirajmo diskretnu numeričku varijablu rukovodstvo *baze podataka* djelatnici.xls *iz primjera 4.5 koja prima vrijednosti iz skupa* $\{0, 1, \ldots, N\}$*, gdje je* N *najveći mogući broj godina radnog staža koje djelatnik može provesti na rukovodećoj poziciji. Frekvencije i relativne frekvencije zabilježenih vrijednosti varijable* rukovodstvo *za djelatnike iz promatranog uzorka dane su u tablici 4.6.*

rukovodstvo	frek.	rel. frek.	rukovodstvo	frek.	rel. frek.
0	20	0.20	8	3	0.03
1	15	0.15	9	3	0.03
2	13	0.13	10	2	0.02
3	10	0.10	11	2	0.02
4	12	0.12	12	1	0.01
5	10	0.10	13	1	0.01
6	1	0.01	14	2	0.02
7	3	0.03	17	2	0.02

Tablica 4.6: Tablica frekvencija i relativnih frekvencija svih zabilježenih vrijednosti varijable rukovodstvo.

Vidimo da se varijabla rukovodstvo *na promatranom uzorku djelanika realizirala sa* 16 *različitih vrijednosti čije su relativne frekvencije grafički prikazane stupčastim dijagramom 4.5.*



Slika 4.5: Stupčasti dijagram relativnih frekvencija svih zabilježenih vrijednosti varijable rukovodstvo.

Ako numerička varijabla prima mnogo međusobno različitih vrijednosti, za prikazivanje skupa izmjerenih vrijednosti obično nam neće puno pomoći frekvencije te stupčasti ili kružni dijagrami napravljeni na osnovu svake pojedine izmjerene vrijednosti. Takvi se slučajevi često javljaju ako podaci dolaze iz kontinuiranih numeričkih varijabli. Problem ćemo ilustrirati sljedećim primjerom.

Primjer 4.11. (djelatnici.xls)
Promotrimo varijablu placa_prije *iz baze podataka* djelatnici.xls *koja sadrži godišnje plaće prije reorganizacije poslovanja poduzeća za uzorak od* 100 *djelatnika. Ovdje možemo pretpostaviti da se radi o kontinuiranoj numeričkoj varijabli koja može primiti vrijednosti iz intervala* $\langle 0, x]$. *Kao gornju granicu intervala, tj.* x, *možemo uzeti realan broja za koji pretpostavljamo da u danim uvjetima predstavlja najveću moguću godišnju plaću u tom poduzeću.*

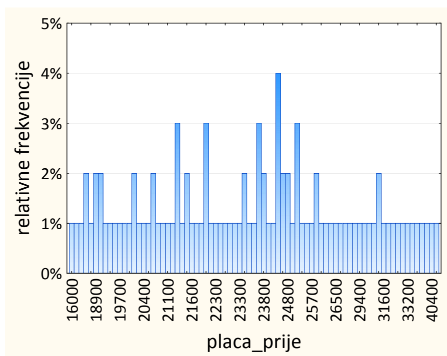
Dio tablice frekvencija i relativnih frekvencija na temelju svih realiziranih vrijednosti prikazan je tablicom 4.7. Od 100 *realizacija zabilježeno je čak* 77 *različitih vrijednosti, a pojedine se vrijednosti pojavljuju s izrazito malim frekvencijama:* 1, 2, 3 *ili* 4. *Iz takvih je tablica teško očitavati frekvencije koje nas zapravo zanimaju (npr. frekvencija ispitanika s godišnjom plaćom manjom od* 20000*). Zbog toga se u tablice uobičajeno dodaje i stupac kumulativnih frekvencija i kumulativnih relativnih frekvencija, kako je prikazano tablicom 4.7.*

Stupčasti dijagram (slika 4.6) ili kružni dijagram frekvencija (relativnih frekvencija) na kojemu su prikazane sve realizirane vrijednosti s odgovarajućom frekvencijom (relativnom frekvencijom), u takvim slučajevima nije osobito informativan.

Ako imamo podatke $(x_1, \ldots, x_n)$, *onda je* **kumulativna frekvencija** *podatka* x_i, $i = 1, \ldots, n$, *broj svih podataka iz* $(x_1, \ldots, x_n)$ *koji su manji ili jednaki* x_i. *Analogno se definira i kumulativna relativna frekvencija podatka.*

iznos plaće	frekvencija	kumulativna frek.	relativna frekvencija	kumulativna rel. frek.
16000	1	1	0.01	0.01
⋮	⋮	⋮	⋮	⋮
19800	1	15	0.01	0.15
20000	1	16	0.01	0.16
20200	2	18	0.02	0.18
⋮	⋮	⋮	⋮	⋮
42400	1	100	0.01	1

Tablica 4.7: Tablica frekvencija i relativnih frekvencija svih sto zabilježenih vrijednosti varijable placa_prije.



Slika 4.6: Stupčasti dijagram relativnih frekvencija svih realiziranih vrijednosti varijable placa_prije.

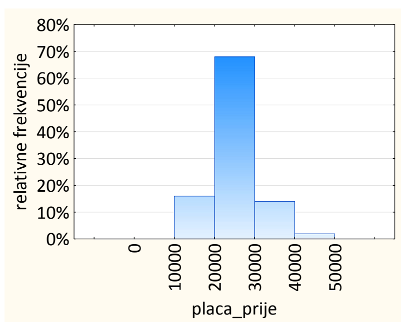
U svrhu dobivanja preglednih i korisnih stupčastih dijagrama za podatke iz kontinuiranih numeričkih varijabli, izmjerene je vrijednosti potrebno kategorizirati. To znači da velik skup podataka podijelimo u nekoliko disjunktnih intervala po kriteriju za koji smatramo da će nam dati željene rezultate. Stupčasti dijagram tada smještamo u koordinatni sustav tako da prikazujemo stupiće nad tim intervalima s površinom koja odgovara relativnoj frekvenciji podataka sadržanih u odgovarajućem intervalu. Duljina intervala informacija je koja je također prikazana stupčastim dijagramom jer odgovara širini stupića. Takav stupčasti dijagram zovemo **histogram**.

Primjer 4.12. (djelatnici.xls)
Promotrimo ponovno varijablu placa_prije *iz baze podataka* djelatnici.xls*. Razvrstajmo vrijednosti u disjunktne intervale duljine* 10000 *počevši od nule. Tako dobiveni tablični prikaz frekvencija i relativnih frekvencija dan je tablicom 4.8, a pripadni histogram slikom 4.7. Takav histogram jasno*

ilustrira činjenicu da najviše djelatnika u uzorku ima godišnju plaću od 20000 *do* 30000 *novčanih jedinica, dok je plaća iz intervala* 40000 *do* 50000 *rijetkost. Intervale za kategorizaciju u takvim i sličnim slučajevima obično radimo tako da bi zadovoljili potrebe za prezentiranjem informacija koje želimo istaknuti.*

iznos plaće	frekvencija	relativna frekvencija
$[0, 10000\rangle$	0	0
$[10000, 20000\rangle$	15	0.15
$[20000, 30000\rangle$	69	0.69
$[30000, 40000\rangle$	14	0.14
$[40000, 50000\rangle$	2	0.02

Tablica 4.8: Tablica frekvencija i relativnih frekvencija kategoriziranih izmjerenih vrijednosti varijable placa_prije.



Slika 4.7: Histogram relativnih frekvencija kategoriziranih izmjerenih vrijednosti varijable placa_prije.

Za numeričke varijable možemo definirati numeričke karakteristike koje imaju logičnu interpretaciju i mogu se iskoristiti s ciljem prikazivanja skupa podataka. Ovdje ćemo definirati neke od najčešće korištenih numeričkih karakteristika skupa podataka.

Aritmetička sredina podataka

Aritmetička sredina niza izmjerenih vrijednosti $x_1, x_2, \ldots, x_n$ varijable X definirana je izrazom

$$\overline{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Aritmetička sredina numerička je karakteristika koja pripada mjerama centralne tendencije, tj. mjeri "srednju vrijednost" podataka.

Primjer 4.13. *Neka su izmjerene vrijednosti jedne varijable sljedeće:*

$$1.2, 2.1, 3.2, 4.3, 5.4, 6.5, 7.6, 8.7, 9.8.$$

S obzirom da ih ima ukupno 9*, aritmetička sredina (eng. mean) tog skupa izmjerenih vrijednosti jest*

$$\frac{1.2 + 2.1 + 3.2 + 4.3 + 5.4 + 6.5 + 7.6 + 8.7 + 9.8}{9} \approx 5.42.$$

Medijan podataka

Da bismo razumjeli i odredili medijan potrebno je prvo poredati izmjerene vrijednosti $x_1, x_2, \ldots, x_n$ varijable X po veličini (u rastućem poretku, tj. od manjeg prema većem). Medijan je također jedna mjera centralne tendencije numeričkih podataka, a ima značenje izmjerene vrijednosti koja se nalazi na sredini niza podataka kada je on uređen po veličini, tj. barem pola podataka manje je ili jednako medijanu, a istovremeno je barem pola podataka veće ili jednako medijanu. Način njegovog izračuna ovisi o tome imamo li **neparan** ili **paran** broj izmjerenih vrijednosti varijable. Ukoliko imamo **neparan broj** izmjerenih vrijednosti, onda postoji vrijednost koja je na srednjoj poziciji u uređenom skupu, pa nju definiramo kao medijan.

Primjer 4.14. *Neka su izmjerene vrijednosti jedne varijable sljedeće:*

$$1, 2, 5, 6, 5, 1, 2, 7, 2, 2, 3.$$

Prvo te vrijednosti poredamo po veličini:

$$1, 1, 2, 2, 2, \mathbf{2}, 3, 5, 5, 6, 7.$$

S obzirom da ih ima ukupno 11*, medijan je vrijednost koja je na šestoj poziciji u tako dobivenom nizu, tj. broj* 2.

Ukoliko imamo **paran broj** izmjerenih vrijednosti varijable, onda ne postoji podatak koji je na srednjoj poziciji jer srednju poziciju "zauzimaju" dva podatka. Medijan se tada definira kao polovište tih dvaju podataka (tj. aritmetička sredina tih dvaju podataka).

Primjer 4.15. *Neka su izmjerene vrijednosti jedne varijable sljedeće:*

$$1, 2, 5, 6, 5, 1, 2, 7, 2, 2, 3, 3.$$

Prvo te vrijednosti poredamo po veličini:

$$1, 1, 2, 2, 2, \mathbf{2}, \mathbf{3}, 3, 5, 5, 6, 7.$$

S obzirom da ima 12 *podataka, "sredinu" čine šesti i sedmi podatak, tj. vrijednosti* 2 *i* 3*. Medijan je tog skupa podataka sredina tih dvaju brojeva, tj. medijan je* $(2 + 3)/2 = 2.5$.

Postotna vrijednost podataka, donji i gornji kvartil

Medijan odgovara pedeset postotnoj vrijednosti s obzirom da je barem 50% podataka manje ili jednako medijanu i barem 50% podataka veće ili jednako od medijana. Postotna vrijednost za neki izabrani broj $p \in \langle 0, 100\rangle$, označimo je x'_p, definira se poštujući zahtjev da je barem $p\%$ izmjerenih vrijednosti manje ili jednako x'_p, dok je barem $(100 - p)\%$ vrijednosti veće ili jednako x'_p. Dvadeset pet postotna vrijednost zove se donji kvartil, a sedamdeset pet postotna vrijednost zove se gornji kvartil. Analogno, kao i kod računanja medijana, ako se na traženoj poziciji za računanje postotne vrijednosti nalaze dva podatka u uređenom skupu izmjerenih vrijednosti, postotnu vrijednost određujemo kao njihovu sredinu. Donji i gornji kvartil mjere su koje pripadaju grupi mjera raspršenosti podataka.

Primjer 4.16. *Neka su izmjerene vrijednosti jedne varijable sljedeće:*

$$1, 2, 5, 6, 6, 1, 3, 7, 3, 3, 3, 3.$$

Prvo te vrijednosti poredamo po veličini:

$$1, 1, 2, 3, 3, 3, 3, 3, 5, 6, 6, 7.$$

*Želimo li odrediti donji kvartil, potrebno je prvo odrediti četvrtinu podataka (*25%*). S obzirom da imamo* 12 *podataka, četvrtinu (*25%*) čine tri podatka. Treći je podatak u gornjem skupu broj* 2*, a četvrti* 3*. Donji kvartil jest* 2.5*. Deveti je broj u gornjem skupu podataka broj* 5*, a deseti* 6*, stoga je gornji kvartil* 5.5.

Najmanja i najveća vrijednost, raspon podataka

Raspon podataka mjera je koja pokazuje koliko su numerički podaci raspršeni, tj. to je jedna od mjera raspršenosti podataka. Definiran je kao razlika najveće i najmanje vrijednosti u skupu mjerenih vrijednosti varijable (tj. razlika maksimalne i minimalne izmjerene vrijednosti varijable). Ako su $x_1, x_2, \ldots, x_n$ izmjerene vrijednosti varijable, označimo najmanju od njih (minimum) $x_{\min}$, a najveću (maksimum) $x_{\max}$.

Primjer 4.17. *Neka su izmjerene vrijednosti jedne varijable sljedeće:*

$$1, 2, 5, 6, 5, 1, 2, 7, 2, 2, 3, 3.$$

Vidimo da je vrijednost 1 *najmanja izmjerena vrijednost, a* 7 *najveća. Prema tome, raspon ovog skupa izmjerenih vrijednosti je* $7 - 1 = 6$.

U mnogim je primjerima zanimljivo promatrati **maksimalno odstupanje izmjerenih vrijednosti varijable od "prosjeka", tj. aritmetičke sredine** izmjerenih vrijednosti. Ta je numerička karakteristika definirana kao veći od brojeva $(\overline{x}_n - x_{\min})$ i $(x_{\max} - \overline{x}_n)$, tj. broj

$$\max\{(\overline{x}_n - x_{\min}), (x_{\max} - \overline{x}_n)\}.$$

Primjer 4.18. *Neka su* $1, 2, 5, 6, 5, 1, 2, 7, 2, 2, 3, 3$ *izmjerene vrijednosti neke varijable. Tada je*

$$x_{\min} = 1, \quad x_{\max} = 7, \quad \overline{x}_n = \frac{1+2+5+6+5+1+2+7+2+2+3+3}{12} = 3.25.$$

Maksimalno odstupanje izmjerenih vijednosti te varijable od njihovog prosjeka jest

$$\max\{3.25 - 1, 7 - 3.25\} = \max\{2.25, 3.75\} = 3.75.$$

Varijanca i standardna devijacija podataka

Varijanca i standardna devijacija također pripadaju grupi mjera raspršenosti podataka. One karakteriziraju raspršenost podataka oko aritmetičke sredine. Varijanca niza izmjerenih vrijednosti $x_1, x_2, \ldots, x_n$ varijable definirana je izrazom:

$$\overline{s}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2,$$

a standardna devijacija jest kvadratni korijen varijance, tj.

$$\overline{s}_n = \sqrt{\overline{s}_n^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2}.$$

Primjer 4.19. *Neka su izmjerene vrijednosti jedne varijable sljedeće:*

$$1.2, 2.1, 3.2, 4.3, 5.4, 6.5, 7.6, 8.7, 9.8.$$

Iz primjera 4.13 znamo da je aritmetička sredina tog skupa podataka približno jednaka 5.42. *Varijanca tog skupa podataka jest*

$$\overline{s}_n^2 = \frac{1}{9}\sum_{i=1}^{9}(x_i - 5.42)^2 \approx 7.87,$$

a standardna devijacija

$$\overline{s}_n = \sqrt{\frac{1}{9}\sum_{i=1}^{9}(x_i - 5.42)^2} \approx 2.81.$$

Mod podataka

Mod je vrijednost iz niza izmjerenih vrijednosti varijable kojoj pripada najveća frekvencija, tj. izmjerena je najviše puta. Mod ne mora biti jedinstven.

Primjer 4.20. *Neka su izmjerene vrijednosti jedne varijable sljedeće:*

$$1, 2, 5, 6, 5, 1, 2, 7, 2, 2, 3, 3.$$

Vidimo da je vrijednost 2 *izmjerena najviše puta (četiri puta), pa je* 2 *mod tog skupa podataka.*

Primjer 4.21. *Neka su izmjerene vrijednosti jedne varijable sljedeće:*

$$1, 2, 5, 6, 5, 3, 1, 2, 7, 2, 2, 3, 3.$$

Vidimo da su najviše puta izmjerene dvije vrijednosi - 2 i 3 izmjerene su točno četiri puta. Dakle, mod tog skupa podataka nije jedinstven.

Korištenjem numeričkih karakteristika numeričkih varijabli skup mjerenih vrijednosti može se prikazati grafički pomoću **kutijastog dijagrama** (eng. *box plot*, *boxplot* ili *box-and-whisker plot*).

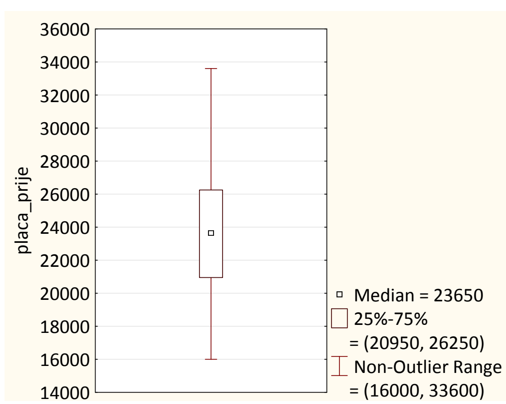
Kutijastim dijagramom prikazujemo odnos pet numeričkih karakteristika skupa izmjerenih vrijednosti: minimalnu vrijednost, donji kvartil, medijan, gornji kvartil i maksimalnu vrijednost.

Primjer 4.22. (djelatnici.xls)
Numeričke karakteristike skupa izmjerenih vrijednosti varijable placa_prije *iz baze podataka* djelatnici.xls *prikazane su u tablici 4.9.*

veličina uzorka	aritmetička sredina	mod	frek. moda	st. dev.	varijanca
100	24522	24600	4	5105.801	26069208.1
minimum	donji kvartil	medijan	gornji kvartil	maksimum	raspon
16000	20950	23650	26250	42400	26400

Tablica 4.9: Deskriptivna statistika varijable placa_prije.

Odnos minimuma, donjeg kvartila, medijana, gornjeg kvartila i maksimuma izmjerenih vrijednosti varijable placa_prije *prikazan je kutijastim dijagramom 4.8.*



Slika 4.8: Kutijasti dijagram na bazi medijana za izmjerene vrijednosti varijable placa_prije.

Iz tablice 4.9 i kutijastog dijagrama 4.8 možemo izvesti sljedeće i slične zaključke:

- *najniža godišnja plaća u uzorku iznosi* 16000*, a najviša* 42400*,*
- *bar* 25% *ispitanika iz uzorka ima plaću manju ili jednaku* 20950*,*
- *bar* 25% *ispitanika iz uzorka ima plaću veću ili jednaku* 26250*,*
- *bar* 50% *ispitanika iz uzorka ima plaću manju ili jednaku medijanu, tj.* 23650*,*
- *bar* 50% *ispitanika iz uzorka ima plaću veću ili jednaku* 23650*.*

Na kutijastom dijagramu mogu se označiti i takozvane **stršeće vrijednosti** ako postoje, a radi se o podacima koji su po svojoj vrijednosti značajno veći ili manji u odnosu na druge izmjerene vrijednosti promatrane varijable. Pojavljivanje je stršećih vrijednosti najčešće vezano uz jedan od sljedećih razloga:

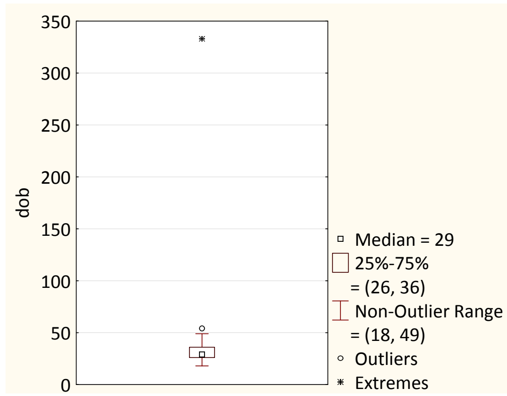
- podatak je ili netočno izmjeren ili krivo unesen u bazu podataka,
- podatak dolazi iz druge populacije (ne iz populacije koju promatramo u kontekstu problema kojega proučavamo) - npr. ako u varijablu čije su izmjerene vrijednosti godišnje plaće 1000 poreznih obveznika u Hrvatskoj upišemo godišnju plaću Microsoftovog managera iz SAD-a, taj će podatak biti stršeća vrijednost,
- podatak je točno izmjeren i unesen u bazu, ali predstavlja rijetku pojavu u populaciji - npr. ako se u varijabli čije su izmjerene vrijednosti koncentracije glukoze u krvi za 1000 osoba nađe točno izmjerena vrijednost 46.7, taj ćemo podatak smatrati stršećom vrijednošću jer se radi o vrlo visokoj koncentraciji glukoze koja se rijetko pojavljuje.

Primjer 4.23. (djelatnici.xls)
Varijabla dob *iz baze podataka* djelatnici.sta *za svakog ispitanika iz uzorka djelatnika promatranog poduzeća sadrži informaciju o dobi u godinama. Iz deskriptivne statistike te varijable (tablica 4.10) vidimo da je dob od* 333 *godine stršeći podatak i zaključujemo da se radi o podatku koji je pogrešno upisan u bazu podataka.*

veličina uzorka	aritmetička sredina	mod	frek. moda	st. dev.	varijanca
100	33.83	28	12	31.05	964.28
minimum	**donji kvartil**	**medijan**	**gornji kvartil**	**maksimum**	**raspon**
18	26	29	36	333	315

Tablica 4.10: Deskriptivna statistika varijable dob.

Osim iz tablice 4.10, stršeću vrijednost među izmjerenim vrijednostima varijable dob *mogli smo detektirati i pomoću kutijastog dijagrama na bazi medijana.*



Slika 4.9: Kutijasti dijagram na bazi medijana s prikazom stršećih vrijednosti za izmjerene vrijednosti varijable dob.

Kao što vidimo iz kutijastog dijagrama 4.9, i dob od 54 *godine prepoznata je kao stršeći podatak. Budući da je sasvim razumljivo da promatrano poduzeće može imati djelatnika starog* 54 *godine, taj podatak smatramo točno izmjerenim i točno upisanim u bazu podataka, no radi se o dobi koja se rijetko pojavljuje u populaciji djelatnika tog poduzeća. Tako iz kategorizirane tablice frekvencija i relativnih frekvencija (tablica 4.11) varijable* dob *vidimo da je takvih djelatnika samo* 0.03%.

dob	frekvencija	relativna frekvencija
$[18, 28\rangle$	37	0.37
$[28, 38\rangle$	44	0.44
$[38, 48\rangle$	15	0.15
$[48, 58\rangle$	3	0.03
$[58, 333]$	1	0.01

Tablica 4.11: Kategorizirana tablica frekvencija i relativnih frekvencijna varijable dob.

4.1.3 Metode opisivanja ordinalnih varijabli

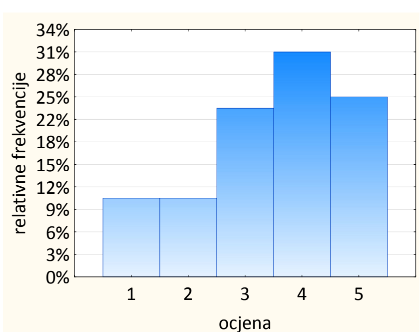
Ordinalne se varijable najčešće zadaju tako da mogu primiti samo nekoliko međusobno različitih vrijednosti i one su po svojoj prirodi kvalitativne, zbog čega su metode opisivanja kvalitativnih varijabli primjenjive u potpunosti i na ordinalne varijable. S obzirom da se vrijednosti ordinalinih varijabli izražavaju brojčano, često se u primjeni može vidjeti da se za ordinalne varijable izražavaju, komentiraju i koriste numeričke karakteristike podataka. Ponekad to ima smisla, ali moramo upozoriti da treba dobro razmisliti u svakom pojedinom slučaju je li i kako je moguće iskoristiti informaciju koju daje odabrana numerička karakteristika.

Primjer 4.24. *U hrvatskom su obrazovnom sustavu ocjene su od* 1 *do* 5*. Prilikom ocjenjivanja znanja na pisanim testovima nerijetko se ocjena* 1 *postiže za manje od* 40% *mogućih bodova, dok su ostale ocjene raspoređene na intervalu od* 40% *do* 100% *tako da svakoj ocjeni odgovara bodovni interval iste duljine. Pretpostavimo li da je prosjek svih postignutih ocjena 1.5, možemo li taj broj tumačiti u odnosu na znanje mjereno brojem bodova koje se ocjenjuje? Ipak, odredimo li da je medijan 1.5, možemo zaključiti da je barem* 50% *učenika postiglo manje od* 40% *bodova na testu.*

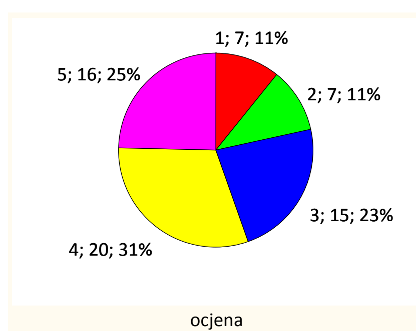
Primjer 4.25. (ocjena.xls)
Varijabla ocjena *baze podataka* ocjena.xls *ordinalna je varijabla koja sadrži ocjene na usmenom ispitu iz jednog kolegija za uzorak od* 65 *studenata iste generacije. Dakle, varijabla* ocjena *prima vrijednosti iz skupa* $1, 2, 3, 4, 5$*. Frekvencije i relativne frekvencije zabilježenih vrijednosti varijable* ocjena *dane su u tablici 4.12. Relativne frekvencije grafički su prikazane stupčastim dijagramom, a frekvencije i relativne frekvencije kružnim dijagramom na slici 4.10.*

ocjena	frekvencija	relativna frekvencija
1	7	$7/65 \approx 0.11$
2	7	$7/65 \approx 0.11$
3	15	$3/13 \approx 0.23$
4	20	$4/13 \approx 0.31$
5	16	$16/65 \approx 0.25$

Tablica 4.12: Tablica frekvencija i relativnih frekvencija svih zabilježenih vrijednosti varijable ocjena.



(a) stupčasti dijagram



(b) kružni dijagram

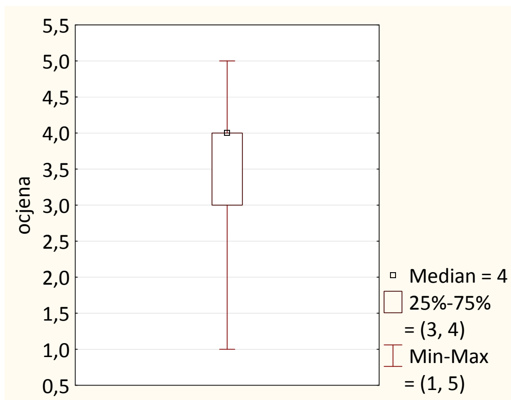
Slika 4.10: Grafički prikazi frekvencija i relativnih frekvencija svih zabilježenih vrijednosti varijable ocjena.

Numeričke karakteristike podataka varijable ocjena *prikazane su u tablici 4.13.*

veličina uzorka	aritmetička sredina	mod	frek. moda	st. dev.	varijanca
65	3.48	4	20	1.63	1.28
minimum	**donji kvartil**	**medijan**	**gornji kvartil**	**maksimum**	**raspon**
1	3	4	4	5	4

Tablica 4.13: Deskriptivna statistika varijable ocjena.

Odnosi minimuma, donjeg kvartila, medijana, gornjeg kvartila i maksimuma izmjerenih vrijednosti varijable ocjena *prikazani su kutijastim dijagramom 4.11.*



Slika 4.11: Kutijasti dijagram na bazi medijana za izmjerene vrijednosti varijable ocjena.

Aritmetičku sredinu, varijancu i standardnu devijaciju ne bismo komentirali jer ne znamo koliko se stvarno znanje krije iza svake ocjene, ali na temelju ostalih numeričkih karakteristika možemo izvesti sljedeće i slične zaključke:

- *na tom uzorku postignuta je i najniža i najviša moguća ocjena,*
- *bar 25% studenata iz uzorka ocijenjeno je ocjenom 3 ili manjom od 3, dok je bar 75% studenata ocijenjeno ocjenom 3 ili većom,*
- *bar 50% studenata iz uzorka ocijenjeno je ocjenom 4 ili 5.*

4.2 Statistički model

Pretpostavka je statističkog zaključivanja da se donose zaključci o varijablama od kojih baram jedna ima slučajan karakter, a na temelju prikupljenih realizacija za te varijable. Zbog toga je slučajna varijabla (slučajni vektor) matematički objekt koji je temelj svakoga statističkog zaključivanja.

Da bi primjenjivali metode statističkog zaključivanja, prva je pretpostavka postavljanje statističkog modela koji koristimo za modeliranje prikupljenih podataka. Zadati statistički model znači opisati poznate karakteristike slučajnog vektora za kojega smatramo da naši podaci čine jednu realizaciju. S obzirom da je slučajni vektor određen svojom funkcijom distribucije, statističkim je modelom opisano ono što je unaprijed poznato o funkciji distribucije slučajnog vektora kojim modeliramo podatke, odnosno **statistički model jest familija funkcija distribucije koja se uzima u obzir za zaključivanje u danom problemu**.

Najjednostavniji su modeli koji se koriste u statistici razni modeli **jednostavnoga slučajnog uzorka**.

Definicija 4.1. *Statistički model zovemo model* **jednostavnoga slučajnog uzorka** *iz funkcije distribucije F ako za slučajni vektor $(X_1, \ldots, X_n)$, čiju realizaciju čine podaci $(x_1, \ldots, x_n)$, vrijedi:*

- *slučajne varijable $X_1, \ldots, X_n$ nezavisne su,*
- *sve slučajne varijable $X_1, \ldots, X_n$ imaju istu funkciju distribucije F.*

Kod takvih modela promatranu veličinu smatramo slučajnom varijablom s funkcijom distribucije F, a vrijednosti varijable izmjerene na jedinkama iz uzorka nezavisne su realizacije te slučajne varijable.

Zbog jednostavnosti izražavanja u nastavku teksta ćemo koristiti termin **slučajni uzorak** za slučajni vektor $(X_1, \ldots, X_n)$ statističkog modela, a termin **uzorak** za njegovu realizaciju $(x_1, \ldots, x_n)$, tj. podatke.

U nastavku ćemo opisati nekoliko karakterističnih statističkih modela koji se vrlo često koriste u praksi, kao i primjere za koje su ti modeli prikladni.

4.2.1 Problem proporcije

Da bismo pojasnili probleme koji se mogu razmatrati u okviru naziva "problem proporcije" i definirali statistički model koji koristimo, poslužit ćemo se primjerom.

Primjer 4.26. *Imamo proizvodnu traku koja proizvodi proizvod A, ali s određenim (nepoznatim) postotkom škarta. Nepoznati postotak škarta označimo s θ.*

Skupo je kontrolirati svaki proizvod koji je proizveden na toj traci, pa za zaključivanje o postotku škarta uzimamo slučajni uzorak od n proizvoda.

Uzimanje jednog proizvoda u uzorak rezultira slučajnom varijablom

$$X = \begin{pmatrix} 0 & 1 \\ 1-\theta & \theta \end{pmatrix}, \ \ \theta \in [0,1],$$

koja daje jedinicu ako je proizvod škart, a u suprotnom nulu. Ako pod proporcijom smatramo omjer broja neispravnih proizvoda proizvedenih na toj traci i ukupnog broja proizvoda proizvedenih na toj traci u danom periodu, onda vjerojatnost $P\{X=1\}=\theta$ *odgovara proporciji škarta u populaciji.*

S obzirom da se na traci stalno proizvode novi proizvodi, ako ne dolazi do bitnih promjena u procesu proizvodnje tijekom vremena, možemo smatrati da svako uzimanje proizvoda u uzorak predstavlja realizaciju slučajne varijable jednako distribuirane kao X *te da su sve te slučajne varijable međusobno nezavisne.*

Podaci dobiveni u opisanom postupku čine uređenu n*-torku nula i jedinica koja je realizacija slučajnog vektora* $(X_1,\ldots,X_n)$ *n. j. d. slučajnih varijabli distribuiranih kao* X *(vidi Bernoullijevu shemu).*

Skup svih mogućih ishoda slučajnog vektora opisanog u prethodnom primjeru jest

$$\mathcal{R}(X_1,\ldots,X_n) = \{(x_1,\ldots,x_n) : x_i \in \{0,1\}\}.$$

Distribucija je zadana s

$$\begin{aligned} p(x_1,\ldots,x_n;\theta) &= P\{X_1=x_1,\ldots,X_n=x_n\} \\ &= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}, \ \ (x_1,\ldots,x_n) \in \mathcal{R}(X_1,\ldots,X_n). \end{aligned}$$

Tako nastao slučajni vektor zovemo **jednostavni slučajni uzorak iz Bernoullijeve distribucije** s parametrom $\theta \in [0,1]$.

Ako s F_θ označimo funkciju distribucije tog slučajnog uzorka za dani θ, onda statistički model u okviru kojega se bavimo zaključivanjem o proporciji čini familija funkcija distribucije:

$$\mathcal{P} = \{F_\theta : \theta \in [0,1]\}.$$

Zovemo ga **statistički model jednostavnoga slučajnog uzorka iz Bernoullijeve distribucije**.

Uočimo da je taj model $\mathcal{P}$ indeksiran parametrom θ, pa takav model zovemo **parametarski**.

Model slučajnog uzorka iz Bernoullijeve distribucije možemo koristiti u mnogo drugih praktičnih primjera.

Primjer 4.27. *U primjeru 4.26 uzorak od n proizvoda konstruiran je uzimanjem proizvoda s proizvodne trake. To možemo interpretirati kao odabir konačnog uzorka iz beskonačne populacije. U slučaju odabira konačnog uzorka iz konačne populacije, npr. uzorka od n vijaka iste vrste iz skladišta koje sadrži N takvih vijaka, moramo razlikovati dva slučaja:*

(1) *izvlačenje jednog po jednog vijka iz konačnog skupa vijaka, pri čemu se izvučeni element vraća u skup,*

(2) *izvlačenje jednog po jednog vijka iz konačnog skupa vijaka, pri čemu se izvučeni element ne vraća u skup.*

Slučaj (1)

Označimo s v broj neispravnih vijaka u tom skladištu. To znači da je proporcija neispravnih vijaka $\theta = v/N$. Označimo s X slučajnu varijablu koja se realizira nulom u slučaju da pri odabiru jednog vijka iz skladišta izvučemo ispravan, a jedinicom u slučaju da izvučemo neispravan vijak. Dakle, vidimo da je $P\{X = 1\} = v/N$, a $P\{X = 0\} = 1 - (v/N)$, tj. distribucija slučajne varijable X dana je tablicom

$$X = \begin{pmatrix} 0 & 1 \\ 1-(v/N) & v/N \end{pmatrix}.$$

Ako iz tog skladišta na slučajan način odaberemo n vijaka na način opisan u točki (1) ovog primjera i za svaki izvučeni vijak zabilježimo je li ispravan (0) ili neispravan (1), dobijemo jednu uređenu n-torku nula i jedinica koju shvaćamo kao jednu realizaciju slučajnog vektora $(X_1, \ldots, X_n)$ sa slikom $\mathcal{R}(X) = \{(x_1, \ldots, x_n) : x_i \in \{0, 1\}\}$. Zbog načina provođenja izvlačenja opisanog pod (1), komponente su slučajnog vektora $(X_1, \ldots, X_n)$ međusobno nezavisne i sve su jednako distribuirane kao slučajna varijabla X. Distribucija tog slučajnog vektora zadana je izrazom

$$p(x_1, \ldots, x_n; v) = \prod_{i=1}^{n} P\{X_i = x_i\} = \prod_{i=1}^{n} \left(\frac{v}{N}\right)^{x_i} \left(1 - \frac{v}{N}\right)^{n-x_i}. \tag{4.1}$$

Prema tome, $(X_1, \ldots, X_n)$ jest jednostavni slučajni uzorak iz Bernoullijeve distribucije s parametrom $\theta = v/N$, a familija pripadnih funkcija distribucije $\{F_\theta : \theta \in [0, 1]\}$ određenih distribucijom (4.1) pripadni je parametarski statistički model. Iz tog primjera vidimo da se parametarski statistički modeli za jednostavni slučajni uzorak iz populacije koja nije konačna i za jednostavni slučajni uzorak iz konačne populacije konstruiran na način opisan pod (1) podudaraju.

Slučaj (2)

Opišimo sada konstrukciju slučajnog uzorka iz konačne populacije na način opisan u točki (2) ovog primjera, tj. bez vraćanja u uzorak prethodno izvučenih elemenata. Slučajne varijable $X_1, \ldots, X_n$, kojima modeliramo ishode izvlačenja vijaka pri čemu nas zanima je li izvučeni vijak ispravan ili neispravan, nisu nezavisne. Slučajna varijabla X_1 jednako je distribuirana kao slučajna varijabla X opisana u prethodnom dijelu primjera, no slučajne varijable $X_2, \ldots X_n$ nemaju takvu distribuciju. Naime,

$$X_2|_{X_1=a_1} = \begin{pmatrix} 0 & 1 \\ 1 - \frac{v-a_1}{N-1} & \frac{v-a_1}{N-1} \end{pmatrix}, \quad a_1 \in \{0, 1\},$$

$$X_3|_{X_1=a_1, X_2=a_2} = \begin{pmatrix} 0 & 1 \\ 1 - \frac{v-a_1-a_2}{N-2} & \frac{v-a_1-a_2}{N-2} \end{pmatrix}, \quad a_1, a_2 \in \{0, 1\},$$

odnosno općenito za $i \in \{3, \ldots, N\}$

$$X_i|_{X_1=a_1,\ldots,X_{i-1}=a_{i-1}} = \begin{pmatrix} 0 & 1 \\ 1 - \frac{v-\sum_{j=1}^{i-1} a_j}{N-(i-1)} & \frac{v-\sum_{j=1}^{i-1} a_j}{N-(i-1)} \end{pmatrix}, \quad a_1, \ldots, a_{i-1} \in \{0, 1\}.$$

Distribucija slučajnog vektora $(X_1, \ldots, X_n)$ dana je na sljedeći način:

$$p(x_1, \ldots, x_n; v) = P\{X_1 = x_1, \ldots X_n = x_n\} =$$
$$= P\{X_1 = x_1, \ldots X_{n-1} = x_{n-1} | X_n = x_n\} P\{X_n = x_n\} = \ldots =$$
$$= P\{X_1 = x_1\} P\{X_2 = x_2 | X_1 = x_1\} \prod_{i=3}^{n} P\{X_i = x_i | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}\}. \tag{4.2}$$

Očito distribucija slučajne varijable X_1 i uvjetne distribucije slučajnih varijabli $X_2, \ldots, X_n$ ovise o nepoznatoj proporciji neispravnih vijaka u velikom skladištu, tj. o parametru $\theta = v/N \in [0,1]$. Prema tome, familija pripadnih funkcija distribucije $\{F_\theta : \theta \in [0,1]\}$ određenih distribucijom (4.2) pripadni je parametarski statistički model.
Uočimo da distribuciju (4.2) možemo aproksimirati distribucijom (4.1) ukoliko je

$$\frac{v - \sum_{j=1}^{i-1} a_j}{N - (i-1)} \approx \frac{v}{N}.$$

Riječima rečeno, za veliki N i mali v, parametarski statistički model temeljen na odabiru n elemenata iz populacije veličine N bez vraćanja prethodno izvučenih elemenata u tu populaciju možemo aproksimirati parametarskim statističkim modelom temeljenim na odabiru n elemenata iz iste populacije s vraćanjem prethodno izvučenih elemenata u tu populaciju.

4.2.2 Problem očekivanja i varijance normalne distribucije

Za mnogo kvantitativnih slučajnih pojava može se pretpostaviti da imaju normalnu distribuciju. Činjenica je to koju se može potkrijepiti centralnim graničnim teoremima. Naime, u centralnom graničnom teoremu (vidi poglavlje 3.5.2) navedeno je da se suma n. j. d. slučajnih varijabli koje imaju konačnu varijancu asimptotski ponaša po zakonu normalne distribucije.

Te jake pretpostavke o nezavisnosti i jednakoj distribuiranosti slučajnih varijabli koje se sumiraju mogu se u velikoj mjeri relaksirati, a da zaključak o asimptotski normalnom ponašanju sume ostane istinit (vidi npr. [26]).

Kao posljedicu takvih rezultata imamo činjenicu da vrlo često pojave, čiji je slučajan karakter rezultat sume puno slučajnih utjecaja koje ne možemo posebno analizirati, imaju distribuciju toliko blisku normalnoj distribuciji da u statističkoj analizi ne možemo primijetiti razlike između stvarne distribucije i aproksimativne normalne distribucije.

Za takve pojave provodimo statističko zaključivanje kao da su normalno distribuirane.

Primjer 4.28. *Beskonačna traka pakira šećer u pakovanja na kojima je deklarirana neto masa od 1 kg. Međutim, dovoljno preciznom vagom moguće je utvrditi da neto masa šećera pokazuje odstupanja od 1 kg. Po karakteru slučajne greške koja nastaje pri pakiranju šećera na tekućoj*

vrpci, može se smatrati da masa šećera u pakiranjima delariranim kao 1 *kg zapravo ima normalnu distribuciju.*
Uzimanje jednog pakiranja i precizno vaganje rezultira podatkom koji je realizacija normalne slučajne varijable s očekivanjem μ *i varijancom* σ^2*. Uzimamo li s trake* n *pakiranja i važemo im sadržaj, imamo podatke* $(x_1, x_2, \dots, x_n)$ *koji su jedna realizacija n. j. d. slučajnog vektora* $(X_1, X_2, \dots, X_n)$*.*

Slučajni vektor $(X_1, X_2, \dots, X_n)$ zovemo **jednostavni slučajni uzorak iz normalne distribucije s parametrima μ i σ^2**, ako su $X_1, X_2, \dots, X_n$ nezavisne i jednako distribuirane slučajne varijable s distribucijom koja je $\mathcal{N}(\mu, \sigma^2)$.

Ako je $F_{(\mu,\sigma^2)}$ funkcija distribucije normalne slučajne varijable za dani (μ, σ^2), onda je funkcija distribucije jednostavnoga slučajnog uzorka iz te distribucije (zbog nezavisnosti):

$$\mathbf{F}_{(\mu,\sigma^2)}(x_1, \dots, x_n) = \prod_{i=1}^{n} F_{(\mu,\sigma^2)}(x_i).$$

Dakle, statistički model u okviru kojega se bavimo zaključivanjem o očekivanju i varijanci normalne distribucije jest familija funkcija distribucije

$$\mathcal{P} = \{\mathbf{F}_{(\mu,\sigma^2)} : (\mu, \sigma^2) \in \mathbb{R} \times \langle 0, \infty\rangle\}, \tag{4.3}$$

a zovemo ga **statistički model jednostavnoga slučajnog uzorka iz normalne distribucije**.

Uočimo da je taj model $\mathcal{P}$ također indeksiran parametrom, ovoga puta dvodimenzionalnim (μ, σ^2), pa je i takav model parametarski.

Model slučajnog uzorka iz normalne distribucije možemo koristiti u mnogo praktičnih primjera.

Primjer 4.29. *Prosječnu godišnju količinu kiše koja padne na određenom području ima smisla promatrati kao slučajnu veličinu, tj. modelirati slučajnom varijablom. Naime, neka su* X_i, $i \in \{1, \dots, 365\}$*, slučajne varijable kojima modeliramo količinu kiše koja na promatranom području padne tijekom i-tog dana u neprijestupnoj godini. Tada prosječnu količinu kiše koja na tom području padne tijekom jedne neprijestupne godine modeliramo slučajnom varijablom*

$$Y = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad n = 365,$$

za koju zbog centralnog graničnog teorema možemo smatrati da ima normalnu distribuciju s parametrima μ *i* σ^2*. Dakle, ako raspolažemo podacima o prosječnoj godišnjoj količini kiše na nekom području za* 100 *neprijestupnih godina, zapravo raspolažemo jednom realizacijom slučajnog vektora* $(Y_1, \dots, Y_{100})$*. Ako su istinite pretpostavke o nezavisnosti i jednakoj distribuiranosti slučajnih varijabli* $Y_1, \dots, Y_{100}$*, ima smisla u zaključivanju o prosječnoj godišnjoj količini kiše na tom području koristiti statistički model jednostavnoga slučajnog uzorka iz normalne distribucije s očekivanjem* μ *i varijancom* σ^2*.*

4.2.3 Jednostavna linearna regresija

Da bismo lakše objasnili statistički model, poslužit ćemo se jednim primjerom.

Primjer 4.30. *U jednom istraživanju želi se odrediti ovisi li prosječna dnevna potrošnja goriva zaposlene osobe u Hrvatskoj, koja je vlasnik automobila i ima vozačku dozvolu, o udaljenosti mjesta stanovanja osobe od radnog mjesta. Pod pojmom "prosječna dnevna potrošnja goriva" ovdje podrazumijevamo ukupnu potrošnju goriva u godini dana podijeljenu s* 365.

Udaljenost mjesta stanovanja od radnog mjesta za svaku pojedinu osobu može se smatrati točno određenom, tj. determinističkom veličinom. Međutim, prosječnu dnevnu potrošnju goriva svakako treba tretirati kao slučajnu veličinu. S obzirom na to da se radi o prosječnoj potrošnji, zahvaljujući centralnom graničnom teoremu, za potrebe statističkog zaključivanja moći ćemo se osloniti na činjenicu da je ta veličina normalno distribuirana.

Statistički model za zaključivanje o vezi između determinističke veličine x (udaljenost mjesta stanovanja od radnog mjesta) i slučajne veličine Y_x (prosječna dnevna potrošnja goriva), izgradit ćemo u ovom primjeru postavljanjem linearne veze između tih veličina:

$$Y_x = ax + b + \varepsilon_x, \quad a, b \in \mathbb{R},$$

gdje ε_x predstavlja slučajnu varijablu greške modela za dani x.

U navedenom primjeru, kao i u mnogim drugim primjerima sličnog tipa, prikupljanjem podataka u uzorak imamo n uređenih parova brojeva

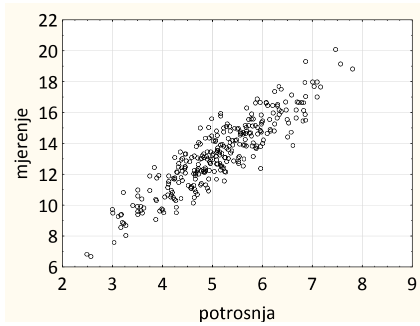
$$(x_1, y_1), \ldots, (x_n, y_n)$$

koji predstavljaju realizacije uređenih parova koje ćemo označiti

$$(x_1, Y_1), \ldots, (x_n, Y_n),$$

a kod kojih je druga komponenta slučajna varijabla. Uočimo da slučajne varijable iz tih parova, tj. $Y_1, \ldots, Y_n$ ne moraju biti jednako distribuirane. Naprotiv, pretpostavka je da njihova distribucija ovisi o iznosu prve koordinate iz para.

Primjer 4.31. (automobili.xls)
Varijabla potrosnja *baze podataka* automobili.sta *sadrži podatke o potrošnji goriva novog modela automobila pri brzini od* 110 *km/h za* 300 *nezavisnih mjerenja, dok varijabla* mjerenje *sadrži vrijednosti nekog parametra izmjerenog na tehničkom pregledu tog automobila nakon takve vožnje, a za kojeg se pretpostavlja da bi kod tehnički ispravnog automobila trebao biti linearno povezan s prosječnom potrošnjom automobila pri velikim brzinama. U kontekstu jednostavne linearne regresije izmjerene vrijednosti varijable* potrosnja *označavamo s $x_1, \ldots x_{300}$, a izmjerene vrijednosti varijable* mjerenje *označavamo s $(y_1, \ldots, y_{300})$ i smatramo ih realizacijama nezavisnih slučajnih varijabli $Y_1, \ldots, Y_{300}$ kojima modeliramo rezultate provedenih mjerenja. Parove $(x_1, y_1), \ldots, (x_{300}, y_{300})$ izmjerenih vrijednosti varijabli* potrosnja *i* mjerenje *grafički prikazujemo dijagramom raspršenosti (eng. scatterplot) 4.12.*



Slika 4.12: Dijagram raspršenosti za izmjerene vrijednosti varijabli potrosnja i mjerenje.

U postupku ćemo modeliranja pretpostaviti sljedeće:

- $EY_i = ax_i + b$ za svaki $i = 1, \ldots, n$, tj. ako je $\varepsilon_i = Y_i - ax_i - b$, onda je $E\varepsilon_i = 0$.

- $Y_1, \ldots, Y_n$ jesu međusobno nezavisne i imaju istu varijancu, označimo je σ^2. Ako je ε vektor $\varepsilon = [\varepsilon_1, \ldots, \varepsilon_n]^\tau$, onda je matrica kovarijanci tog vektora oblika $\sigma^2 I$ (I je oznaka za jediničnu matricu n-tog reda).

- $Y_1, \ldots, Y_n$ su normalno distribuirane slučajne varijable. Dakle, $Y_i \sim \mathcal{N}(ax_i + b, \sigma^2)$.

Statistički model za zaključivanje o parametrima a, b i σ^2 definirat ćemo korištenjem familije dopuštenih funkcija distribucije za slučajni vektor $(Y_1, \ldots, Y_n)$. Njegova funkcija distribucije određena je funkcijom gustoće

$$f(y_1, \ldots, y_n; a, b, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \prod_{i=1}^n e^{-\frac{(y_i - ax_i - b)^2}{2\sigma^2}} =$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^n (y_i - ax_i - b)^2}$$

u kojoj su tri nepoznata parametra a, b i σ^2, dok su $x_1, \ldots, x_n$ izmjerene vrijednosti prve komponente uređenih parova podataka.

Takav statistički model zovemo **klasičan jednostavan linearan regresijski model**.

4.3 Procjena parametra

U parametarskim statističkim modelima osnovu statističkog zaključivanja čini zaključivanje o parametrima. Jedan je od osnovnih problema njihova procjena na temelju realizacije uzorka iz danog statističkog modela. Za procjene parametra koristimo procjenitelje.

Definicija 4.2. *Neka je $\mathcal{P} = \{F_\theta : \theta \in \mathbf{\Theta} \subseteq \mathbb{R}^k\}$ parametarski statistički model, θ k-dimenzionalan parametar, a $\mathbf{\Theta} \subseteq \mathbb{R}^k$ prostor parametra, tj. skup svih dozvoljenih vrijednosti nepoznatog parametra θ. Neka je $(X_1, \ldots, X_n)$ slučajni vektor s distribucijom iz $\mathcal{P}$ i $t : \mathbb{R}^n \to \mathbf{\Theta}$. Slučajni vektor $\mathbf{T} = t(X_1, \ldots, X_n)$ jest* **procjenitelj** *za θ.*[1]

Primjer 4.32. *U modelu je jednostavnoga slučajnog uzorka iz Bernoullijeve distribucije s parametrom $p \in [0, 1]$ relativna frekvencija jedinice jedan procjenitelj za p.*

Primjer 4.33. *U modelu jednostavnoga slučajnog uzorka iz $\mathcal{N}(\mu, \sigma^2)$ označimo*

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\overline{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2,$$

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

Tada je $(\overline{X}_n, \overline{S}_n^2)$ jedan procjenitelj za (μ, σ^2), a $(\overline{X}_n, S_n^2)$ drugi procjenitelj za taj parametar.

Problem je kako izabrati funkciju t procjenitelja u pojedinom statističkom modelu. Takvim pitanjima bavi se matematička statistika. Ovdje ćemo samo navesti nekoliko svojstava koja su vrlo poželjna da ih procjenitelj parametra ima i intuitivno ilustrirati zašto su ta svojstva korisna.

Definicija 4.3. *Neka je $\mathcal{P} = \{F_\theta : \theta \in \mathbf{\Theta} \subseteq \mathbb{R}^k\}$ parametarski statistički model, a θ k-dimenzionalan parametar. Procjenitelj T* **nepristran** *je ako za svaki $\theta \in \mathbf{\Theta}$ vrijedi*

$$E_\theta T = E_\theta\, t(X_1, \ldots, X_n) = \theta.$$

Ukoliko je nepoznati parametar jednodimenzionalan, a predloženi procjenitelj nepristran, poželjno je da takav procjenitelj ima što manju varijancu. Naime, varijanca je jedna mjera raspršenosti oko očekivanja pa je za očekivati da će se procjena s većom vjerojatnošću realizirati u malim okolinama oko stvarne vrijednosti parametra

[1] Uočite da distribucija procjenitelja ovisi o vrijednosti nepoznatog parametra θ. Da bi tu činjenicu naglasili, vjerojatnost skupa $\{\mathbf{T} \in A\}$ označavamo s $P_\theta\{\mathbf{T} \in A\}$, a očekivanje od $\mathbf{T}$ s $E_\theta \mathbf{T}$.

(tj. očekivanja procjenitelja) ako je varijanca manja. Zbog toga, uspoređujući dva nepristrana procjenitelja, kažemo da je **efikasniji** onaj koji ima manju varijancu. Osim toga, poželjno je da varijanca nepristranog procjenitelja konvergira u nulu s porastom veličine uzorka pa za jednodimenzionalan parametar definiramo svojstvo konzistentnosti sljedećom definicijom.

Definicija 4.4. *Neka je $\mathcal{P} = \{F_\theta : \theta \in \Theta \subseteq \mathbb{R}\}$ parametarski statistički model, a θ jednodimenzionalan parametar. Procjenitelj T* **konzistentan** *je ako za pripadni niz procjenitelja $T_n = t(X_1, \dots, X_n)$, $n \in \mathbb{N}$, za svaki $\varepsilon > 0$ i $\theta \in \Theta$ vrijedi*

$$\lim_{n\to\infty} P_\theta\{|T_n - \theta| \geq \varepsilon\} = 0.$$

Primjer 4.34. *Neko poduzeće sve svoje finalne proizvode raspoređuje po kvaliteti u tri skupine: I, II i III kategoriju. Iskustveno je poznato da su vjerojatnosti da finalni proizvod bude I kategorije i da finalni proizvod bude II kategorije jednake. Ovu vjerojatnost označimo s θ. Dakle, slučajna varijabla koja opisuje kategoriju kvalitete proizvoda može se zadati na sljedeći način:*

$$X = \begin{pmatrix} 1 & 2 & 3 \\ \theta & \theta & 1-2\theta \end{pmatrix}.$$

Parametar θ može primiti bilo koju vrijednost iz intervala $[0, 1/2]$ da bi ta slučajna varijabla bila dobro definirana. Problem koji se postavlja pred statističara procjena je parametra θ na temelju n-dimenzionalnog slučajnog uzorka.

U skladu s prethodnim primjerom za očekivati je da se relativna frekvencija neke kategorije može iskoristiti u svrhu procjene s obzirom da θ ima značenje vjerojatnosti da se realizira proizvod prve kategorije, ali također i vjerojatnost da se realizira proizvod druge kategorije. Tako se može razmišljati o procjeniteljima:

$$T_1 = \frac{f_1}{n}, \quad T_2 = \frac{f_2}{n},$$

gdje je f_1 frekvencija jedinice, a f_2 frekvencija dvojke u uzorku.

Uočimo da su oba procjenitelja nepristrana. Naime, $f_1 \sim \mathcal{B}(n, \theta)$ i $f_2 \sim \mathcal{B}(n, \theta)$ pa je

$$E_\theta\, T_1 = E_\theta\, T_2 = \theta.$$

Možemo također izračunati i varijance:

$$\mathrm{Var}_\theta\, T_1 = \mathrm{Var}_\theta\, T_2 = \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n}.$$

Korištenjem se slabog zakona velikih brojeva lako vidi da su oba procjenitelja konzistentna. Na temelju tih analiza ne možemo preferirati niti jednog od navedenih dvaju procjenitelja. Međutim, razmotrimo kao varijantu i procjenitelja koji je vezan uz relativnu frekvenciju treće kategorije proizvoda (f_3):

$$T_3 = \frac{1}{2}\left(1 - \frac{f_3}{n}\right).$$

S obzirom da je $f_3 \sim \mathcal{B}(n, 1-2\theta)$, $E_\theta\, T_3 = \theta$, to je i nepristran procjenitelj za θ. Varijanca je

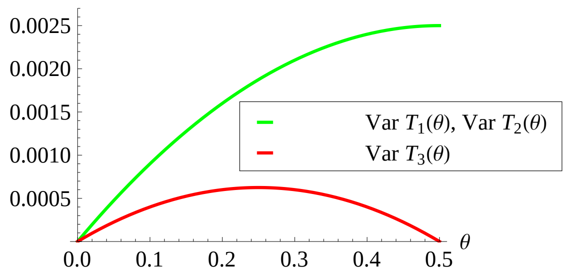
$$\mathrm{Var}_\theta\, T_3 = E_\theta(T_3 - \theta)^2 = \frac{1}{4n^2}\mathrm{Var}_\theta\, f_3 = \frac{1}{4n^2} n(1-2\theta)2\theta = \frac{\theta}{2n}(1-2\theta).$$

Dakle, to je također konzistenatan procjenitelj. Usporedimo li varijance od T_1 i T_2 s varijancom od T_3 za svaki $\theta \in \Theta$ i $n \in \mathbb{N}$, vidimo da je

$$\operatorname{Var}_\theta T_3(\theta, n) \leq \operatorname{Var}_\theta T_1(\theta, n) = \operatorname{Var}_\theta T_2(\theta, n),$$

pa je T_3 efikasniji procjenitelj za θ (slika 4.13). Napomenimo da taj rezultat nije neočekivan s obzirom da procjenitelj T_3 u sebi sadrži informacije koje nose oba prethodna procjenitelja jer vrijedi:

$$T_3 = \frac{1}{2}(T_1 + T_2).$$



Slika 4.13: Varijance procjenitelja T_1, T_2 i T_3 za $n = 100$.

4.3.1 Procjena proporcije

Neka je $(X_1, \ldots, X_n)$ jednostavni slučajni uzorak iz Bernoullijeve distribucije

$$X = \begin{pmatrix} 0 & 1 \\ 1-\theta & \theta \end{pmatrix}, \quad \theta \in [0, 1].$$

Tada je relativna frekvencija jedinice

$$T = \frac{f_1}{n}$$

jedan nepristran i konzistentan procjenitelj za parametar θ koji ima značenje vjerojatnosti realizacije jedinice.

Zaista, s obzirom da slučajna varijabla koja daje broj uspjeha u n nezavisnih Bernoullijevih pokusa ima binomnu distribuciju s parametrima n i θ, to je $f_1 \sim \mathcal{B}(n, \theta)$. Vrijedi:

$$E_\theta \frac{f_1}{n} = \frac{n\theta}{n} = \theta.$$

Konzistentnost je posljedica slabog zakona velikih brojeva.

Primjer 4.35. *Promotrimo igraću kockicu sa svojstvom da se pri jednom bacanju broj* $i \in \{1, \dots, 6\}$ *okrene s vjerojatnošću* p_i, *pri čemu je*

$$p_i > 0, \quad \forall i \in \{1, \dots, 6\}, \qquad \text{i} \qquad \sum_{i=1}^{6} p_i = 1.$$

Želimo procijeniti vjerojatnost da se pri bacanju te kockice okrene šestica, stoga označimo s 1 *događaj "okrenula se šestica", a s* 0 *događaj "nije se okrenula šestica". U tom kontekstu rezultat bacanja kockice modeliramo Bernoullijevom slučajnom varijablom*

$$X = \begin{pmatrix} 0 & 1 \\ 1 - p_6 & p_6 \end{pmatrix},$$

gdje je p_6 *nepoznati parametar kojega ćemo procijeniti.*

Pretpostavimo da smo kockicu bacili nezavisno sto puta zaredom. Realizaciju tih 100 *bacanja modeliramo slučajnim vektorom* $(X_1, \dots X_{100})$ *čije su komponente nezavisne i jednako distribuirane kao slučajna varijabla* X, *tj.* $(X_1, \dots X_{100})$ *je jednostavni slučajni uzorak iz Bernoullijeve distribucije s parametrom* p_6. *Nadalje, pretpostavimo da se u tih* 100 *bacanja šestica realizirala* 17 *puta. To znači da je realizacija slučajnog uzorka* $(X_1, \dots X_{100})$ *uređena* 100*-torka koja se sastoji od* 17 *jedinica i* 83 *nule. Slijedi da je relativna rekvencija pojavljivanja šestice, broj* 17/100, *procjena vjerojatnosti* p_6. *Analogno bismo postupali pri procjeni bilo koje od preostalih vjerojatnosti* p_i, $i \in \{1, \dots, 5\}$.

Primjer 4.36. (djelatnici.xls)
Varijabla dob *baze podataka* djelatnici.xls *sadrži informacije o godinama starosti za svakoga od* 100 *djelatnika iz reprezentativnog uzorka zaposlenika tvornice A. Označimo s* θ *vjerojatnost da je slučajno odabrani djelatnik te tvornice stariji od* 30 *godina te procijenimo* θ. *Znamo da svaki djelatnik te tvornice ima ili najviše* 30 *godina (oznaka* 0*) ili je stariji od* 30 *godina (oznaka* 1*) pa rezultat ispitivanja starosti u tom kontekstu možemo modelirati Bernoullijevom slučajnom varijablom*

$$X = \begin{pmatrix} 0 & 1 \\ 1 - \theta & \theta \end{pmatrix}, \quad \theta \in [0, 1].$$

Prema tome, podatke iz varijable dob *shvaćamo kao jednu realizaciju jednostavnog slučajnog uzorka* $(X_1, \dots, X_{100})$ *iz te Bernoullijeve distribucije i na temelju te realizacije parametar* θ *procjenjujemo relativnom frekvencijom jedinica, tj. relativnom frekvencijom djelatnika iz uzorka koji su stariji od* 30 *godina. Kako takvih djelatnika u uzorku ima* 46, *parametar* θ *procjenjujemo brojem* $46/100 = 0.46$.

4.3.2 Procjena očekivanja

Neka je $(X_1, X_2, \dots, X_n)$ jednostavni slučajni uzorak iz distribucije F_μ, gdje je $\mu \in \mathbb{R}$ nepoznato očekivanje. Dakle, $(X_1, X_2, \dots, X_n)$ je vektor n. j. d. slučajnih varijabli iz distribucije F_μ. Pretpostavimo da varijanca distribucije F_μ također postoji i označimo je sa σ^2.

Ako razmatramo problem procjene očekivanja, možemo iskoristiti **aritmetičku sredinu uzorka**

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

koja je nepristran i konzistentan procijenitelj za očekivanje. Naime,

$$E_\mu \overline{X}_n = \mu, \quad \mathrm{Var}_\mu \, \overline{X}_n = \frac{\sigma^2}{n},$$

a konzistentnost slijedi iz slabog zakona velikih brojeva.

Primjer 4.37. *Ocjenu znanja studenata nekog fakulteta na usmenom ispitu iz jednog kolegija modeliramo diskretnom slučajnom varijablom X sa slikom $\mathcal{R}(X) = \{1, 2, 3, 4, 5\}$ i funkcijom distribucije F_μ, gdje je μ nepoznato očekivanje koje želimo procijeniti. Pretpostavimo da od svih studenata fakulteta koji su ocijenjeni na usmenom ispitu tog kolegija odabiremo reprezentativan uzorak od 65 studenata. Frekvencije ocjena na tom uzorku dane su tablicom frekvencija 4.14.*

ocjena	frekvencija
1	7
2	7
3	15
4	20
5	16

Tablica 4.14: Tablica frekvencija ocjena studenata na usmenom ispitu iz jednog kolegija.

Ocjene prikupljene na uzorku, predstavljene kao uređena 65-torka, čine jednu realizaciju jednostavnog slučajnog uzorka $(X_1, \dots, X_{65})$ iz distribucije F_μ. Procjenitelj za očekivanje od X jest slučajna varijabla $\overline{X}_{65}$. Njezinu realizaciju na uzorku $(x_1, \dots, x_{65})$ označimo s $\overline{x}_{65}$. Prema tome, procjena očekivanja slučajne varijable X na temelju tablice frekvencija 4.14 jest

$$\overline{x}_{65} = \frac{1 \cdot 7 + 2 \cdot 7 + 3 \cdot 15 + 4 \cdot 20 + 5 \cdot 16}{65} \approx 3.48.$$

Primjer 4.38. (kolokvij.xls)
U bazi podataka kolokvij.sta *nalaze se rezultati dvaju kolokvija iz nekog kolegija (varijable* kolokvij_1 *i* kolokvij_2*) te ukupan broj bodova ostvaren na kolokvijima (varijabla* bodovi_ukupno*) za reprezentativan uzorak od 70 studenata koji su pristupili tim kolokvijima. Na svakom od kolokvija bilo je moguće ostvariti najviše 100 bodova. Dakle, na obama kolokvijima bilo je moguće ostvariti najviše 200 bodova.*
Usmjerimo se na ukupan broj bodova ostvaren na obama kolokvijima. Njega modeliramo kontinuiranom slučajnom varijablom X s funkcijom distribucije F_μ, gdje je μ nepoznato očekivanje koje želimo procijeniti.
Postignuti ukupni bodovi za promatrani uzorak od 70 studenata (podaci iz varijable bodovi_ukupno*) jedna su realizacija jednostavnog slučajnog uzorka $(X_1, \dots, X_{70})$ iz distribucije F_μ. Prema tome, procjenitelj $\overline{X}_{70}$ za očekivanje od X realizira se aritmetičkom sredinom podataka iz varijable* bodovi_ukupno*, tj. procjena je očekivanja od X $\overline{x}_{70} \approx 96.9$.*

Primjer 4.39. (djelatnici.xls)
Varijabla placa_prije *baze podataka* djelatnici.xls *sadrži iznos godišnje plaće za reprezentativan uzorak od 100 djelatnika tvornice A. Godišnju plaću djelatnika te tvornice modeliramo kontinuiranom slučajnom varijablom X s funkcijom distribucije F_μ, gdje je μ nepoznato očekivanje koje želimo procijeniti.*

Iznosi godišnjih plaća za promatrani uzorak od 100 *djelatnika (podaci iz varijable* placa_prije*) jedna su realizacija jednostavnog slučajnog uzorka* $(X_1, \ldots, X_{100})$ *iz distribucije* F_μ*. Prema tome, procjenitelj* $\overline{X}_{100}$ *za očekivanje od* X *realizira se aritmetičkom sredinom podataka iz varijable* placa_prije*, tj. procjena je očekivanja od* X $\overline{x}_{100} \approx 24522$.

4.3.3 Procjena varijance

Neka je $(X_1, X_2, \ldots, X_n)$ jednostavni slučajni uzorak iz distribucije F_θ, gdje je $\theta = \sigma^2 > 0$ nepoznata varijanca. Dakle, $(X_1, X_2, \ldots, X_n)$ je vektor n. j. d. slučajnih varijabli iz distribucije F_θ. Označimo s μ očekivanje te distribucije.

Ako razmatramo problem procjene varijance, možemo iskoristiti **varijancu uzorka**

$$\overline{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

Međutim, taj procjenitelj nije nepristran. Naime,

$$\begin{aligned}\overline{S}_n^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}[(X_i-\mu) - (\overline{X}_n - \mu)]^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 + \frac{1}{n}\sum_{i=1}^{n}(\overline{X}_n-\mu)^2 - \frac{2}{n}(\overline{X}_n-\mu)\sum_{i=1}^{n}(X_i-\mu).\end{aligned}$$

Budući da je

$$\sum_{i=1}^{n}(\overline{X}_n-\mu)^2 = n(\overline{X}_n-\mu)^2, \quad \sum_{i=1}^{n}(X_i-\mu) = n(\overline{X}_n-\mu),$$

slijedi da je

$$\overline{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - (\overline{X}_n-\mu)^2.$$

Dakle,

$$E_\theta\overline{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}E_\theta[(X_i-\mu)^2] - E_\theta[(\overline{X}_n-\mu)^2] = \frac{1}{n}n\sigma^2 - \frac{\sigma^2}{n},$$

tj.

$$E_\theta\overline{S}_n^2 = \frac{n-1}{n}\theta.$$

Ako napravimo malu korekciju varijance uzorka i definiramo novog procjenitelja, **korigiranu varijanca uzorka**

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2,$$

njegovo očekivanje bit će upravo jednako varijanci obilježja koje proučavamo, pa ćemo tako dobiti zadovoljeno svojstvo nepristranosti.
Zaista,

$$E_\theta S_n^2 = E_\theta \left(\frac{n}{n-1} \overline{S}_n^2 \right) = \theta. \tag{4.4}$$

Može se pokazati (vidi npr. [21]) da je S_n^2 ujedno i konzistentan procjenitelj.

Primjer 4.40. (kolokvij.xls)
Neka je X diskretna slučajna vrijabla iz primjera 4.38 kojom smo modelirali ukupan broj bodova studenta postignutih na dvama kolokvijima iz nekog kolegija. Znamo da je procjena očekivanja od X dana s $\overline{x}_{70} \approx 96.9$. Za procjenu varijance koristimo standardne procjenitelje: varijancu uzorka $\overline{S}_{70}^2$ i korigiranu varijancu uzorka S_{70}^2. Za realizaciju uzorka iz funkcije distribucije od X, danu u varijabli bodovi_ukupno, *$\overline{S}_{70}^2$ realizira se procjenom*

$$\overline{s}_{70}^2 = 43.56,$$

a S_{70}^2 procjenom

$$s_{70}^2 = 44.19.$$

Primjer 4.41. (djelatnici.xls)
Standardni tablični kalkulatori i programski paketi za statističku analizu imaju ugrađenu funkciju za procjenu varijance definiranu na temelju korigirane varijance uzorka, tj. na temelju procjenitelja S_n^2. Procijenimo na taj način varijancu neprekidne slučajne varijable X kojom modeliramo godišnju plaću zaposlenika u jednoj tvornici (primjer 4.39). Za zabilježene vrijednosti slučajne varijable X na uzorku od 100 zaposlenika te tvornice (varijabla placa_prije*) korigirana varijanca uzorka S_{100}^2 realizira se procjenom $s_{100}^2 \approx 26069208.1 = 5105.8^2$.*

4.3.4 Procjena funkcije distribucije

Neka je $(X_1, X_2, \ldots, X_n)$ jednostavni slučajni uzorak iz nepoznate funkcije distribucije F slučajne varijable X. Statistički model takvog slučajnog uzorka ne možemo smatrati parametarskim s obzirom da familija dozvoljenih funkcija distribucije nije indeksirana nepoznatim parametrom, nego je potpuno neodređena. Takav statistički model zvat ćemo **neparametarski**.

U takvom modelu opisat ćemo jedan način procjene funkcije distribucije F. Preciznije, za svaki pojedini $x \in \mathbb{R}$ procjenu vrijednosti funkcije distribucije, $F(x)$.

S obzirom da je $F(x)$ po definiciji $P\{X_i \leq x\}$, za svaki $i \in \{1, \ldots, n\}$ i za dani x zapravo imamo problem procjene vjerojatnosti uspjeha (uspjeh znači da se realizirao događaj $\{X_i \leq x\}$) pa se možemo poslužiti rezultatima iz procjene proporcije.

Dakle, neka je x fiksan realan broj, a F funkcija distribucije slučajne varijable X. Definirajmo slučajnu varijablu

$$I_{\{X\leq x\}}=\begin{cases}1\ ,\ X\leq x\\ 0\ ,\ X>x.\end{cases}$$

Njezina je distribucija:

$$I_{\{X\leq x\}}=\begin{pmatrix}0 & 1\\ 1-F(x) & F(x)\end{pmatrix},\quad F(x)\in[0,1].$$

Uočimo da ovdje $F(x)$ igra ulogu parametra distribucije slučajne varijable $I_{\{X\leq x\}}$.

Iz problema procjene proporcije znamo da je dobar procjenitelj za parametar Bernoullijeve distribucije relativna frekvencija jedinice, a u slučaju varijable $I_{\{X\leq x\}}$ to je relativna frekvencija skupa $\{X\leq x\}$. Označimo

$$\hat{F}(x)=\frac{f_{\{X\leq x\}}}{n}=\frac{1}{n}\sum_{i=1}^{n}I_{\{X_i\leq x\}}.$$

Tada je

$$E(\hat{F}(x))=F(x)$$

za svaki pojedini $x\in\mathbb{R}$ i

$$\mathrm{Var}\,(\hat{F}(x))=\frac{F(x)(1-F(x))}{n},$$

pa možemo govoriti o nepristranosti i konzistentnosti procjenitelja $\hat{F}(x)$ za svaki pojedini $x\in\mathbb{R}$. Tako definiran procjenitelj funkcije distribucije na temelju realizacije jednostavnog slučajnog uzorka zove se **empirijska funkcija distribucije**.

Primjer 4.42. (djelatnici.xls)
Diskretna numerička varijabla rukovodstvo *baze podataka* djelatnici.xls *iz primjera 4.5 prima vrijednosti iz skupa* $\{0,1,\ldots,N\}$*, gdje je* N *najveći mogući broj godina radnog staža koje djelatnik može provesti na rukovodećoj poziciji. Frekvencije i relativne frekvencije zabilježenih vrijednosti varijable* rukovodstvo *za djelatnike iz promatranog uzorka dane su u tablici 4.6. Procijenimo funkciju distribucije diskretne slučajne varijable* X *kojom modeliramo varijablu* rukovodstvo*. U tu svrhu promotrimo kumulativne frekvencije zabilježenih vrijednosti te varijable (tablica 4.15).*

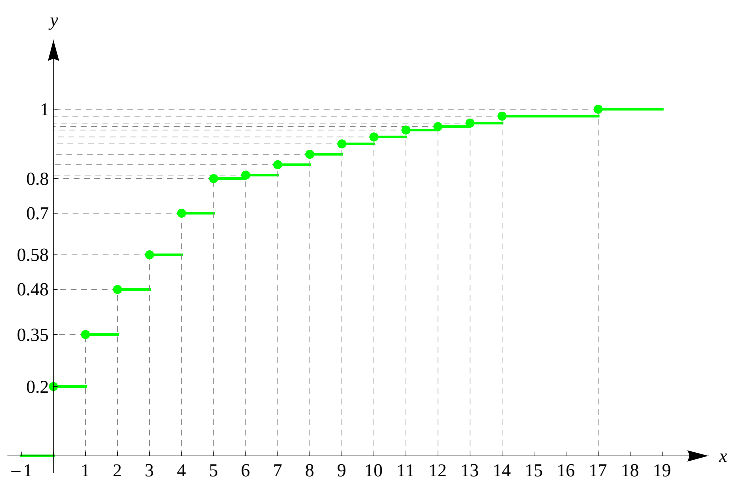
rukovodstvo	frek.	kumulativna frek.	rukovodstvo	frek.	kumulativna frek.
0	20	0.20	8	3	0.87
1	15	0.35	9	3	0.9
2	13	0.48	10	2	0.92
3	10	0.58	11	2	0.94
4	12	0.7	12	1	0.95
5	10	0.8	13	1	0.96
6	1	0.81	14	2	0.98
7	3	0.84	17	2	1

Tablica 4.15: Tablica frekvencija i kumulativnih frekvencija svih zabilježenih vrijednosti varijable rukovodstvo.

Iz tablice 4.15, preciznije, iz stupca s kumulativnim frekvencijama, slijedi da je procjena funkcije distribucije slučajne varijable X, tj. njezina empirijska funkcija distribucije, definirana izrazom

$$\hat{F}(x) = \frac{f_{\{X \leq x\}}}{100} = \begin{cases} 0 & , \quad x < 0 \\ 0.2 & , \; 0 \leq x < 1 \\ 0.35 & , \; 1 \leq x < 2 \\ 0.48 & , \; 2 \leq x < 3 \\ 0.58 & , \; 3 \leq x < 4 \\ 0.7 & , \; 4 \leq x < 5 \\ \vdots & \vdots \qquad \vdots \\ 1 & , \quad x \geq 17. \end{cases}$$

Graf funkcije $\hat{F}(x)$ prikazan je na slici 4.14.

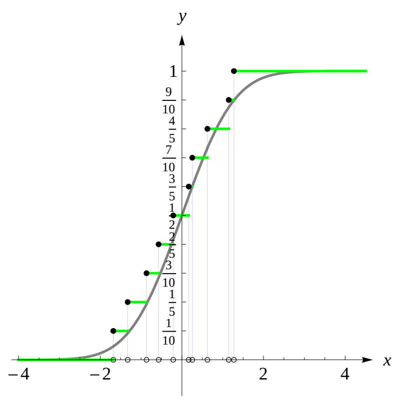


Slika 4.14: Empirijska funkcija distribucije slučajne varijable X iz primjera 4.44.

Primjer 4.43. *Funkcija distribucije standardne normalne slučajne varijable definirana je pravilom*

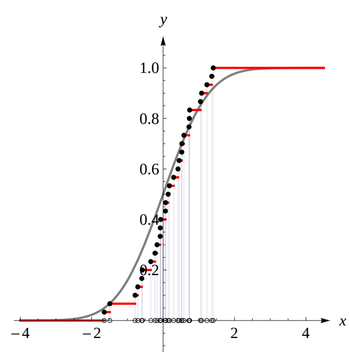
$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt, \quad x \in \mathbb{R}.$$

Graf te funkcije prikazan je na slici 2.20 u drugom poglavlju. U ovom ćemo primjeru na temelju simuliranih vrijednosti iz standardne normalne distribucije (tj. simulirane realizacije n-dimenzionalnog jednostavnog slučajnog uzorka iz standardne normalne distribucije) ilustrirati ponašanje procjene $\hat{F}$ za F u ovisnosti o veličini uzorka.
Na temelju jedne simulacije 10 *vrijednosti iz te distribucije dobili smo graf funkcije $\hat{F}(x)$ prikazan na slici 4.15 (stepenasti graf).*

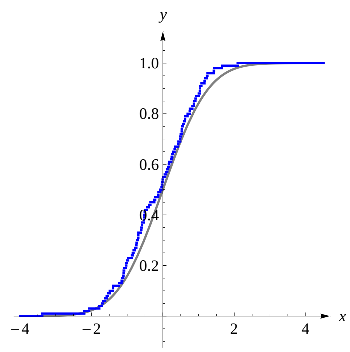


Slika 4.15: Grafovi funkcije distribucije i empirijske funkcije distribucije ($n = 10$) standardne normalne slučajne varijable.

Grafovi empirijskih funkcija distribucije standardne normalne slučajne varijable za 30 *i* 100 *simuliranih vrijednosti prikazani su na slikama 4.16 i 4.17.*



Slika 4.16: Grafovi funkcije distribucije i empirijske funkcije distribucije ($n = 30$) standardne normalne slučajne varijable.



Slika 4.17: Grafovi funkcije distribucije i empirijske funkcije distribucije ($n = 100$) standardne normalne slučajne varijable.

Korištenjem nekog programskog paketa simulirajte 1000 *podataka iz standardne normalne distribucije i nacrtajte graf empirijske funkcije distribucije. Usporedite ju sa stvarnom funkcijom distribucije.*

Primjer 4.44. (djelatnici.xls)
Procijenimo funkciju distribucije neprekidne slučajne varijable X iz primjera 4.39 kojom modeliramo godišnju plaću djelatnika jedne tvornice. Varijabla placa_prije *sadrži mnogo različitih realizacija slučajne varijable X. Uočimo da su frekvencije svih realizacija male. Empirijska funkcija distribucije temeljena na svih* 77 *različitih realizacija slučajne varijable X može se lako očitati korištenjem relativnih kumulativnih frekvencija. Tako, npr.* $\hat{F}(18400) = 0.05$ *(tablica 4.16).*

plaća	frekvencija	kumulativna frekvencija
16000	1	1
18100	1	2
18300	1	3
18400	2	5
⋮	⋮	⋮
24500	1	57
24600	4	61
⋮	⋮	⋮
42400	1	100

Tablica 4.16: Tablica kumulativnih frekvencija plaća djelatnika iz primjera 4.39.

4.3.5 Procjena parametara u jednostavnoj linearnoj regresiji

Ideja za procjenu parametara u linernom regresijskom modelu (poglavlje 4.2.3) temelji se na jednostavnom zahtjevu da treba minimizirati sumu kvadrata odstupanja eksperimentalnih od teorijskih vrijednosti.

Neka je $(x_1, y_1), \ldots, (x_n, y_n)$ dana realizacija slučajnog uzorka iz regresijskog problema. Onda su $y_1, \ldots, y_n$ tzv. eksperimentalne vrijednosti, tj. realizacije slučajnog vektora $(Y_1, \ldots, Y_n)$ kojega želimo opisati modelom

$$Y_i = ax_i + b + \varepsilon_i, \quad i = 1, \ldots, n, \tag{4.5}$$

kao što je navedeno u poglavlju 4.2.3. S obzirom da ε_i predstavljaju slučajne varijable koje opisuju grešku modela, teorijske su vrijednosti $y(x_i) = ax_i + b$. Jedan procjenitelj za nepoznati vektorski parametar (a, b) može se dobiti minimizacijom izraza

$$S(a, b) = \sum_{i=1}^{n} (y(x_i) - y_i)^2$$

po $(a, b) \in \mathbb{R} \times \mathbb{R}$.

Uvedimo oznake:

$$\mathbf{x} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}, \quad \beta = [a, b]^\tau, \quad \mathbf{y} = [y_1, \ldots, y_n]^\tau.$$

Rješenje je toga minimizacijskog problema

$$\hat{\beta} = (\mathbf{x}^\tau \mathbf{x})^{-1} \mathbf{x}^\tau \mathbf{y}. \tag{4.6}$$

Uvrstimo li slučajni vektor $\mathbf{Y} = [Y_1, \ldots Y_n]^\tau$ u izraz za $\hat{\beta}$ umjesto $\mathbf{y}$, imamo procjenitelja za nepoznati parametar β.

Može se pokazati da je tako definiran procjenitelj za β nepristran. Za dokaze i ostala svojstva procjenitelja vidi npr. [21] ili [29].

Primjer 4.45. (automobili.xls)
Baza podataka automobili.xls *opisana je u primjeru 4.12. U istom primjeru opisano je i značenje izmjerenih vrijednosti varijabli* potrosnja *i* mjerenje *u jednostavnom linearnom regresijskom modelu: $x_1, \ldots x_{300}$ prosječne su potrošnje promatranog tipa automobila pri vožnji po autocesti brzinom 110 km/h (varijabla* potrosnja*), a $y_1, \ldots, y_{300}$ realizacije su slučajnog vektora $(Y_1, \ldots, Y_{300})$*

kojim modeliramo rezultate mjerenja nekog parametra na tehničkom pregledu tog automobila nakon te vožnje, a za kojega se pretpostavlja da bi kod tehnički ispravnog automobila trebao biti linearno povezan s prosječnom potrošnjom automobila pri velikim brzinama. Uvrštavanjem podataka iz baze automobili.xls *u izraz 4.6 i primjenom nekog statističkog softvera možemo dobiti procjene parametara a i b regresijskog modela 4.5 u ovom primjeru:*

$$\widehat{a} = 2.138, \quad \widehat{b} = 2.349.$$

Slijedi da modelom određen rezultat mjerenja promatranog parametra na tehničkom pregledu tog automobila nakon vožnje u opisanim uvjetima računamo pomoću pravila

$$y(x) = \widehat{a}x + \widehat{b} = 2.138x + 2.349.$$

4.4 Procjena parametra pouzdanim intervalom

U ovom poglavlju pretpostavit ćemo da je statistički model za podatke parametarski s **jednodimenzionalnim** parametrom θ kojega želimo procijeniti. Parametar procjenjujemo korištenjem procjenitelja T. S obzirom da je parametar jednodimenzionalan, procjenitelj je slučajna varijabla. Procjena je na temelju podataka samo jedna realizacija procjenitelja. Koristeći distribuciju procjenitelja u danom statističkom modelu moguće je dobiti i druge informacije o stvarnoj vrijednosti procijenjenog parametra, ne samo jedan broj koji predstavlja procjenu.

U ovom poglavlju opisat ćemo koncept procjene parametra **pouzdanim intervalom** i interpretaciju takve procjene te ilustrirati procjenu pouzdanim intervalom na nekoliko primjera.

Pouzdani interval za procjenu parametra po svojoj definiciji nije interval s realnim brojevima kao granicama nego interval kojemu su granice slučajne varijable pa ga možemo zvati **slučajni interval**.

Definicija 4.5. *Neka je $\gamma \in \langle 0, 1\rangle$, a slučajni vektor $(X_1, \ldots, X_n)$ s distribucijom iz parametarske familije $\mathcal{P} = \{F_\theta : \theta \in \Theta \subseteq \mathbb{R}\}$ neka određuje statistički model. Ako postoje dva procjenitelja $D = d(X_1, \ldots, X_n)$ i $G = g(X_1, \ldots, X_n)$ za paramear θ sa svojstvima:*

- $d(x_1, \ldots, x_n) \leq g(x_1, \ldots, x_n)$, *za sve* $(x_1, \ldots, x_n) \in \mathcal{R}(X_1, \ldots, X_n)$,
- $P_\theta\{D \leq \theta \leq G\} \geq \gamma$, *za sve* $\theta \in \Theta$,

onda kažemo da je slučajni interval $[D, G]$ pouzdani interval za θ pouzdanosti γ.

Kao što se može prepoznati iz definicije, pouzdani interval određuje se na temelju zahtjeva da se stvarna vrijednost parametra nalazi u slučajnom intervalu s vjerojatnošću barem γ.

Ako na temelju podataka izračunamo realizacije procjenitelja D i G koji su granice slučajnog intervala, dobit ćemo običan interval s realnim brojevima kao granicama. Dakle, realizacija slučajnog intervala običan je interval realnih brojeva. Iz definicije pouzdanog intervala pouzdanosti γ jasno je da će, pod pretpostavkom adekvatnosti statističkog modela, u približno $100\gamma\,\%$ slučajeva izračunati interval realnih brojeva sadržavati stvarnu vrijednost parametra.

Dakle, interval pouzdanosti γ jest slučajni interval, tj. granice su mu slučajne varijable. Jedna realizacija intervala pouzdanosti γ, određena na osnovu podataka (realizacije slučajnog vektora statističkog modela), običan je interval realnih brojeva. Uobičajeno je u praksi i tu realizaciju pouzdanog intervala također zvati pouzdani interval. Međutim, važno je znati razliku između pouzdanog intervala kao slučajnog intervala i njegove realizacije - običnog intervala realnih brojeva.

U nastavku ilustriramo postupak određivanja pouzdanih intervala na nekoliko primjera. Zainteresirani čitatelj više o tom problemu može pronaći u literaturi koja opširnije obrađuje teme iz matematičke statistike kao npr. [2], [9], [16], [21], [22], [24].

4.4.1 Procjena očekivanja pouzdanim intervalom za velike uzorke

Neka je $(X_1, \ldots, X_n)$ jednostavni slučajni uzorak iz zadane distibucije, označimo je F_μ, za koju je varijanca poznata i iznosi σ^2, a očekivanje je μ nepoznato i želimo ga procijeniti iz jedne realizacije uzorka.

Aritmetička sredina uzorka $\overline{X}_n$ jest jedan nepristran i konzistentan procjenitelj za očekivanje. Osim toga, na temelju centralnog graničnog teorema poznato je da standardizirana aritmetička sredina, tj.

$$\sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma},$$

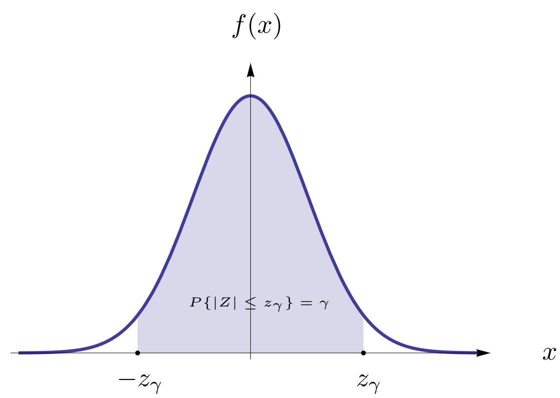
ima asimptotski standardnu normalnu distribuciju. Označimo sa $Z \sim \mathcal{N}(0,1)$. Korištenjem tih rezultata konstruirat ćemo pouzdani interval za očekivanje na temelju jednostavnoga slučajnog uzorka iz distribucije F_μ.

Neka je $\gamma \in \langle 0, 1\rangle$ izabrana pouzdanost i z_γ broj za koji vrijedi

$$P\{|Z| \leq z_\gamma\} = \gamma.$$

Uočimo da vrijednost γ predstavlja površinu ispod grafa funkcije gustoće standardne normalne distribucije nad intervalom $[-z_\gamma, z_\gamma]$ (slika 4.18), tj.

$$P\{|Z| \leq z_\gamma\} = \frac{1}{\sqrt{2\pi}} \int\limits_{-z_\gamma}^{z_\gamma} e^{-x^2/2}\, dx = \gamma.$$



Slika 4.18: Vjerojatnost $P\{|Z| \leq z_\gamma\}$.

S obzirom da distribuciju od $\frac{\overline{X}_n - \mu}{\sigma}\sqrt{n}$ možemo dobro aproksimirati standardnom normalnom distribucijom za velike n, vrijedi:

$$P_\mu\left\{-z_\gamma \leq \frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} \leq z_\gamma\right\} = P_\mu\left\{\overline{X}_n - z_\gamma\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n + z_\gamma\frac{\sigma}{\sqrt{n}}\right\}.$$

Dakle, vrijedi:

$$P_\mu\left\{\mu \in \left[\overline{X}_n - z_\gamma\frac{\sigma}{\sqrt{n}}, \overline{X}_n + z_\gamma\frac{\sigma}{\sqrt{n}}\right]\right\} \approx \gamma,$$

pa slučajni interval $\left[\overline{X}_n - z_\gamma\frac{\sigma}{\sqrt{n}}, \overline{X}_n + z_\gamma\frac{\sigma}{\sqrt{n}}\right]$ možemo smatrati pouzdanim intervalom za μ pouzdanosti γ ukoliko imamo veliku veličinu uzorka n.

Postupak procjene očekivanja pouzdanim intervalom na temelju realizacije $(x_1, \ldots, x_n)$ jednostavnog slučajnog uzorka iz distribucije s varijancom σ^2, tj. standardnom devijacijom σ, možemo opisati sljedećim koracima:

- odrediti broj z_γ za koji vrijedi da je $P\{|Z| \leq z_\gamma\} = \gamma$, gdje je $Z \sim \mathcal{N}(0,1)$,
- izračunati interval po formuli

$$\left[\overline{x}_n - z_\gamma \frac{\sigma}{\sqrt{n}},\ \overline{x}_n + z_\gamma \frac{\sigma}{\sqrt{n}}\right]. \tag{4.7}$$

Interval izračunat tim postupkom u približno $100\gamma\%$ slučajeva sadržat će stvarnu vrijednost očekivanja μ.

U praksi najčešće ne znamo stvarnu vrijednost standardne devijacije σ. S obzirom da je taj rezultat izveden za velike uzorke, može se također dokazati da će korištenje procjenitelja za standardnu devijaciju (korijena korigirane varijance uzorka) umjesto standardne devijacije u izvodu formule za pouzdani interval, dati analognu asimptotsku distribuciju. Iz tog se razloga u praksi koristi procjena standardne devijacije umjesto stvarne vrijednosti standardne devijacije ako je ona nepoznata.

Primjer 4.46. (kolokvij.xls)
Baza podataka kolokvij.xls *ukratko je opisana u primjeru 4.38. Procijenimo intervalom pouzdanosti* $\gamma = 0.95$ *očekivanje diskretne slučajne varijable* X *kojom modeliramo ukupan broj bodova ostvaren na tim dvama kolokvijima. Da bismo izračunali realizaciju traženog intervala pouzdanosti, trebaju nam sljedeće vrijednosti*

$$n = 70, \quad \overline{x}_{70} \approx 96.9, \quad s_{70} \approx 44.19, \quad z_{0.95} \approx 1.96,$$

gdje su $\overline{x}_{70}$ *i* s_{70} *redom procjene očekivanja i standardne devijacije slučajne varijable* X*. Korištenjem izraza 4.7 slijedi da je traženi interval*

$$[86.36, 107.44].$$

Procjena istog očekivanja intervalom pouzdanosti $\gamma = 0.97$ *zahtijeva poznavanje vrijednosti* $z_{0.97} \approx 2.17$*. Tražena je realizacija intervala pouzdanosti* $\gamma = 0.97$ *interval*

$$[85.19, 108.61].$$

Uočimo da se za veću vrijednost γ *interval pouzdanosti realizira širim intervalom realnih brojeva nego za manju vrijednost* γ*.*

Primjer 4.47. (djelatnici.xls)
Procijenimo intervalom pouzdanosti $\gamma = 0.97$ *očekivanje neprekidne slučajne varijable iz primjera 4.39 kojom modeliramo godišnju plaću djelatnika jedne tvornice (varijabla* placa_prije*). Da bismo izračunali realizaciju traženog intervala pouzdanosti, trebaju nam sljedeće vrijednosti:*

$$n = 100, \quad \overline{x}_{100} \approx 24522, \quad s_{100} \approx 5105.8, \quad z_{0.97} \approx 2.17,$$

gdje su $\overline{x}_{100}$ *i* s_{100} *redom procjene očekivanja i standardne devijacije slučajne varijable* X *određene u primjerima 4.39 i 4.41. Korištenjem izraza 4.7 slijedi da je traženi interval*

$$[23414, 25630].$$

4.4.2 Procjena proporcije pouzdanim intervalom za velike uzorke

U statističkom modelu koji smo nazvali "problem proporcije" zaključujemo o vjerojatnosti da se realizira broj 1 na temelju n-dimenzionalnoga jednostavnog slučajnog uzorka iz Bernoullijeve distribucije.

Nepristran i konzistentan procjenitelj koji u tu svrhu koristimo jest relativna frekvencija jedinice, tj.

$$\frac{f_1}{n}.$$

Slično prethodnom poglavlju, centralni granični teorem ovdje omogućuje određivanje asimptotske distribucije standardizirane relativne frekvencije jedinice, pa ćemo na temelju tog rezultata konstruirati pouzdani interval za proporciju.

Neka je $(X_1, \ldots, X_n)$ jednostavni slučajni uzorak iz Bernoullijeve distribucije s parametrom $\theta \in \langle 0, 1\rangle$. Tada je

$$\frac{f_1}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Osim toga, $f_1 \sim \mathcal{B}(n, \theta)$, pa je

$$E_\theta \frac{f_1}{n} = \theta, \quad \mathrm{Var}_\theta \frac{f_1}{n} = \frac{\theta(1-\theta)}{n}.$$

Prema centralnom graničnom teoremu sada slijedi da slučajna varijabla

$$\sqrt{n}\frac{\frac{f_1}{n} - \theta}{\sqrt{\theta(1-\theta)}}$$

ima asimptotski standardnu normalnu distribuciju. Označimo sa $Z \sim \mathcal{N}(0, 1)$ i z_γ odaberimo kao u prethodnom poglavlju, tj. $P\{|Z| \leq z_\gamma\} = \gamma$. Sada vrijedi:

$$\gamma \approx P_\theta\left\{\left|\sqrt{n}\frac{\frac{f_1}{n} - \theta}{\sqrt{\theta(1-\theta)}}\right| \leq z_\gamma\right\}.$$

Nejednadžba

$$\left|\sqrt{n}\frac{\frac{f_1}{n} - \theta}{\sqrt{\theta(1-\theta)}}\right| \leq z_\gamma$$

bit će zadovoljena ako je $\theta \in [\theta_1, \theta_2]$, gdje su θ_1 i θ_2 rješenja jednadžbe

$$\theta^2(n + z_\gamma^2) - \theta(2f_1 + z_\gamma^2) + \frac{f_1^2}{n} = 0.$$

Pod pretpostavkom da je n velik, u rješenjima je te kvadratne jednadžbe n mnogo veći od z_γ pa se zanemarivanjem članova koji sadrže z_γ^2 dobije:

$$\theta_1 \approx \frac{f_1}{n} - z_\gamma\sqrt{\frac{1}{n}\frac{f_1}{n}(1-\frac{f_1}{n})}, \quad \theta_2 \approx \frac{f_1}{n} + z_\gamma\sqrt{\frac{1}{n}\frac{f_1}{n}(1-\frac{f_1}{n})}.$$

Za danu je realizaciju toga jednostavnog slučajnog uzorka uobičajeno relativnu frekvenciju jedinice označiti s $\hat{p}$, a relativnu frekvenciju nule s $\hat{q}$. Korištenjem takvih oznaka i dobivenih rezultata navodimo postupak za računanje pouzdanog intervala za proporciju na temelju realizacije jednostavnoga slučajnog uzorka iz Bernoullijeve distribucije:

- odrediti broj z_γ za koji vrijedi da je $P\{|Z| \leq z_\gamma\} = \gamma$, gdje je $Z \sim \mathcal{N}(0,1)$,
- izračunati interval po formuli:

$$\left[\hat{p} - z_\gamma\sqrt{\frac{\hat{p}\hat{q}}{n}}, \hat{p} + z_\gamma\sqrt{\frac{\hat{p}\hat{q}}{n}}\right]. \tag{4.8}$$

Primjer 4.48. *Neka je X diskretna slučajna varijabla kojom modeliramo ocjene znanja studenata jednog fakulteta na usmenom ispitu iz nekog kolegija (primjer 4.37). Procijenimo intervalom pouzdanosti $\gamma = 0.95$ proporciju studenata koji usmeni ispit polažu ocjenom izvrstan. Ovdje se realizacija ocjene izvrstan smatra "uspjehom", a realizacija bilo koje druge ocjene "neuspjehom", pa iz tablice 4.14 slijedi da je*

$$\hat{p} = \frac{16}{65}, \quad \hat{q} = 1 - \frac{16}{65} = \frac{49}{65}.$$

Kako je $n = 65$, a $z_{0.95} \approx 1.96$, primjenom izraza (4.8) slijedi da je realizacija traženog intervala pouzdanosti interval

$$[0.14, 0.35].$$

Primjer 4.49. (kolokvij.xls)
Neka je X diskretna slučajna varijabla kojom modeliramo ukupan broj bodova studenta ostvaren na dvama kolokvijima iz nekog kolegija (primjer 4.38). Procijenimo intervalom pouzdanosti $\gamma = 0.95$ proporciju studenata koji su na kolokvijima ukupno ostvarili barem 100 od mogućih 200 bodova. Ovdje se ukupan broj bodova veći ili jednak 100 smatra "uspjehom", a ukupan broj bodova manji od 100 "neuspjehom", pa iz podataka dostupnih u varijabli bodovi_ukupno *slijedi da je*

$$\hat{p} = \frac{39}{70} \approx 0.56, \quad \hat{q} = 1 - \frac{39}{70} = \frac{31}{70} \approx 0.44.$$

Kako je $n = 70$, a $z_{0.95} \approx 1.96$, primjenom izraza 4.8 slijedi da je realizacija traženog intervala pouzdanosti interval

$$[0.44, 0.67].$$

Primjer 4.50. (djelatnici.xls)
Neka je X neprekidna slučajna varijabla iz primjera 4.39 kojom modeliramo godišnju plaću djelatnika jedne tvornice. Procijenimo intervalom pouzdanosti $\gamma = 0.98$ proporciju djelatnika koji

godišnje zarađuju više od 30000. *Ovdje se realizacija godišnje plaće veće od* 30000 *smatra "uspjehom", a realizacija godišnje plaće manje ili jednake* 30000 *"neuspjehom". Iz podataka za varijablu* placa_prije *slijedi da je*

$$\hat{p} = 0.16, \quad \hat{q} = 1 - 0.16 = 0.84.$$

Budući da je $n = 100$, *a* $z_{0.98} \approx 2.33$, *primjenom izraza* 4.8 *slijedi da je traženi interval*

$$[0.075, 0.245].$$

4.5 Testiranje statističkih hipoteza

Statistička hipoteza slutnja je koja se odnosi na statistički model. Cilj je testiranja hipoteze odlučiti o istinitosti ili neistinitosti te slutnje.

Primjer 4.51. *O proporciji pušača izabranog područja može se zaključivati na temelju reprezentativnog uzorka koji predstavlja jednu realizaciju jednostavnog slučajnog uzorka iz Bernoullijeve distribucije s parametrom* $\theta \in \langle 0, 1\rangle$. *Nepoznati je parametar* θ *upravo proporcija pušača populacije. Jedna je slutnja/hipoteza koju možemo testirati npr. da je proporcija pušača u toj populaciji manja od 20%.*

Postupak testiranja hipoteza uvijek počinje prevođenjem slutnje koja nas zanima u statističku hipotezu. To znači formiranje statističkog modela u okviru kojega ćemo zaključivati i izricanje hipoteze u terminima koji se odnose na odabrani statistički model. S obzirom da je statistički model $\mathcal{P}$ familija funkcija distribucije koje su dozvoljene prilikom zaključivanja o zadanom problemu, očigledno je da postavljena hipoteza zapravo određuje jedan podskup od $\mathcal{P}$. U prethodnom je primjeru statistički model odabran, a hipotezu možemo iskazati kao podskup distribucija modela za koji vrijedi nejednakost $\theta \leq 0.2$.

Statističku hipotezu standardno označavamo s $\mathcal{H}$. Testirati hipotezu znači donijeti odluku o tome hoćemo li $\mathcal{H}$ odbaciti ili ne.

Odluku u postupku testiranja hipoteza donosimo na temelju odabranog kriterija i realizacije uzorka. Odabrati kriterij zapravo znači definirati pravilo za odbacivanje hipoteze. S obzirom da se odluka donosi na temelju realizacije slučajnog vektora $(X_1, \ldots, X_n)$ statističkog modela, pravilo mora biti iskazano u terminima tog slučajnog vektora . U tu svrhu koristimo pojam "statistika".

Definicija 4.6. *Neka je* $(X_1, \ldots, X_n)$ *slučajni vektor statističkog modela* $\mathcal{P}$ *na temelju kojega donosimo zaključke i neka je* $t : \mathcal{R}(X_1, \ldots, X_n) \to \mathbb{R}^k$ *zadana funkcija. Slučajni vektor/varijablu* $T = t(X_1, \ldots, X_n)$ *zovemo* **statistika** *za taj statistički model.*

Uočimo da je i procjenitelj jedna statistika, ali s dodatnim zahtjevom na skup vrijednosti funkcije t.

Slično kao kod definiranja pouzdanih intervala za procjenu nepoznatog parametra, kod odabira pravila za testiranje hipoteza koristimo svojstva statistike kao slučajnog vektora/varijable, a prilikom donošenja odluke koristimo odabranu statistiku i pravilo te realizaciju statistike na podacima.

Pravilo po kojemu odbacujemo postavljenu hipotezu podijelit će skup svih mogućih realizacija slučajnog vektora statističkog modela na dva disjunktna dijela. Uobčajeno ih označavamo s C_r i C_r^c, pri čemu je C_r dio koji odgovara odbacivanju postavljene hipoteze i kojega zovemo **kritično područje**.
S obzirom da postavljena hipoteza $\mathcal{H}$ izdvaja podskup distribucija iz statističkog modela $\mathcal{P}$, uobičajeno je uvesti i drugu hipotezu u postupku testiranja. Ta hipoteza obuhvatit će sve one distribucije koje nisu sadržane u $\mathcal{H}$. Zbog toga često govorimo o testiranju dviju hipoteza u statističkom testu. Jednu od njih zovemo **nul-hipoteza** i označavamo s $\mathcal{H}_0$, a drugu **alternativna hipoteza** i označavamo s $\mathcal{H}_1$. Alternativna hipoteza jest ona koju prihvaćamo u slučaju odbacivanja nul-hipoteze. Za statistički model $\mathcal{P}$ vrijedi:

$$\mathcal{P} = \mathcal{H}_0 \cup \mathcal{H}_1.$$

4.5.1 Pogreške statističkog testa

Odluka koja je donesena statističkim testom može biti pogrešna ili ispravna. Pritom se mogu dogoditi dva tipa pogrešne odluke:

pogreška I. tipa: **odbaciti $\mathcal{H}_0$ ako je ona istinita,**
pogreška II. tipa: **ne odbaciti $\mathcal{H}_0$ ako je $\mathcal{H}_1$ istinita.**

Vjerojatnosti pogreške prvog tipa i pogreške drugog tipa ovise o stvarnoj distribuciji slučajne varijable o kojoj testiramo hipotezu, tj. te su vjerojatnosti zapravo funkcije koje distribuciji iz $\mathcal{P}$ pridružuju broj, ali one nisu definirane na istoj domeni. Vjerojatnost pogreške prvog tipa može se računati za svaku distribuciju iz $\mathcal{H}_0$, dok se vjerojatnost pogreške drugog tipa računa za svaku distribuciju iz $\mathcal{H}_1$. Ilustrirajmo taj postupak na primjeru.

Primjer 4.52. *Pretpostavimo da želimo zaključivati o proporciji pušača izabranog područja na temelju jednostavnoga slučajnog uzorka iz Bernoullijeve distribucije s parametrom $\theta \in \langle 0, 1\rangle$ (primjer 4.51). Neka je $\mathcal{P} = \{F_\theta : F_\theta \text{ Bernoullijeva}, \theta \in \langle 0,1\rangle\}$ pripadni statistički model. Taj model parametarski je, pa i hipoteze možemo izražavati u terminima parametra. Imamo slutnju da je* *proporcija pušača manja ili jednaka* 20%. *Statistička hipoteza koja opisuje ovu slutnju jest podskup od $\mathcal{P}$ za koji vrijedi $\theta \leq 0.2$. Dakle,*

$$\mathcal{H}_0: \ \theta \leq 0.2,$$
$$\mathcal{H}_1: \ \theta > 0.2.$$

Za taj parametarski model podjela $\mathcal{P}$ na $\mathcal{H}_0$ i $\mathcal{H}_1$ ekvivalentna je podjeli skupa dozvoljenih vrijednosti parametra $\langle 0,1\rangle$ na dva dijela: $\langle 0, 0.2]$ i $\langle 0.2, 1\rangle$. Vjerojatnost pogreške prvog tipa i vjerojatnost pogreške drugog tipa sada možemo promatrati kao funkcije od nepoznatog parametra. Ako je odabran postupak za odbacivanje nul-hipoteze, znači da je definirano kritično područje C_r, pa se te funkcije mogu definirati na sljedeći način:

- *vjerojatnost pogreške prvog tipa α:*

$$\alpha : \langle 0, 0.2] \to [0,1], \ \ \alpha(\theta) = P_\theta\{(X_1, \dots X_n) \in C_r\},$$

- *vjerojatnost pogreške drugog tipa β:*

$$\beta : \langle 0.2, 1\rangle \to [0,1], \ \ \beta(\theta) = P_\theta\{(X_1, \dots X_n) \in C_r^c\}.$$

U prehodnom primjeru uočimo da se funkcija β može prikazati kao

$$\beta(\theta) = 1 - P_\theta\{(X_1, \dots X_n) \in C_r\}.$$

Objedinjeni prikaz obiju funkcija vjerojatnosti pogreške postiže se definiranjem funkcije jakosti testa.

Funkcija jakosti testa, u oznaci π, definirana je za svaku distribuciju iz $\mathcal{P}$ (odnosno za svaku dozvoljenu vrijednost parametra ako je statistički model parametarski) kao

$$\pi(F) = P_F\{(X_1, \dots, X_n) \in C_r\}, \quad F \in \mathcal{P},$$

odnosno za parametarski statistički model

$$\pi(\theta) = P_\theta\{(X_1, \dots, X_n) \in C_r\}, \quad \theta \in \Theta.$$

Uočimo da je

$$\alpha(F) = \pi(F), \quad \forall F \in \mathcal{H}_0,$$
$$\beta(F) = 1 - \pi(F), \quad \forall F \in \mathcal{H}_1.$$

Ilustrirajmo primjerom funkciju jakosti testa i njezinu vezu s vjerojatnostima pogreške.

Primjer 4.53. *U problemu zaključivanja o proporciji pušača na nekom području (primjeri 4.51 i 4.52) želimo testirati hipotezu da je proporcija pušača manja ili jednaka* 20%. *Statistički model i hipoteze opisani su u primjeru 4.52:*

$$\mathcal{H}_0: \ \theta \leq 0.2,$$
$$\mathcal{H}_1: \ \theta > 0.2.$$

Odaberimo sljedeće pravilo odbacivanja nul-hipoteze:

C_r: odbacit ćemo nul-hipotezu ako je relativna frekvencija pušača u uzorku veća od 22%.

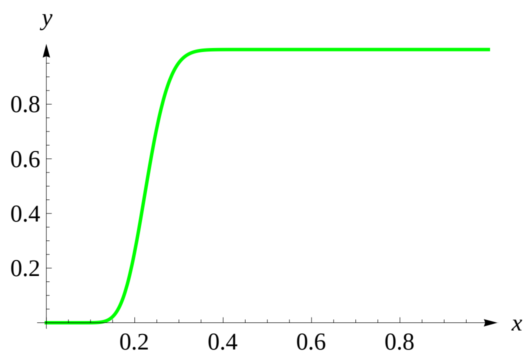
Uz standardnu oznaku f_1 za frekvenciju pušača, funkcija je jakosti za takav test dana izrazom

$$\pi(\theta) = P_\theta\{(X_1, \ldots, X_n) \in C_r\} = P_\theta\left\{\frac{f_1}{n} > 0.22\right\}.$$

Pod pretpostavkam statističkog modela znamo da je $f_1 \sim \mathcal{B}(n, \theta)$. Dakle, vrijedi:

$$P_\theta\left\{\frac{f_1}{n} > 0.22\right\} = P_\theta\{f_1 > 0.22n\} = \sum_{k>0.22n} \binom{n}{k} \theta^k (1-\theta)^{n-k}.$$

Graf te funkcije za n=100 prikazan je slikom 4.19.



Slika 4.19: Funkcija jakosti testa za kritično područje $\frac{f_1}{n} > 0.22$.

S obzirom da je $\pi(0.2) = 0.261067$, vidimo da je maksimalna vjerojatnost pogreške prvog tipa

$$\max_{\theta \in \mathcal{H}_0} \alpha(\theta) = 0.261067,$$

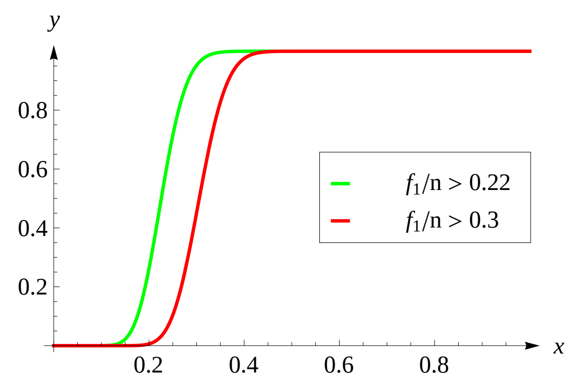
dok je maksimalna vjerojatnost pogreške drugog tipa za tako definirano kritično područje

$$\max_{\theta \in \mathcal{H}_1} \beta(\theta) = 1 - 0.261067 = 0.738933$$

Odabirom drugog kritičnog područja tako da granicu za odbacivanje nul-hipoteze pomaknemo u veće vrijednosti, npr.

C_r: odbacit ćemo nul-hipotezu ako je relativna frekvencija pušača u uzorku veća od 30%,

možemo smanjiti maksimalnu vjerojatnost pogreške prvog tipa, ali će se povećati maksimalna vjerojatnost pogreške drugog tipa. Grafovi funkcija jakosti testa za oba navedena kritična područja i $n = 100$ prikazani su slikom 4.20.



Slika 4.20: Funkcije jakosti testa za kritična područja $\frac{f_1}{n} > 0.22$ i $\frac{f_1}{n} > 0.3$.

Za novo je kritično područje

$$\max_{\theta \in \mathcal{H}_0} \alpha(\theta) = 0.00605934$$

$$\max_{\theta \in \mathcal{H}_1} \beta(\theta) = 1 - 0.00605934 = 0.993941.$$

Povećanje maksimalne vjerojatnosti jedne pogreške ako želimo smanjiti maksimalnu vjerojatnost druge pogreške, ilustrirano prethodnim primjerom, nije neuobičajno. Dapače, može se pokazati da je nemoguće odabrati kritično područje koje bi minimiziralo maksimalnu vjerojatnost obiju pogrešaka po svim mogućim kritičnim područjima. Zbog toga se odabir kritičnog područja vrši tako da se dopušta istraživaču izbor maksimalne vjerojatnosti pogreške prvog tipa koju želi prihvatiti. Te se vrijednosti uglavnom biraju između brojeva 0.01, 0.05 ili 0.1. Odabrana maksimalna vjerojatnost pogreške prvog tipa zove se **razina značajnosti testa** ili **nivo signifikantnosti testa** i standardno označava s α. Uz izabranu razinu značajnosti testa, test se dizajnira uz nastojanje da se maksimalna vjerojatnost pogreške drugog tipa učini što manjom. Dakle, maksimalna je vjerojatnost pogreške drugog tipa temelj za teorijsku analizu testa i odabir test-statistike, no najčešće se ne iskazuje u primjeni testa. Primjena se oslanja na činjenicu da je odabrani test kreiran kao optimalan za dani skup pretpostavki.

Uzimajući u obzir da ćemo biti u mogućnosti birati maksimalnu vjerojatnost pogreške prilikom odbacivanja nul-hipoteze, to je informacija koju u primjeni testa referiramo. Npr. reći ćemo da **odbacujemo nul-hipotezu na nivou značajnosti** α, tj. da odbacujemo nul-hipotezu uz vjerojatnost najviše α da smo pri tome pogriješili. Ako pravilo testa primijenjeno na podatke sugerira da ne odbacimo nul-hipotezu, prilikom primjene testa obično nemamo dostupnu informaciju koliko iznosi maksimalna vjerojatnost da smo pogriješili. Zato ćemo tada reći kako podaci ne podupiru tvrdnju da $\mathcal{H}_0$ treba odbaciti.

Takav neravnopravan odnos između nulte i alternativne hipoteze prilikom kreiranja statističkog testa upućuje na činjenicu da nije svejedno kako smo izbrali nultu i alternativnu hipotezu i pripadni test. Ukoliko je moguće, uputno je u primjeni birati statistički test kojemu alternativna hipoteza odgovara tvrdnji koju želimo dokazati.

Iako je teorija odabira optimalne test-statistike vrlo razvijena (vidi npr. [18], [2], [21]) ovdje se nećemo baviti metodama za odabir optimalnih statistika za testiranje hipoteza odabranog modela. U nastavku samo ilustriramo intuitivnu metodu odabira postupka za testiranje hipoteza na nekoliko primjera bez analize maksimalne vjerojatnosti pogreške drugog tipa. Zainteresirani čitatelj više o tom problemu može pronaći u literaturi koja opširnije obrađuje teme iz matematičke statistike kao npr. [2], [9], [16], [21], [22], [24].

4.5.2 Testiranje hipoteze o očekivanju za velike uzorke

U ovom poglavlju ilustrirat ćemo jedan od statističkih testova koji možemo koristiti prilikom testiranja hipoteza o iznosu očekivanja distribucije slučajne varijable na temelju koje je formiran statistički model jednostavnoga slučajnog uzorka. Problem ilustriramo sljedećim primjerom.

Primjer 4.54. *Pretpostavimo da želimo provjeriti je li očekivana vrijednost vremena čekanja u redu studentske menze u vrijeme ručka veća od pet minuta. U tu svrhu, od sto slučajno izabranih studenata koji odlaze na ručak u studentsku menzu prikupljamo podatke o vremenu čekanja. Tako dolazimo do podataka* $(x_1, \ldots, x_{100})$*. Na temelju tih podataka aritmetičkom sredinom procijenili smo očekivanje slučajne varijable iz koje potječu podaci - procjena je iznosila* 6.5 *minuta. Znajući iz prethodnih proučavanja te slučajne varijable da je njezina varijanca* 25*, želimo testirati slutnju da je očekivano vrijeme čekanja veće od 5 minuta.*

Test kojim možemo testirati takvu slutnju izvest ćemo na temelju statističkog modela jednostavnog slučajnog uzorka $(X_1, \ldots, X_n)$ iz distribucije F kojoj je varijanca poznata i iznosi σ^2, a očekivanje μ jest nepoznato.

Postavimo nultu i alternativnu hipotezu na sljedeći način:

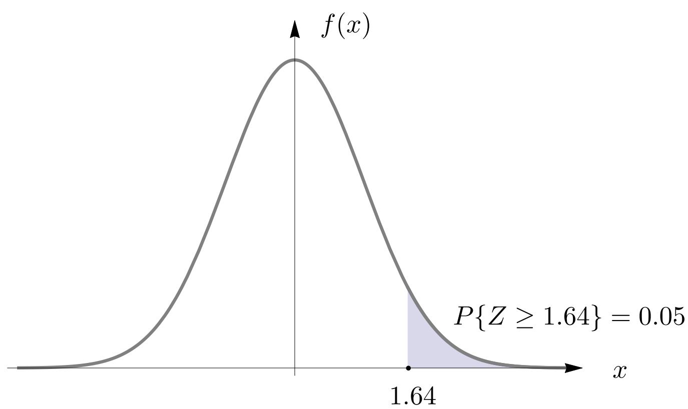
$$\mathcal{H}_0\colon\ \mu = \mu_0$$
$$\mathcal{H}_1\colon\ \mu > \mu_0.$$

Ako je $\mathcal{H}_0$ istinita hipoteza, a uzorak velik, onda je distribucija artmetičke sredine uzorka približno normalna s očekivanjem μ_0 i varijancom σ^2/n (centralni granični teorem!). Dakle, pod pretpostavkom je istinitosti nul-hipoteze distribucija od

$$Z' = \frac{\overline{X}_n - \mu_0}{\sigma}\sqrt{n}$$

približno standardna normalna i očekuje se realizacija od Z' blizu ili manje od nule (slika 4.21). Uočimo primjerice da se realizacije veće ili jednake 1.64 pojavljuju s vjerojatnošću približno 0.05, tj. da je

$$P\{Z' \geq 1.64\} \approx 0.05.$$

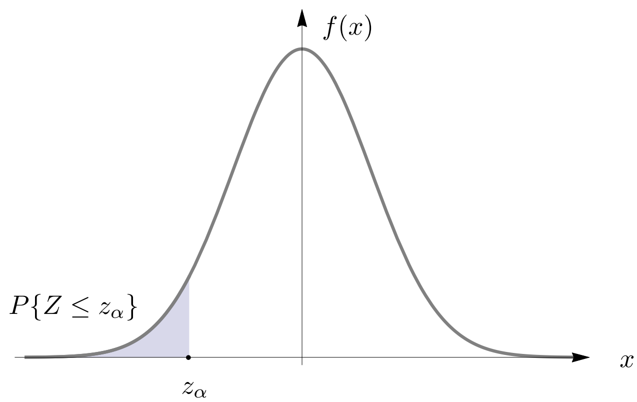


Slika 4.21: Površina ispod grafa funkcije gustoće standardne normalne distribucije nad intervalom $[1.64, \infty)$.

Pretpostavimo da se u našem slučaju Z' realizirala brojem $\hat{z}$. U uvjetima istinitosti hipoteze $\mathcal{H}_0$ očekujemo realizacije od Z' bliske 0 ili manje od 0. Veći iznosi realizacija od Z' idu u prilog alternativnoj hipotezi, ali to ne znači da se ne mogu dogoditi i ako je H_0 istinita hipoteza. Naime, ako je H_0 istinita hipoteza, vjerojatnost da se slučajna varijabla Z' realizira brojem većim ili jednakim $\hat{z}$ iznosi približno $P\{Z \geq \hat{z}\}$.

Sada zaključujemo na sljedeći način. Ako je $\mathcal{H}_0$ istinita hipoteza, realizacije veće ili jednake $\hat{z}$ mogu se pojaviti, a vjerojatnost je za to približno $P\{Z \geq \hat{z}\}$. Dakle, ako odbacimo nul-hipotezu, vjerojatnost je da ćemo time pogriješiti najviše $P\{Z \geq \hat{z}\}$. Ukoliko je taj broj manji od standardno prihvaćenih vrijednosti za maksimalnu vjerojatnost pogreške prvog tipa (tj. nivoa značajnosti testa), hipotezu $\mathcal{H}_0$ možemo odbaciti. U suprotnom kažemo da nemamo dovoljno argumenata za odbacivanje hipoteze $\mathcal{H}_0$.

Na sličan bismo način bismo proveli postupak testiranja u slučaju da je alternativna hipoteza oblika $\mathcal{H}_1 : \mu < \mu_0$. Tada vjerojatnost $p = P_{\mu_0}\{Z' \leq \hat{z}\} \approx P\{Z \leq \hat{z}\}$, $Z \sim \mathcal{N}(0,1)$, uspoređujemo s razinom značajnosti α, koja je u tom slučaju površina ispod grafa funkcije gustoće standardne normalne distribucije nad intervalom $(-\infty, z_\alpha]$ (slika 4.22).



Slika 4.22: Površina ispod grafa funkcije gustoće standardne normalne distribucije nad intervalom $(-\infty, z_\alpha]$.

Objašnjeni postupak općenito zapisujemo na sljedeći način:

- Nul-hipoteza:
$$\mathcal{H}_0 : \quad \mu = \mu_0,$$
- Test-statistika:
$$Z' = \frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}}.$$
Ovdje je n veličina uzorka, $\overline{X}_n$ aritmetička sredina uzorka, a σ poznata standardna devijacija.
- U uvjetima istinitosti nul-hipoteze očekujemo da je realizacija od Z' (označit ćemo je $\hat{z}$) blizu ili veća od 0 jer varijabla Z' ima približno standardnu normalnu distribuciju. Ako označimo $Z \sim \mathcal{N}(0,1)$, na osnovu realizacije $\hat{z}$ statistike Z' na podacima možemo odrediti takozvanu p-vrijednost kao:
$$p = P\{Z \geq \hat{z}\}, \text{ ako je alternativna hipoteza oblika } \mathcal{H}_1 : \mu > \mu_0,$$

$p = P\{Z \leq \hat{z}\}$, ako je alternativna hipoteza oblika $\mathcal{H}_1 : \mu < \mu_0$.

- Tako izračunatu p-vrijednost uspoređujemo s razinom značajnosti α. U slučaju da je $p < \alpha$, odbacujemo nul-hipotezu na razini značajnosti α. Ako je $p > \alpha$, zaključujemo da nemamo dovoljno informacija koje bi poduprle odluku o odbacivanju nul-hipoteze.

Treba napomenuti da je pretpostavka o poznatoj varijanci iz te procedure nerealna za primjene. S obzirom da je taj rezultat izveden za velike uzorke, može se također dokazati da će korištenje korigirane varijance uzorka umjesto varijance u izvodu formule za test-statistiku dati analognu asimptotsku distribuciju. Iz tog se razloga u praksi koristi procjena za varijancu umjesto stvarne vrijednosti varijance ako je ona nepoznata.

Primjer 4.55. (kolokvij.xls)
Promotrimo ponovno diskretnu slučajnu varijablu X kojom modeliramo ukupan broj bodova koje je student ostvario na dvama kolokvijima iz nekog kolegija. Iz primjera 4.38 znamo da je procjena njezinog očekivanja realan broj 96.9. Korištenjem istih podataka testirajmo hipotezu da je očekivani ukupni broj bodova ostvaren na tim dvama kolokvijima manji od 100. U tu svrhu postavimo statističke hipoteze na sljedeći način:

$$\mathcal{H}_0 : \ \mu = 100,$$
$$\mathcal{H}_1 : \ \mu < 100.$$

Da bismo donijeli odluku na nivou značajnosti $\alpha = 0.05$, računamo vrijednost $\hat{z}$ test-statistike Z' temeljenu na podacima iz varijable bodovi_ukupno*:*

$$\hat{z} = \frac{\overline{x}_{70} - \mu_0}{s_{70}/\sqrt{70}} = \frac{96.9 - 100}{44.19/\sqrt{70}} = -0.59.$$

Odavde upotrebom kalkulatora vjerojatnosti slijedi da je pripadna p-vrijednost

$$p = P\{Z < \hat{z}\} = 0.28,$$

što je veće od zadane razine značajnosti $\alpha = 0.05$. Zaključujemo da, na razini značajnosti $\alpha = 0.05$, ne možemo tvrditi da je očekivani ukupni broj bodova ostvaren na tim dvama kolokvijima manji od 100.

Primjer 4.56. (djelatnici.xls)
Neka je X neprekidna slučajna varijabla iz primjera 4.39 kojom modeliramo godišnju plaću djelatnika jedne tvornice. U primjeru 4.39 pokazali smo da je $\overline{x}_{100} = 24522$ procjena njezinog očekivanja, a u primjeru 4.41 da je $s^2_{100} = 26069208.1$ procjena varijance na temelju odabranog uzorka.
Želimo provjeriti je li očekivanje od X veće od 23000 na razini značajnosti $\alpha = 0.05$. U tu svrhu postavljamo statističke hipoteze:

$$\mathcal{H}_0 : \ \mu = 23000,$$
$$\mathcal{H}_1 : \ \mu > 23000.$$

Da bismo donijeli odluku, računamo vrijednost $\hat{z}$ test-statistike Z' na temelju podataka za varijablu placa_prije*:*

$$\hat{z} = \frac{\overline{x}_{100} - \mu_0}{s_{100}/10} = \frac{24522 - 23000}{\sqrt{26069208.1}/10} \approx 2.98.$$

Odavde upotrebom kalkulatora vjerojatnosti slijedi da je pripadna p-vrijednost

$$p = P\{Z > \hat{z}\} = 0.0014,$$

što je manje od zadane razine značajnosti $\alpha = 0.05$. Stoga zaključujemo da, na nivou značajnosti $\alpha = 0.05$, prihvaćamo alternativnu hipotezu, tj. tvrdimo da je očekivanje godišnje plaće djelatnika u promatranoj tvornici veće od 23000.

4.5.3 Testiranje hipoteze o proporciji za velike uzorke

O proporciji ćemo, u ovom primjeru statističkog testa, zaključivati na temelju n-dimenzionalnoga jednostavnog slučajnog uzorka iz Bernoullijeve distribucije. Neka je slučajni pokus modeliran Bernoullijevom slučajnom varijablom s tablicom distribucije

$$X = \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}, \quad p \in (0,1), \quad q = 1 - p.$$

Testirat ćemo hipotezu o vrijednosti parametra p koji ima značenje vjerojatnosti realizacije "uspjeha" (tj. jedinice) u jednom izvođenju danog pokusa. U tom postupku koristimo relativnu frekvenciju realiziranih "uspjeha" (jedinica) kao procjenitelj za nepoznatu vjerojatnost p i označavamo ju s $\widehat{p}$.

Postavimo nul-hipotezu na sljedeći način:

$$\mathcal{H}_0 : \quad p = p_0.$$

Iskoristimo test-statistiku

$$Z' = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}.$$

U uvjetima istinitosti nul-hipoteze centralni granični terem garantira da slučajna varijabla Z' ima približno standardnu normalnu distribuciju za velike n. Ako je $Z \sim \mathcal{N}(0,1)$, na osnovu realizacije $\hat{z}$ test-statistike Z' na našem uzorku možemo odrediti p-vrijednost kao:

- $p = P\{Z \geq \hat{z}\}$ ako je alternativna hipoteza oblika $\mathcal{H}_1 : p > p_0$,
- $p = P\{Z \leq \hat{z}\}$ ako je alternativna hipoteza oblika $\mathcal{H}_1 : p < p_0$.

Tako izračunatu p-vrijednost uspoređujemo s razinom značajnosti α. U slučaju da je $p < \alpha$, na razini značajnosti α odbacujemo nul-hipotezu $\mathcal{H}_0$. Ako je $p > \alpha$, nemamo dovoljno informacija koje bi poduprle odluku o odbacivanju nul-hipoteze.

Veličina uzorka za provođenje tog testa treba biti najmanje takva da interval

$$\left[p_0 - 3\sqrt{\frac{p_0(1-p_0)}{n}}, p_0 + 3\sqrt{\frac{p_0(1-p_0)}{n}}\right]$$

ne sadrži ni 0 ni 1.

Primjer 4.57. *Iskoristimo podatke o ocjeni studenata na usmenom dijelu ispita izabranog kolegija iz primjera 4.37. Zanima nas je li proporcija studenata koji taj usmeni ispit polažu ocjenom izvrstan manja od* 0.2*. Iz tablice 4.14 znamo da je relativna frekvencija studenata koji su ispit položili izvrsnom ocjenom* $\hat{p} = 0.16$*. Postavljamo statističke hipoteze:*

$$\mathcal{H}_0 : p = 0.2,$$
$$\mathcal{H}_1 : p < 0.2.$$

Da bismo donijeli odluku, računamo vrijednost $\hat{z}$ *test-statistike* Z'*:*

$$\hat{z} = \frac{0.16 - 0.2}{\sqrt{\frac{0.2(1-0.2)}{65}}} = -0.81.$$

Odavde slijedi da je pripadna p-vrijednost

$$p = P\{Z < \hat{z}\} = 0.21,$$

što je veće od zadane razine značajnosti $\alpha = 0.05$*. Zaključujemo da na razini značajnosti* $\alpha = 0.05$ *ne možemo poduprijeti hipotezu da je proporcija studenata koji taj usmeni ispit polažu ocjenom izvrstan manja od* 20%*.*

Primjer 4.58. (kolokvij.xls)
Vratimo se bazi podataka kolokvij.xls *i testirajmo je li proporcija studenata koji su na kolokvijima ostvarili ukupno barem* 100 *bodova veća od* 50% *na razini značajnosti* $\alpha = 0.05$*. Iz primjera 4.49 znamo da je relativna frekvencija takvih studenata* $\hat{p} = 0.56$*, pa su statističke hipoteze oblika*

$$\mathcal{H}_0 : p = 0.5,$$
$$\mathcal{H}_1 : p > 0.5.$$

Računamo vrijednost $\hat{z}$ *test-statistike* Z'*:*

$$\hat{z} = \frac{0.56 - 0.5}{\sqrt{\frac{0.25}{70}}} = 1.004.$$

Odavde slijedi da je pripadna p-vrijednost

$$p = P\{Z > \hat{z}\} = 0.16$$

što je veće od zadane razine značajnosti $\alpha = 0.05$*. Stoga zaključujemo da ne odbacujemo nul-hipotezu na razini značajnosti* $\alpha = 0.05$*, tj. na nivou značajnosti* $\alpha = 0.05$ *ne možemo tvrditi da više od* 50% *studenata postiže barem* 100 *bodova na kolokvijima.*

Primjer 4.59. (djelatnici.xls)
Vratimo se primjeru 4.39 i testirajmo je li proporcija zaposlenika koji imaju godišnju plaću manju od 30000 *veća od 75% na razini značajnosti* $\alpha = 0.05$. *Iz tablice frekvencija za varijablu* placa_prije *vidimo da je relativna frekvencija takvih djelatnika* $\hat{p} = 0.86$, *pa su statističke hipoteze oblika*

$$\mathcal{H}_0 : p = 0.75,$$
$$\mathcal{H}_1 : p > 0.75.$$

Računamo vrijednost $\hat{z}$ *test-statistike* Z'*:*

$$\hat{z} = \frac{0.86 - 0.75}{\sqrt{\frac{0.75(1-0.75)}{100}}} = 2.54.$$

Odavde slijedi da je pripadna p-vrijednost

$$p = P\{Z > \hat{z}\} = 0.0055,$$

što je manje od zadane razine značajnosti $\alpha = 0.05$. *Stoga zaključujemo da odbacujemo nul-hipotezu na razini značajnosti* $\alpha = 0.05$, *tj. na razini značajnosti* $\alpha = 0.05$ *možemo tvrditi da više od 75% djelatnika te tvornice ostvaruje godišnju plaću manju od* 30000.

4.5.4 Testiranje hipoteze o jednakosti očekivanja

Do sada prezentirani testovi kreirani su na temelju najjednostavnijega statističkog modela, tj. modela jednostavnog slučajnog uzorka iz zadane jednodimenzionalne distribucije. U ovom poglavlju navodimo dva testa koja se na prvi pogled ne uklapaju u takav model, ali se prikladnim pristupom problemu mogu svesti na njega.

Zavisni uzorci

Za ilustraciju problema navodimo prvo primjer.

Primjer 4.60. *Farmaceutska tvrtka želi testirati učinkovitost novog lijeka na smanjenje razine triglicerida u krvi. U tu svrhu odabire osobe u reprezentativnu skupinu za testiranje i izvodi mjerenje razine triglicerida prije uzimaja lijeka. Iste osobe koriste lijek u zadano vrijeme. Nakon toga se ponovo mjeri razina triglicerida. Dobiveni podaci prikazani su sljedećom tablicom:*

pacijent	prije	poslije
1	x_1	y_1
2	x_2	y_2
$\vdots$	$\vdots$	$\vdots$
n	x_n	y_n

Istraživači bi htjeli dobiti odgovor na pitanje dolazi li do promjene razine triglicerida nakon uzimanja lijeka.

Analogni problemi pojavljuju se često u primjenama. Kažemo da u takvim slučajevima analiziramo istu karakteristiku (varijablu) prije i nakon tretmana.

Statistički model za zaključivanje možemo postaviti korištenjem slučajnog vektora tako da parove (x_i, y_i), $i = 1, \ldots, n$, smatramo međusobno nezavisnim realizacijama slučajnog vektora (X, Y), gdje komponenta X opisuje varijablu na populaciji prije tretmana, a Y varijablu na populaciji nakon tretmana. Tako nastao slučajni uzorak $((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n))$ niz je nezavisnih jednako distribuiranih slučajnih vekotra čije komponente X_i i Y_i, $i \in \{1, \ldots, n\}$, ne moraju biti nezavisne, što komplicira statistički model. Međutim, problem analize razlika u distribuciji varijable prije i poslije tretmana može se postaviti i tako da formiramo slučajnu varijablu razlike:

$$D = X - Y.$$

Ako je očekivanje slučajne varijable D različito od nule, to je svakako dokaz da su razlike u distribuciji između X i Y bitne. U tu svrhu proširimo tablicu podataka novom varijablom "razlike" i promatramo statistički model jednostavnoga slučajnog uzorka iz distribucije slučajne varijable D.

br	tretman 1	tretman 2	razlike
1	x_1	y_1	$d_1 = x_1 - y_1$
2	x_2	y_2	$d_2 = x_1 - y_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	x_n	y_n	$d_n = x_1 - y_n$

Postavimo hipotezu:

$$\mathcal{H}_0 : \mu_D = 0.$$

Ukoliko je uzorak dovoljno velik, u poglavlju 4.5.2 opisali smo proceduru za testiranje te hipoteze u kombinaciji s alternativnom hipotezom $\mu_D > 0$ ili $\mu_D < 0$. U tom slučaju za varijancu od D možemo koristiti korigiranu varijancu uzorka iz distribucije od D, kao što je navedeno u poglavlju 4.5.2.

Primjer 4.61. (djelatnici.xls)
U bazi podataka djelatnici.xls *raspolažemo podacima o godišnjim plaćama za uzorak od* 100 *djelatnika jedne tvornice (tvornice A). Varijabla* placa_prije *sadrži iznose godišnjih plaća za djelatnike iz uzorka prije reorganizacije poslovnog sustava, a varijabla* placa_poslije *iznose godišnjih plaća za isti uzorak djelatnika nakon reorganizacije. Budući da se ovdje radi o istom uzorku djelatnika čije godišnje plaće pratimo prije i poslije reorganizacije posla, kažemo da se radi o zavisnim uzorcima. Označimo s $X^{(1)}$ slučajnu varijablu kojom modeliramo godišnju plaću djelatnika tvornice A prije reorganizacije, a s $X^{(2)}$ slučajnu varijablu kojom modeliramo godišnju plaću djelatnika te tvornice poslije reorganizacije. Zanima nas razlikuju li se na nivou značajnosti $\alpha = 0.05$ očekivanja plaća djelatnika te tvornice prije i poslije reorganizacije posla. Prije nego postavimo potrebne statističke hipoteze, procijenimo, na temelju realizacija jednostavnih slučajnih uzoraka iz distribucija od $X^{(1)}$ i $X^{(2)}$, očekivanja i varijance od $X^{(1)}$, $X^{(2)}$ i $D = X^{(1)} - X^{(2)}$. Procjene su dane u tablici 4.17.*

slučajna varijabla	procjena očekivanja	procjena varijance
$X^{(1)}$	24522	26069208.1
$X^{(2)}$	24986.85	26252789.5
D	−464.85	116302.63

Tablica 4.17: Procjene očekivanja i varijanci slučajnih varijabli $X^{(1)}$ i $X^{(2)}$.

Postavimo statističke hipoteze:

$$\mathcal{H}_0\colon\ \mu_D = 0,$$
$$\mathcal{H}_1\colon\ \mu_D < 0.$$

Da bismo donijeli odluku o tome odbacujemo li na nivou značajnosti $\alpha = 0.05$ nul-hipotezu ili ne, računamo vrijednost $\hat{z}$ test-statistike Z' temeljem procjena očekivanja i varijance slučajne varijable D i vrijednosti $\mu_0 = 0$:

$$\hat{z} = \frac{-464.85}{\sqrt{116302.63/10}} = -13.63.$$

Odavde, upotrebom kalkulatora vjerojatnosti, slijedi da je pripadna p-vrijednost

$$p = P\{Z < \hat{z}\} < 10^{-6},$$

što je manje od zadanog nivoa značajnosti $\alpha = 0.05$, stoga odbacujemo nul-hipotezu na nivou značajnosti $\alpha = 0.05$ i prihvaćamo alternativnu hipotezu, tj. tvrdimo da je očekivanje godišnje plaće djelatnika tvornice A prije reorganizacije bilo manje od očekivanja godišnje plaće djelatnika iste tvornce nakon reorganizacije.

Nezavisni uzorci

Za ilustraciju problema prvo navedimo jedan primjer.

Primjer 4.62. *Ured za kvalitetu na nekom fakultetu želi provjeriti je li došlo do bitne promjene u distribuciji ocjene iz kolegija Matematika 1 koju su ostvarili studenti generacije 2009./2010. u odnosu na generaciju 2008./2009. U tu svrhu prikuplja podatke o ocjenama na uzorcima studenata i modelira ocjenu iz tog kolegija kao slučajnu varijablu i to X za generaciju 2008./2009., a Y za 2009./2010. Također, pretpostavlja da su X i Y nezavisne slučajne varijable. Prikupljeni podaci o ocjenama $(x_1, \ldots, x_n)$ i $(y_1, \ldots, y_m)$ realizacije su jednostavnih slučajnih uzoraka iz distrbucije od X, odnosno iz distribucije od Y.*

Statistički model opisan u tom primjeru sastoji se od distribucija slučajnog vektora $(X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m)$ nezavisnih, ali ne nužno jednako distribuiranih slučajnih varijabli. Međutim, prvi dio vektora, tj. $(X_1, X_2, \ldots, X_n)$ jest jednostavni slučajni uzorak iz distribucije slučajne varijable X, dok je drugi dio, $(Y_1, Y_2, \ldots, Y_m)$, jednostavni slučajni uzorak iz distribucije slučajne varijable Y.

Jednostavan način potvrde postojanja razlike u distribucijama potvrda je postojanja razlike u očekivanjima tih distribucija. Ovdje ćemo navesti jedan test kojim možemo testirati upravo takve hipoteze, tj. hipoteze o razlici u očekivanjima.

U tu svrhu pretpostavimo da su veličine uzoraka (n i m) velike te uočimo da su $\overline{X}_n$ i $\overline{Y}_m$ dvije slučajne varijable s asimptotski normalnim distribucijama. Osim toga, $E(\overline{X}_n) = \mu_X$, dok je $E(\overline{Y}_m) = \mu_Y$, $\operatorname{Var}\overline{X}_n = \frac{\sigma_X^2}{n}$, $\operatorname{Var}\overline{Y}_m = \frac{\sigma_Y^2}{m}$, a $\overline{X}_n$ i $\overline{Y}_m$ nezavisne su slučajne varijable.

Neka je $D = \overline{X}_n - \overline{Y}_m$. Uočimo da je

$$E\,D = \mu_X - \mu_Y, \quad \operatorname{Var} D = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}.$$

Sada možemo testirati slutnju o postojanju razlika u očekivanjima korištenjem distribucije slučajne varijable D. Postavimo statističku hipotezu:

$$\mathcal{H}_0 : \mu_D = 0.$$

Statistička teorija dokazuje da, u uvjetima istinitosti $\mathcal{H}_0$, standardizirani oblik slučajne varijable D, tj. slučajna varijabla

$$Z' = \frac{D}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$$

ima asimptotski standardnu normalnu distribuciju. Ukoliko varijance nisu unaprijed poznate (što je u praksi slučaj), moramo modificirati Z' korištenjem varijance uzorka kao procjenitelja za varijancu, pa koristimo statistiku

$$Z'' = \frac{D}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}},$$

koja također ima asimptotski standardnu normalnu distribuciju ako je $\mathcal{H}_0$ istinita hipoteza.

Korištenjem statistike Z'' testiramo hipotezu $\mathcal{H}_0$ u kombinaciji s alternativnom hipotezom $\mathcal{H}_1{:}\mu_D > 0$ ili $\mathcal{H}_1{:}\mu_D < 0$ analognim postupkom kao u poglavlju 4.5.2.

Primjer 4.63. (djelatnici.xls)
U bazi podataka djelatnici.xls *raspolažemo podacima o godišnjim plaćama za uzorke od po* 100 *djelatnika iz dviju konkurentskih tvornica - tvornice A (varijable* placa_prije *i* placa_poslije*, pogledati primjer 4.61) i tvornice B (varijabla* placa_konkurencija*). Budući da se radi o uzorcima djelatnika iz dviju različitih tvornica, zaključujemo da se radi o nezavisnim uzorcima.*
Označimo ponovno s $X^{(1)}$ i $X^{(2)}$ slučajne varijable kojima modeliramo godišnje plaće djelatnika tvornice A prije i poslije reorganizacije posla, redom, te s Y slučajnu varijablu kojom modeliramo godišnju plaću djelatnika tvornice B. U varijablama placa_prije*,* placa_poslije *i* placa_konkurencija *sadržane su realizacije jednostavnih slučajnih uzoraka*

$$(X_1^{(1)}, \dots, X_{100}^{(1)}),\ (X_1^{(2)}, \dots, X_{100}^{(2)}) \text{ i } (Y_1, \dots Y_{100})$$

iz distribucija slučajnih varijabli $X^{(1)}$, $X^{(2)}$ i Y, redom. Zanima nas razlikuje li se, na nivou značajnosti $\alpha = 0.01$, očekivanje plaće djelatnika tvornice A prije reorganizacije posla od očekivanja plaće djelatnika tvornice B. Potražimo također i odgovor na analogno pitanje za očekivanje plaće djelatnika tvornice A nakon reorganizacije.
Procjene očekivanja i varijanci slučajnih varijabli $X^{(1)}$ i $X^{(2)}$ dane su u tablici 4.17. Procjene očekivanja i varijanci slučajnih varijabli Y, $D_1 = \overline{X}^{(1)}_{100} - \overline{Y}_{100}$ i $D_2 = \overline{X}^{(2)}_{100} - \overline{Y}_{100}$ dane su u tablici 4.18.

slučajna varijabla	procjena očekivanja	procjena varijance
Y	25432.24	316895.23
D_1	−910.24	26004549.7
D_2	−445.39	26215378.2

Tablica 4.18: Procjene očekivanja i varijanci slučajnih varijabli Y, D_1 i D_2.

Postavimo statističke hipoteze:

$$\mathcal{H}_0\colon\ \mu_{D_i} = 0,$$
$$\mathcal{H}_1\colon\ \mu_{D_i} < 0,$$

gdje je $i \in \{1,2\}$. Da bismo donijeli odluku o tome odbacujemo li na nivou značajnosti $\alpha = 0.01$ nul-hipoteze ili ne, računamo vrijednosti $\hat{z}_i$ test-statistike Z'' na temelju procjena očekivanja i varijanci slučajnih varijabli D_1 i D_2:

$$\hat{z}_1 = \frac{-910.24}{\sqrt{26004549.7/100}} = -1.78, \qquad \hat{z}_2 = \frac{-445.39}{\sqrt{26215378.2/100}} = -0.87.$$

Odavde, upotrebom kalkulatora vjerojatnosti, slijedi da su pripadne p-vrijednosti redom

$$p_1 = P\{Z < \hat{z}_1\} = 0.038, \qquad p_2 = P\{Z < \hat{z}_2\} = 0.19.$$

Budući da je p_2 veći od 0.01 na nivou značajnosti $\alpha = 0.01$, ne možemo tvrditi da je očekivanje godišnje plaće djelatnika tvornice A nakon reorganizacije manje od očekivanja godišnje plaće djelatnika tvornice B. Osim toga, p_1 je također veći od 0.01, pa takvu tvrdnju ne možemo poduprijeti niti ako uspoređujemo godišnje plaće djelatnika tvornice A prije reorganizacije s godišnjim plaćama djelatnika tvornice B. Uočimo da p_2 ima veću "težinu" od p_1 u neodbacivanju hipoteze $\mathcal{H}_0$.

4.6 Zadaci

Zadatak 4.2. (tlak.xls)
Baza podataka tlak.xls sadrži podatke o krvnom tlaku dobivene anketiranjem reprezentativnog uzorka pacijenata jedne liječničke ordinacije:

varijable spol i dob sadrže informacije o spolu i broju godina za svakog ispitanika,

varijable sistolicki-tlak i dijastolicki-tlak sadrže vrijednosti sistoličkog i dijastoličkog tlaka za svakog ispitanika, varijabla tlak klasificira vrijednosti sistoličkog i dijastoličkog tlaka u tri kategorije: N - nizak tlak, O - normalan tlak, P - povišen tlak,

varijabla puls sadrži broj otkucaja srca u minuti (puls) za svakog ispitanika,

varijabla opce-stanje sadrži subjektivnu ocjenu (u standardnoj skali od 1 do 5) vlastitog zdravstvenog stanja svakog ispitanika.

Na temelju podataka sadržanih u toj bazi riješite sljedeće zadatke.

a) Odredite tablice frekvencija i relativnih frekvencija, nacrtajte i proanalizirajte histograme frekvencija i relativnih frekvencija te kružni dijagram s prikazom relativnih frekvencija za podatke sadržane u varijabli opce-stanje. Kolike su frekvencija i relativna frekvencija ispitanika koji su svoje opće zdravstveno stanje ocijenili barem ocjenom 4?
Rješenje: 39 ispitanika, tj. njih je 78% svoje opće zdravstveno stanje ocijenilo barem ocjenom 4.

b) Odredite tablice frekvencija i relativnih frekvencija za podatke sadržane u varijabli opce-stanje posebno za kategoriju ispitanika ženskog spola i kategoriju ispitanika muškog spola te nacrtajte pripadne histograme frekvencija i relativnih frekvencija. Također, nacrtajte histograme frekvencija i relativnih frekvencija za podatke sadržane u varijabli opce-stanje kategorizirane po vrijednostima varijable tlak (N, O, P). Proanalizirajte dobivene histograme.

c) Odredite i ukratko protumačite sljedeće numeričke karakteristike podataka sadržanih u varijabli dob: aritmetičku sredinu, medijan, donji i gornji kvartil, mod, raspon i standardnu devijaciju. Je li mod jedinstven? Koliko iznosi maksimalno odstupanje podataka sadržanih u varijabli dob od njihove aritmetičke sredine? Nacrtajte i detaljno proanalizirajte kutijasti dijagram na bazi medijana za podatke sadržane u varijabli dob. Obrazložite svoj odgovor.
Rješenje: $\overline{x}_{50} = 40.22$, $x_{min} = 11$, $x'_{25} = 23$, $x'_{50} = 43$, $x'_{75} = 51$, $x_{max} = 83$, $s_{50} = 17.69$.

d) Nacrtajte i detaljno proanalizirajte kutijasti dijagram na bazi medijana za podatke sadržane u varijabli dob. Obrazložite svoj odgovor.

e) Crtanjem i analizom kutijastog dijagrama na bazi medijana neosjetljivog na stršeće vrijednosti i kutijastog dijagrama na bazi medijana osjetljivog na stršeće vrijednosti donesite zaključak o tome pojavljuju li se među podacima sadržanima u varijabli puls stršeće vrijednosti ili ne. Ako ste se uvjerili u njihovo postojanje, korištenjem kategoriziranih tablica frekvencija odredite sve prisutne stršeće vrijednosti među podacima u varijabli puls. Kako biste neutralizirali njihov utjecaj na numeričke karakteristike podataka?
Rješenje: stršeće su vrijednosti varijable puls 800 i 1006.

Zadatak 4.3. (glukoza.xls)
Baza podataka (glukoza.xls) za reprezentativan uzorak pacijenata jedne internističke klinike sadrži informacije o dobi (varijabla dob), koncentraciji glukoze u krvi (varijabla koncentracija) i tome je li izmjerena koncentracija glukoze normalna ili povišena (varijabla kategorija: N - normalna koncentracije, P - povišena koncentracija). Riješite sljedeće zadatke i sva rješenja interpretirajte u kontekstu promatranog problema.

a) Procijenite proporciju pacijenta promatrane klinike koji su stariji od 30 godina te proporciju pacijenata starih barem 50 ali manje od 60 godina.
Rješenje: proporcija je pacijenta starijih od 30 godina $90/102 = 0.88$, a proporcija pacijenata starih barem 50, ali manje od 60 godina jest $21/102 = 0.21$.

b) Intervalom pozdanosti 95% procijenite proporciju pacijenata koji imaju povišenu koncentraciju glukoze u krvi.
Rješenje: $[0.61, 0.79]$.

c) Procijenite očekivanje, varijancu i standardnu devijaciju slučajne varijable kojom modeliramo koncentraciju glukoze u krvi pacijenata.
Rješenje: $\overline{x}_{102} = 7.69$, $s^2_{102} = 0.75$, $s_{102} = 2.78$.

d) Intervalom pozdanosti 95% procijenite očekivanu koncentraciju glukoze u krvi pacijenata.
Rješenje: $[7.15, 8.24]$.

e) Možete li na nivou značajnosti $\alpha = 0.01$ tvrditi da je proporcija pacijenata koji imaju povišenu koncentraciju glukoze manja od 0.8? Koji ste test koristili i zašto?
Rješenje: $p = 0.0044$.

f) Možete li na nivou značajnosti $\alpha = 0.05$ tvrditi da je očekivana koncentracija glukoze u krvi veća od normalne, tj. 6 mMol/L? Koji ste test koristili i zašto?
Rješenje: $p = 0$.

g) Možete li na nivou značajnosti $\alpha = 0.05$ tvrditi da je očekivana dob pacijenata koji imaju povišenu koncentraciju glukoze u krvi veća od očekivane dobi onih koji imaju normalnu koncentraciju glukoze? Koji ste test koristili i zašto?
Rješenje: $p = 0.59$.

Zadatak 4.4. (zdravlje.xls)
Baza podataka zdravlje.xls sadrži neke podatke o zdravstvenom stanju reprezentativnog uzorka stanovnika jednog mjesta:

varijable godine i spol sadrže podatke o starosti u godinama i spolu ispitanika;

varijabla zdravlje sadrži subjektivne ocjene vlastitog zdravstvenog stanja ispitanika;

varijabla broj-pregleda sadrži informacije o ukupnom broju zdravstvenih pregleda svakog ispitanika u tekućoj kalendarskoj godini;

varijabla dodatno-zdravstveno sadrži podatke o dodatnom zdravstvenom osiguranju svakog ispitanika (1 - ispitanik je dodatno osiguran; 0 - ispitanik nije dodatno osiguran);

varijabla cijena sadrži cijenu u kunama najskupljeg zdravstvenog pregleda svakog ispitanika (u tekućoj kalendarskoj godini).

Riješite sljedeće zadatke i sva rješenje interpretirajte u kontekstu promatranog problema.

a) Procijenite proporciju stanovnika promatranog mjesta koji su svoje zdravstveno stanje ocijenili ocjenom većom od dva, ali manjom of pet.
Rješenje: proporcija stanovnika koji su svoje zdravstveno stanje ocijenili ocjenom većom od dva, ali manjom od pet $30/51 = 0.59$.

b) Intervalom pozdanosti 97% procijenite proporciju stanovnika koji nemaju dodatno zdravstveno osiguranje.
Rješenje: $[0.59, 0.86]$.

c) Procijenite očekivanje, varijancu i standardnu devijaciju slučajne varijable kojom modeliramo subjektivnu ocjenu zdravstvenog stanja stanovnika.
Rješenje: $\overline{x}_{51} = 3.27$, $s^2_{51} = 1.36$, $s_{51} = 1.17$.

d) Intervalom pozdanosti 95% procijenite očekivanu ocjenu zdravstvenog stanja stanovnika.
Rješenje: $[2.95, 3.6]$.

e) Možete li na razini značajnosti $\alpha = 0.01$ tvrditi da je proporcija stanovnika koji imaju dopunsko zdravstveno osiguranje manja od 0.15? Koji ste test koristili i zašto?
Rješenje: $p = 0.421$.

g) Možete li na nivou značajnosti $\alpha = 0.05$ tvrditi da je očekivana dob stanovnika koji nemaju dopunsko zdravstveno osiguranje manja od očekivane dobi onih koji imaju dopunsko zdravstveno osiguranje? Koji ste test koristili i zašto? Što zaključujete ako za razinu značajnosti testa uzmete $\alpha = 0.01$ i zašto?
Rješenje: $p = 0.03$.

Zadatak 4.5. (matematika.xls)
Baza podataka (matematika.xls) sadrži podatke prikupljene anketiranjem reprezentativnog uzorka studenata jedne generacije nakon održanih predavanja, vježbi, kolokvija te usmenog ispita iz jednog matematičkog kolegija. Prikupljeni podaci organizirani su na sljedeći način:

varijabla prosjek sadrži podatke o prosječnoj ocjeni na studiju za svakog od anketiranih studenata,

varijabla polozeno za svakog anketiranog studenta sadrži informaciju o tome je li položio usmeni ispit iz promatranog kolegija (oznaka 1) ili nije (oznaka 0),

varijable predavanja i vjezbe sadrže informaciju o redovitosti pohađanja nastave iz promatranog kolegija (oznaka 1 - student s p/v nije nikada izostao, oznaka 2 - student je s p/v izostao samo jednom, oznaka 3 - student je s p/v izostao barem dva puta),

varijable tezina_kolegija i materijali sadrže subjektivne ocjene (u standardnoj skali od 1 do 5) promatranih studenata za težinu kolegija i dostatnost dostupnih materijala za pripremanje ispita iz promatranog kolegija.

Riješite sljedeće zadatke i sva rješenja interpretirajte u kontekstu promatranog problema.

a) Procijenite proporciju studenata čija je prosječna ocjena na studiju veća od tri.
Rješenje: proporcija studenata čija je prosječna ocjena studiranja veća od tri jest $42/49 = 0.86$.

b) Intervalom pozdanosti 95% procijenite proporciju studenata čija je prosječna ocjena na studiju najviše tri.
Rješenje: $[0.045, 0.24]$.

c) Procijenite očekivanje, varijancu i standardnu devijaciju slučajne varijable kojom modeliramo prosječnu ocjenu na studiju kojeg pohađaju ispitanici.
Rješenje: $\overline{x}_{49} = 3.98$, $s_{49}^2 = 0.565$, $s_{49} = 0.752$.

d) Intervalom pozdanosti 95% procijenite očekivanu prosječnu ocjenu na studiju kojeg pohađaju ispitanici.
Rješenje: $[3.76, 4.19]$.

e) Možete li na razini značajnosti $\alpha = 0.01$ tvrditi da je proporcija studenata koji su redovito pohađali predavanja veća od 0.7? Koji ste test koristili i zašto?
Rješenje: $p = 0.413$.

Zadatak 4.6. (komarci.xls)
Baza podataka komarci.xls sadrži dio rezultata proučavanja komaraca u jednom močvarnom području (dostupni su podaci za 210 mjerenja na istoj lokaciji):

varijable brojM i brojZ redom sadrže broj muških i ženskih jedinki komaraca uhvaćenih u klopku;

varijabla mjesec sadrži mjesečevu mijenu (M - mlađak, U - uštap) za svako mjerenje;

varijabla doba-dana sadrži doba dana u kojem je mjerenje obavljeno (P - predvečerje, N - noć, S - svitanje);

varijabla svjetlost sadrži tip osvjetljenja pri mjerenju;

varijabla temperatura sadrži temperaturu zraka pri kojoj je mjerenje izvršeno;

varijabla rel-vlaznost sadrži relativnu vlažnost zraka za vrijeme mjerenja.

Riješite sljedeće zadatke i sva rješenja interpretirajte u kontekstu promatranog problema.

a) Procijenite proporciju mjerenja u kojima je izbrojeno manje od 50 muških jedinki komaraca te proporciju mjerenja u kojima je izbrojeno više od 50 ženskih jedinki komaraca.
Rješenje: proporcija mjerenja u kojima je izbrojeno manje od 50 muških jedinki komaraca jest $202/210 = 0.962$, a proporcija je mjerenja u kojima je izbrojeno više od 50 ženskih jedinki komaraca je $39/210 = 0.186$.

b) Intervalom pozdanosti 95% procijenite proporciju mjerenja u kojima je izbrojeno barem 50, ali manje od 100 ženskih jedinki komaraca.
Rješenje: $[0.022, 0.083]$.

c) Procijenite očekivanje, varijancu i standardnu devijaciju slučajne varijable kojom modeliramo temperaturu zraka tijekom jednog mjerenja.
Rješenje: $\overline{x}_{210} = 21.89$, $s^2_{210} = 6.76$, $s_{210} = 2.6$.

d) Intervalom pozdanosti 97% procijenite očekivanu temperaturu tijekom jednog mjerenja.
Rješenje: $[21.51, 22.29]$.

e) Možete li na nivou značajnosti $\alpha = 0.05$ tvrditi da je proporcija mjerenja u kojima je izbrojeno manje od 50 muških jedinki komaraca veća od 0.95? Koji ste test koristili i zašto?
Rješenje: $p = 0.2$.

f) Možete li na nivou značajnosti $\alpha = 0.01$ tvrditi da je očekivani broj muških jedinki komaraca izbrojenih u mjerenju manji od 20? Koji ste test koristili i zašto?
Rješenje: $p = 0.99$.

g) Možete li na nivou značajnosti $\alpha = 0.05$ tvrditi da se očekivani broj muških jedinki komaraca izbrojenih u jednom mjerenju razlikuje od očekivanog broja ženskih jedinki izbrojenih u jednom mjerenju? Koji ste test koristili i zašto?
Rješenje: $p = 0.009$.

Zadatak 4.7. (gradjevina.xls)
Baza podataka gradjevina.xls sadrži neke informacije o reprezentativnom uzorku koji se sastoji od 100 srednje velikih građevinskih poduzeća u jednoj tranzicijskoj zemlji:

varijabla godina_osnivanja za svako od 100 poduzeća iz uzorka sadrži godinu kada je poduzeće osnovano,

varijable zaposleni2007, zaposleni2008 i zaposleni2009 sadrže podatke o broju zaposlenika u tih 100 poduzeća u 2007., 2008. i 2009. godini,

varijabla prosjecna_starost sadrži prosjecnu dob zaposlenika u 2009. godini,

varijable motivacija_placa i napredovanje redom sadrže subjektivne ocjene kadrovske službe poduzeća o tome u kolikoj je mjeri visina plaće motivacijski faktor za uspješno obavljanje posla te u kolikoj mjeri uspješno obavljanje poslovnih zadataka utječe na mogućnost napredovanja na bolje radno mjesto unutar poduzeća, varijable placa2007, placa2008 i placa2009 sadrže iznose prosječnih plaća zaposlenika u 2007., 2008. i 2009. godini,

varijable troskovi2007, troskovi2008 i troskovi2009 sadrže iznose troškova poduzeća u 2007., 2008. i 2009. godini,

varijable prihodi2007, prihodi2008 i prihodi2009 sadrže iznose ostvarenih prihoda u 2007., 2008. i 2009. godini,

Riješite sljedeće zadatke i sva rješenja interpretirajte u kontekstu promatranog problema.

a) Odredite tipove i pripadne skupove vrijednosti slučajnih varijabli kojima modeliramo broj zaposlenika, njihovu prosječnu dob, prosječnu godišnju plaću zaposlenika u poduzeću, godišnje troškove poduzeća te godišnje prihode poduzeća.

b) Procijenite vjerojatnost da slučajno odabrano srednje veliko građevinsko poduzeće u promatranoj zemlji ima više od 50 zaposlenika u 2007., 2008. te 2009. godini?
Rješenje: proporcije su 0.83 za 2007., 0.93 za 2008. te 0.95 za 2009. godinu.

c) Procijenite očekivanje i varijancu slučajne varijable kojom se modelira prosječna plaća zaposlenika u srednje velikom građevinskom poduzeću u toj zemlji u 2009. godini.
Rješenje: $\overline{x}_{100} = 600.13$, $s_{100} = 194.63$.

d) Kategorizirajte podatke kojima raspolažete te odlučite ima li smisla modelirati prosječnu godišnju plaću u 2009. godini kao normalnu slučajnu varijablu. Ako smatrate da ima, korištenjem normalne distribucije s procijenjenim vrijednostima očekivanja i varijance odredite vjerojatnost da je u 2009. godini u slučajno odabranom poduzeću srednje veličine u toj zemlji prosječna plaća bila viša od 500 eura. Istu vjerojatnost izračunajte i korištenjem procijenjene (empirijske) distribucije te slučajne varijable te usporedite rezultate.
Rješenje: Iz histograma relativnih frekvencija vidimo da normalna distribucija nije prikladna za modeliranje tih podataka, a to sugeriraju i izračunate tražene vjerojatnosti: iz empirijske distribucije te slučajne varijable, označimo ju s X, slijedi da je $P(X > 500) = 0.66$, a ako X modeliramo kao $\mathcal{N}(600.13, 194.63^2)$, slijedi da je $P\{X > 500\} = 0.3$.)

e) Intervalom pouzdanosti 95 % procijenite proporciju srednje velikih građevinskih poduzeća u toj zemlji u kojima je prosječna mjesečna plaća veća od aritmetičke sredine plaća zabilježenih u varijabli placa2009.
Rješenje: $[0.343, 0.537]$.

f) Procijenite očekivanje, varijancu i standardnu devijaciju slučajne varijable kojom modeliramo prosječnu plaću u 2009. godini.
Rješenje: $\overline{x}_{100} = 600.13$, $s^2_{100} = 37879.1$, $s_{100} = 194.63$.

g) Intervalom pouzdanosti 95 % procijenite očekivanje slučajne varijable kojom se modelira prosječna mjesečna plaća zaposlenika u 2009. godini.
Rješenje: $[561.51, 638.75]$.

h) Možete li na razini značajnosti $\alpha = 0.05$ tvrditi da je proporcija zaposlenika koji u 2009. godini imaju plaću višu od očekivane manja od 0.5? Koji ste test koristili i zašto?
Rješenje: $p = 0.12$.

i) Možete li na razini značajnosti $\alpha = 0.05$ tvrditi da je očekivana plaća u 2009. godini veća od 650 eura? Koji ste test koristili i zašto?
Rješenje: $p = 0.012$.

j) Možete li na razini značajnosti $\alpha = 0.05$ tvrditi da postoji razlika u očekivanoj prosječnoj plaći u građevinskim poduzećima srednje veličine u toj zemlji u 2008. i 2009. godini pod pretpostavkom da razlike prosječnih plaća u 2008. i 2009. godini možemo modelirati normalnom slučajnom varijablom? Koji ste test koristili i zašto?
Rješenje: $p = 0.164$.

k) Možete li na razini značajnosti $\alpha = 0.05$ tvrditi da je proporcija srednje velikih građevinskih poduzeća u toj zemlji, koja imaju više od 150 zaposlenih, veća za 2009. nego za 2008. godinu? Koji ste test koristili i zašto?
Rješenje: $p = 0.4245$.

Poglavlje 5

Dodatak

5.1 Osnove algebre skupova

Skup A podskup je skupa B ($A \subseteq B$) ako je svaki element skupa A ujedno element i skupa B.

Skup A jednak je skupu B ako je $A \subseteq B$ i $B \subseteq A$.

Unija skupova A i B skup je $A \cup B = \{\omega \in \Omega : \omega \in A \vee \omega \in B\}$.

Presjek skupova A i B skup je $A \cap B = \{\omega \in \Omega : \omega \in A \wedge \omega \in B\}$.

Razlika skupova A i B skup je $A \setminus B = \{\omega \in \Omega : \omega \in A \wedge \omega \notin B\}$.

Simetrična razlika skupova A i B skup je $A \triangle B = (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$.

Komplement skupa (događaja) $A \subseteq \Omega$ skup je $A^C = \{\omega \in \Omega : \omega \notin A\}$ kojeg nazivamo suprotan događaj događaja A.

Komplement prostora elementarnih događaja jest prazan skup, tj. $\Omega^C = \emptyset$. Cijeli prostor elementarnih događaja Ω nazivamo **siguran događaj**, a njegov komplement **nemoguć događaj**.

Konačna unija skupova (događaja) $A_1, A_2, \ldots A_n \subseteq \Omega$ skup je

$$A_1 \cup A_2 \cup \ldots \cup A_n = \bigcup_{i=1}^{n} A_i = \{\omega \in \Omega : \omega \in A_1 \vee \omega \in A_2 \vee \ldots \vee \omega \in A_n\} =$$
$$= \{\omega \in \Omega : \exists i \in \{1, 2, \ldots, n\} \text{ t.d. } \omega \in A_i\}.$$

Konačan presjek skupova (događaja) $A_1, A_2, \ldots A_n \subseteq \Omega$ skup je

$$A_1 \cap A_2 \cap \ldots \cap A_n = \bigcap_{i=1}^{n} A_i = \{\omega \in \Omega : \omega \in A_1 \wedge \omega \in A_2 \wedge \ldots \wedge \omega \in A_n\} =$$
$$= \{\omega \in \Omega : \omega \in A_i, \forall i \in \{1, 2, \ldots, n\}\}.$$

Osnovna svojstva skupovnih operacija

1.	$A \cup B = B \cup A$	Komutativnost
2.	$A \cap B = B \cap A$	
3.	$(A \cup B) \cup C = A \cup (B \cup C)$	Asocijativnost
4.	$(A \cap B) \cap C = A \cap (B \cap C)$	
5.	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	Distributivnost
6.	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	
7.	$(A \cup B)^C = A^C \cap B^C$	De Morganovi zakoni
8.	$(A \cap B)^C = A^C \cup B^C$	

Tablica 5.1: Osnovna svojstva skupovnih operacija.

5.2 Osnovni kombinatorni rezultati

Najpoznatiji su kombinatorni principi prebrojavanja:

Princip jednakosti

Neka su S i T konačni skupovi. Ako postoji bijekcija među njima, tada su skupovi S i T jednakobrojni, tj. $k(S) = k(T)$, gdje su $k(S)$ i $k(T)$ oznake za kardinalne brojeve skupova S i T, redom.

Princip sume

Neka su $n \in \mathbb{N}$ i $S_1, \ldots, S_n$ konačni skupovi takvi da je $S_i \cap S_j = \emptyset$, $i \neq j$ (dakle, disjunktni su). Tada je $(S_1 \cup S_2 \cup \ldots \cup S_n)$ konačan skup i vrijedi:

$$k(S_1 \cup S_2 \cup \ldots \cup S_n) = k(S_1) + k(S_2) + \ldots + k(S_n).$$

Princip produkta

Neka su $n \in \mathbb{N}$ i $S_1, \ldots, S_n$ konačni skupovi (ne nužno disjunktni). Tada je njihov Kartezijev produkt konačan skup i vrijedi:

$$k(S_1 \times S_2 \times \ldots \times S_n) = k(S_1) \cdot k(S_2) \cdot \ldots \cdot k(S_n).$$

Teorem 5.1 (Teorem o uzastopnom prebrojavanju)**.** *Neka su $A_1, \ldots, A_n$ konačni skupovi i neka je $T \subset (A_1 \times \ldots \times A_n)$ skup uređenih n-torki $(a_1, \ldots, a_n)$ definiranih na sljedeći način: prva komponenta a_1 može se birati na k_1 načina (dakle, među k_1 različitih elemenata skupa A_1); za svaku već izabranu prvu komponentu drugu komponentu a_2 možemo birati na k_2 različitih načina itd. Za svaki izbor komponenata $a_1, a_2, \ldots a_{n-1}$ n-tu komponentu a_n možemo odabrati na k_n različitih načina. Tada je kardinalni broj skupa T jednak:*

$$k(T) = k_1 \cdot \ldots \cdot k_n.$$

Primjer 5.1. *Ako želimo odabrati jedan par, mladića i djevojku, iz razreda koji se sastoji od 21 djevojke i 2 mladića, broj načina na koji to možemo učiniti jednak je broju elemenata kartezijevog produkta skupa koji se sastoji od 21 elemenata i skupa koji se sastoji od 2 elementa. Dakle, prema principu produkta, to možemo učiniti na 42 načina.*

Uređeni razmještaji nazivaju se **permutacije**, a neuređeni razmještaji **kombinacije**.

Definicija 5.1. *Neka je* $A = \{a_1, \dots a_n\}$ *skup koji se sastoji od* n *elemenata i neka je* $r \in \mathbb{N}$, $r \leq n$. **Varijacija** r**-tog razreda** *u skupu* A *svaka je uređena* r*-torka međusobno različitih elemenata iz skupa* A*. Broj varijacija* r*-tog razreda* n*-članog skupa je*

$$V_n^r = n \cdot (n-1) \cdot \ldots \cdot (n-r+1) = \frac{n!}{(n-r)!}.$$

Primjer 5.2. *Neka je* A *skup koji se sastoji od deset elemenata. Svaka uređena trojka elemenata skupa* A *jedna je varijacija trećeg razreda skupa* A*. Takvih varijacija ima*

$$V_{10}^3 = \frac{10!}{(10-3)!} = 720.$$

Definicija 5.2. *Svaku uređenu* n*-torku skupa od* n *elemenata zovemo* **permutacija***. Broj je permutacija* n*-članog skupa*

$$p_n = V_n^n = n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1 = n!.$$

Napomena 5.1. *Permutacija je u* n*-članom skupu svaka varijacija* n*-tog razreda tog skupa, odnosno permutacija* n*-članog skupa svaka je bijekcija tog skupa na samog sebe.*

Primjer 5.3. *Ako se pitamo na koliko načina pet ljudi može stati u red, zapravo se pitamo koliko permutacija ima peteročlani skup. Broj je permutacija peteročlanog skupa* $V_5^5 = 5! = 120$.

Definicija 5.3. *Neka je* $A = \{a_1, \dots a_n\}$ *skup koji se sastoji od* n *elemenata i neka je* $r \in \mathbb{N}$, $r \leq n$. **Kombinacija** r**-tog razreda** *u skupu* A *svaki je* r*-člani podskup skupa* A*. Broj je kombinacija* r*-tog razreda skupa od* n *elemenata*

$$C_n^r = \binom{n}{r} = \frac{n!}{r! \cdot (n-r)!}.$$

Primjer 5.4. *U razredu ima* 15 *dječaka i* 10 *djevojčica. Na primjer:*

> *Tri dječaka možemo odabrati na onoliko načina koliko ima različitih kombinacija trećeg razreda skupa od* 15 *elemenata, dakle na*
>
> $$C_{15}^3 = \binom{15}{3}$$
>
> *načina.*
>
> *Prema principu produkta slijedi da tri dječaka i dvije djevojčice možemo odabrati na*
>
> $$C_{15}^3 \cdot C_{10}^2 = \binom{15}{3} \cdot \binom{10}{2}$$
>
> *načina.*
>
> *Jednak broj dječaka i djevojčica možemo odabrati na*
>
> $$\sum_{k=1}^{10} C_{15}^k \cdot C_{10}^k = \sum_{k=1}^{10} \binom{15}{k} \cdot \binom{10}{k}$$
>
> *načina.*

Definicija 5.4. *Neka je* $A = \{a_1, \dots a_n\}$ *skup koji se sastoji od* n *elemenata.* **Varijacija s ponavljanjem** r**-tog razreda skupa od** n**-elemenata** *svaka je uređena* r*-torka elemenata iz skupa* A*. Broj je takvih varijacija s ponavljenjem* n^r.

Primjer 5.5. *Ako se pitamo koliko ima binarnih nizova (nizova čiji su elementi samo nule i jedinice), zapravo se pitamo koliko ima različitih varijacija sedmog razreda s ponavljanjem dvočlanog skupa. Takvih varijacija s ponavljanjem ima* $2^7 = 128$.

Definicija 5.5. *Neka je* $A = \{a_1, \dots a_n\}$ *skup koji se sastoji od* n *elemenata.* **Permutacija s ponavljanjem r-tog razreda** *svaka je uređena* r*-torka elemenata iz skupa* A *među kojima je* n_1 *elemenata međusobno jednakih,* n_2 *elemenata međusobno jednakih,* $\dots$, n_k *elemenata međusobno jednakih, pri čemu je* $n_1 + n_2 + \dots + n_k = r$. *Broj je takvih permutacija s ponavljenjem*

$$P_r^{n_1, n_2, \dots, n_k} = \frac{r!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}.$$

Primjer 5.6. *Ako nas zanima koliko se različitih riječi, smislenih i besmislenih, može napisati od slova riječi "matematika", tada nas zapravo zanima broj permutacija s ponavljanjem desetog razreda, pri čemu imamo tri slova A, dva slova M i dva slova T. Takvih permutacija s ponavljanjem ima*

$$P_{10}^{3,2,2} = \frac{10!}{3! \cdot 2! \cdot 2!} = 151200.$$

Dakle, od slova riječi "matematika" može se napisati ukupno 151200 *različitih riječi.*

5.3 Ponovljeni red

U diskretnoj je teoriji vjerojatnosti česta potreba za računanjem sume tzv. ponovljenog reda[1]. Da bismo pojasnili razliku između ponovljenog reda i običnog reda, podsjetimo se da je red uređeni par dvaju nizova $((a_n), (s_n))$ od kojih drugi niz nastaje sumiranjem prvih n članova prvog niza, tj.

$$s_n = a_1 + a_2 + \dots a_n.$$

Pritom kažemo da red konvergira ako konvergira niz parcijalnih suma (s_n) [13].
Ponovljeni red nastaje tako da prvo napravimo jednu particiju[2] skupa prirodnih brojeva $(M_i, i \in \mathbb{N})$ pa posebno zbrojimo one članove niza čiji indeksi pripadaju u M_i, a zatim sve tako dobivene sume, ako takve sume postoje.

Primjer 5.7. *Neka je zadan geometrijski red s kvocijentom* $\frac{1}{2}$ *i prvim članom* $\frac{1}{2}$, *tj.*

$$\sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^n.$$

Taj red konvergentan je i suma mu je 1. *Isti rezultat dobit ćemo ako podijelimo skup indeksa* $\mathbb{N}$ *na parne i neparne brojeve, tj. ako promatramo particiju* $\{N, P\}$ *skupa prirodnih brojeva, pri čemu je*

$$N = \{2n - 1 : n \in \mathbb{N}\}, \quad P = \{2n : n \in \mathbb{N}\}.$$

[1] Pod pojmom "ponovljeni red" ovdje porazumijevamo samo dvostruki red.

[2] Familija skupova $\{M_i, i \in \mathbb{N}\}$ čini particiju skupa A ako su zadovoljena sljedeća svojstva:

1. $M_i \subseteq A$, $\forall i \in \mathbb{N}$,
2. $M_i \cap M_j = \emptyset$ za sve $i \neq j$,
3. $\bigcup_{i \in \mathbb{N}} M_i = A$.

Budući da je $\mathbb{N} = P \cup N$, sumu spomenutog geometrijskog reda možemo izračunati tako da prvo zbrojimo sve članove niza s parnim eksponentima, zatim sve članove niza s neparnim eksponentima, a zatim te dvije sume:

$$\sum_{n\in N} \frac{1}{2^n} = \sum_{n=1}^{\infty} \frac{1}{2^{2n-1}} = \frac{\frac{1}{2}}{1-\frac{1}{4}} = \frac{2}{3},$$

$$\sum_{n\in P} \frac{1}{2^n} = \sum_{n=1}^{\infty} \frac{1}{2^{2n}} = \frac{\frac{1}{4}}{1-\frac{1}{4}} = \frac{1}{3}.$$

Očito vrijedi:

$$\sum_{n\in\mathbb{N}} \frac{1}{2^n} = \sum_{n\in N} \frac{1}{2^n} + \sum_{n\in P} \frac{1}{2^n} = \frac{2}{3} + \frac{1}{3} = 1.$$

Za zadavanje ponovljenog reda treba nam niz (a_n) i particija skupa $\mathbb{N}$: $\{M_i, i \in \mathbb{N}\}$. Ako za svaki $i \in \mathbb{N}$ konvergiraju redovi $\sum\limits_{j\in M_i} a_j$ i sume im označimo

$$A_i = \sum_{j\in M_i} a_j,$$

tada definiramo ponovljeni red kao

$$\sum_{i\in\mathbb{N}} \sum_{j\in M_i} a_j = \sum_{i\in\mathbb{N}} A_i.$$

Primjer 5.8. *Neka je zadana beskonačna matrica realnih brojava:*

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & \cdots \\ a_{21} & a_{22} & a_{23} & a_{24} & \cdots \\ a_{31} & a_{32} & a_{33} & a_{34} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}.$$

Takve matrice često koristimo za zadavanje ponovljenih redova s obzirom da se particije skupa $\mathbb{N}$ često mogu lakše prikazati korištenjem dvaju indeksa. Iz beskonačne matrice možemo definirati mnogo ponovljenih redova. Navedimo nekoliko primjera:

- *Sumirajmo prvo članove matrice po redovima:*

$$A_i = \sum_{j=1}^{\infty} a_{ij}.$$

  Ako te sume postoje, tj. ako je $A_i < \infty$, $\forall i \in \mathbb{N}$, možemo definirati ponovljeni red

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij}.$$

- *Sumirajmo prvo članove matrice po stupcima:*

$$A_j = \sum_{i=1}^{\infty} a_{ij}.$$

  Ako te sume postoje, tj. ako je $A_j < \infty$, $\forall j \in \mathbb{N}$, možemo definirati ponovljeni red

$$\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij}.$$

- *Osim zbrajanja po stupcima ili recima, možemo particiju praviti i na druge načine, npr. po pavilima koji su slikovito prikazani u sljedećoj matrici:*

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & \dots \\ a_{21} & a_{22} & a_{23} & a_{24} & \dots \\ a_{31} & a_{32} & a_{33} & a_{34} & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}.$$

S obzirom da se iz istog niza (a_n) brojeva može definirati mnogo različitih ponovljenih redova, pitanje je koliko se toga u konačnici "preslagivanjem" mijenja. Npr. ako jedan ponovljeni red konvergira, može li se "preslagivanjem" dogoditi da novonastali ponovljeni red ne konvergira, može li se uopće definirati ponovljeni red za bilo koju particiju skupa indeksa, itd. Da tako postavljena pitanja imaju smisla, potvrđuje sljedeći primjer:

Primjer 5.9. *Neka je zadana beskonačna realna matrica*

$$A = \begin{bmatrix} 0 & -1 & 1 & -1 & 1 & -1 \dots \\ 1 & 0 & -1 & 1 & -1 & 1 \dots \\ -1 & 1 & 0 & -1 & 1 & -1 \dots \\ 1 & -1 & 1 & 0 & -1 & 1 \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

Svi redovi koji nastaju sumiranjem elemenata iz jednog retka te matrice divergentni su, tj. za svaki $i \in \mathbb{N}$ divergentni su redovi

$$\sum_j a_{ij}.$$

Dakle, ponovljeni red ne može se definirati tako da se prvo sumira po redovima zadane beskonačne matrice. Slično vrijedi i ako se prvo sumira po stupcima. Međutim, ako particiju pravimo po bilo kojoj od dviju shema označenih u matricama:

$$A = \begin{bmatrix} 0 & -1 & 1 & -1 & 1 & -1 & \dots \\ 1 & 0 & -1 & 1 & -1 & 1 & \dots \\ -1 & 1 & 0 & -1 & 1 & -1 & \dots \\ 1 & -1 & 1 & 0 & -1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & & & \end{bmatrix} \qquad A = \begin{bmatrix} 0 & -1 & 1 & -1 & 1 & -1 & \dots \\ 1 & 0 & -1 & 1 & -1 & 1 & \dots \\ -1 & 1 & 0 & -1 & 1 & -1 & \dots \\ 1 & -1 & 1 & 0 & -1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & & & \end{bmatrix}$$

vidimo da sume uvijek postoje. Dakle, ponovljene redove po tim particijama možemo definirati.

Sljedeći teorem daje nužne i dovoljne uvjete koji jamče da se "preslagivanjem" neće promijeniti suma ponovljenog reda.

Teorem 5.2. *Neka je s $(a_n, n \in \mathbb{N})$ dan niz realnih brojeva i neka je $\{M_i, i \in \mathbb{N}\}$ jedna particija skupa prirodnih brojeva. Red*

$$\sum_{n\in\mathbb{N}} |a_n|$$

konvergira onda i samo onda ako konvergira red

$$\sum_{i\in\mathbb{N}}\sum_{j\in M_i}|a_j|.$$

U slučaju konvegencije sume su im iste.

Dokaz. Pretpostavimo da red

$$\sum_{i\in\mathbb{N}}\sum_{j\in M_i}|a_j|$$

konvergira i označimo njegovu sumu L. Red $\sum\limits_{n\in\mathbb{N}}|a_n|$ red je s pozitivnim članovima i očigledno je njegov niz parcijalnih suma ograničen realnim brojem L. Dakle, i taj red konvergentan je (pogledati npr. [13]). Nadalje, s obzirom da je apsolutno konvergentan red realnih brojeva ujedno i konvergentan (pogledati npr. [13]), slijedi da je i red $\sum\limits_{n\in\mathbb{N}}a_n$ konvergentan.
Pretpostavimo sada da konvergira red

$$\sum_{n\in\mathbb{N}}|a_n|$$

i njegovu sumu označimo s L. Tada za svaki element M_i particije skupa prirodnih brojeva pripadni red

$$\sum_{j\in M_i}|a_j|$$

predstavlja red realnih brojeva s pozitivnim članovima čiji je niz parcijalnih suma ograničen. Dakle, za svaki je $i\in\mathbb{N}$ takav red konvergentan. Označimo:

$$\sum_{j\in M_i}|a_j| = A_i, \quad i\in\mathbb{N}.$$

Dokažimo da i red

$$\sum_{i\in\mathbb{N}}A_i$$

konvergira te da mu je suma L. U tu svrhu označimo $(b_{1i}, i\in\mathbb{N})$ podniz niza $(|a_i|, i\in\mathbb{N})$ koji čine oni elementi s indeksima iz M_1. Analogno, za svaki $k\in\mathbb{N}$ definirajmo odgovarajući podniz $(b_{ki}, i\in\mathbb{N})$ niza $(|a_i|, i\in\mathbb{N})$ koji čine oni elementi s indeksima iz M_k. Uočimo da je tako nastali skup podnizova od $(|a_i|, i\in\mathbb{N})$ pokupio sve njegove članove te da se niti jedan član ne pojavljuje više od jednom. S obzirom da se radi o pozitivnim brojevima, posljedica je toga da za svaki $n\in\mathbb{N}$ vrijedi:

$$\begin{aligned}A_1 + A_2 + \cdots A_n &= \lim_{l\to\infty}\sum_{i=1}^{l}b_{1i} + \lim_{l\to\infty}\sum_{i=1}^{l}b_{2i} + \cdots + \lim_{l\to\infty}\sum_{i=1}^{l}b_{ni}\\ &= \lim_{n\to\infty}\left[\sum_{i=1}^{l}b_{1i} + \cdots + \sum_{i=1}^{l}b_{ni}\right] < L.\end{aligned}$$

Dakle, red s pozitivnim članovima ograničen je, pa je time i konvergentan. Označimo njegovu sumu V. Očigledno je da je i $V\leq L$ s obzirom da je L gornja međa pripadnog niza parcijalnih suma. Pokažimo da je istovremeno i $V\geq L$. Naime, za svaku k-tu parcijalnu sumu reda $\sum\limits_{i=1}^{k}|a_i|$ možemo naći dovoljno veliki n tako da je

$$\sum_{i=1}^{n}A_i \geq \sum_{i=1}^{k}|a_i|.$$

Pripadne će granične vrijednosti onda također zadovoljavati istu nejednakost, što znači da je $V\geq L$. Dakle, mora biti $V = L$.

Teorem 5.3. *Neka je s $(a_n, n \in \mathbb{N})$ dan niz realnih brojeva i neka je $\{M_i, i \in \mathbb{N}\}$ jedna particija skupa prirodnih brojeva. Ako jedan od redova*

$$\sum_{n\in\mathbb{N}} |a_n|, \quad \sum_{i\in\mathbb{N}} \sum_{j\in M_i} |a_j|$$

konvergira, onda konvergiraju i redovi

$$\sum_{n\in\mathbb{N}} a_n, \quad \sum_{i\in\mathbb{N}} \sum_{j\in M_i} a_j$$

i sume su im iste.

<u>**Dokaz.**</u> Koristeći prethodni teorem i činjenicu da je apsolutno konvergentan red realnih brojeva ujedno i konvergentan, da bismo dokazali tvrdnju teorema dovoljno je dokazati da konvergencija reda $\sum\limits_{n\in\mathbb{N}} |a_n|$ povlači konvergenciju reda $\sum\limits_{i\in\mathbb{N}} \sum\limits_{j\in M_i} a_j$ te da je tada

$$\sum_{n\in\mathbb{N}} a_n = \sum_{i\in\mathbb{N}} \sum_{j\in M_i} a_j.$$

U tu svrhu uočimo da iz pretpostavke o konvergenciji reda $\sum\limits_{n\in\mathbb{N}} |a_n|$ možemo zaključiti:

- $\sum\limits_{n\in\mathbb{N}} a_n$ konvergentan je. Označimo:

$$V = \sum_{n\in\mathbb{N}} a_n.$$

- $\sum\limits_{j\in M_i} |a_j|$ konvergentan je za svaki $i \in \mathbb{N}$ pa je time i $\sum\limits_{j\in M_i} a_j$ konvergentan. Označimo:

$$\sum_{j\in M_i} a_j = A_i.$$

Slično kao u dokazu prethodnog teorema, za svaki $i \in \mathbb{N}$ definirajmo odgovarajući podniz $(b_{ik}, k \in \mathbb{N})$ niza $(a_j, j \in \mathbb{N})$ koji čine elementi s indeksima iz M_i, tj.

$$\sum_{k\in\mathbb{N}} b_{ik} = A_i.$$

Uočimo da je tako nastali skup podnizova od $(a_j, j \in \mathbb{N})$ pokupio sve njegove članove te da se niti jedan član ne pojavljuje više od jednom.
Iz apsolutne konvergencije reda $\sum\limits_{n\in\mathbb{N}} a_n$ znamo da za svaki $\varepsilon > 0$ postoji $k_0 \in \mathbb{N}$ takav da je

$$\sum_{j=k_0+1}^{\infty} |a_j| < \varepsilon, \tag{5.1}$$

odakle slijedi da je

$$\left| V - \sum_{j=1}^{k_0} a_j \right| = \left| \sum_{j=k_0+1}^{\infty} a_j \right| \leq \sum_{j=k_0+1}^{\infty} |a_j| < \varepsilon.$$

Neka su n_0 i l_0 dovoljno veliki prirodni brojevi da izraz

$$\sum_{i=1}^{n_0} \sum_{k=1}^{l_0} b_{ik} - \sum_{j=1}^{k_0} a_j$$

predstavlja sumu članova reda $\sum_{j\in\mathbb{N}} a_j$ s indeksima većim od k_0. Tada za svaki $n \geq n_0$ i $l \geq l_0$ vrijedi:

$$\left|\sum_{i=1}^{n}\sum_{k=1}^{l} b_{ik} - \sum_{j=1}^{k_0} a_j\right| < \varepsilon,$$

pa prijelazom na limes po $l \to \infty$ vidimo da je

$$\left|\sum_{i=1}^{n} A_i - \sum_{j=1}^{k_0} a_j\right| < \varepsilon.$$

Koristeći prethodnu nejednakost i nejednakost (5.1) vidimo da je

$$\left|\sum_{i=1}^{n} A_i - V\right| < 2\varepsilon, \quad \forall n > n_0,$$

Dakle, red $\sum_{i\in\mathbb{N}}\sum_{j\in M_i} a_j$ konvergira i suma mu je V.

Literatura

[1] Anderson, T.W. *An introduction to Multivariante Statistical Analysis*, J. Wiley, 1958.

[2] Bain, L.E, Engelhardt, M. *Introduction to Probability and Mathematical statistics*, Duxbury, 2009.

[3] Cohn, D.L. *Measure Theory*, Birkhäuser, Boston-Basel-Stuttgart, 1980.

[4] Chow, Y.S, Teicher, H. *Probability Theory*, Spronger, New York-Heidelberg-Berlin, 1978.

[5] Daniel, W.W., Terrell, J.C. *Business Statistics*, Houghton Mifflin Company, Boston, 1989.

[6] Durrett, R. *Probability: Theory and Examples*, Duxbury Press, 1995.

[7] Elezović, N. *Diskretna vjerojatnost*, Element, Zagreb, 2007.

[8] Elezović, N. *Slučajne varijable*, Element, Zagreb, 2007.

[9] Elezović, N. *Statistika i procesi*, Element, Zagreb, 2007.

[10] Glišič, Z., Peruničić, P. *Zbirka rešenih zadataka iz verovatnoće i matematičke statistike*, Naučna knjiga, Beograd, 1982.

[11] Ilijašević, M., Pauše, Ž. *Riješeni primjeri i zadaci iz vjerojatnosti i statistike*, "Zagreb", Samobor, 1990.

[12] Ivanović, B. *Teorijaska statistika* Jugoslavenski institut za ekonomska istraživanja, Beograd, 1966.

[13] Jukić, D., Scitovski, R. *Matematika I*, Elektrotehnički fakultet, Odjel za matematiku, Prehrambeno tehnološki fakultet, Osijek, 2000.

[14] Jukić, D *Uvod u teoriju mjere i integracije - prvi dio*, Odjel za matematiku, Osijek, 2008.

[15] Ivković, Z.A. *Matematička statistika*, Naučna knjiga, Beograd, 1976.

[16] Jamnik, R. *Matematična statistika*, Državna založba Slovenije, Ljubljana, 1980.

[17] Javor, P. *Uvod u matematičku analizu*, Školska knjiga, Zagreb, 1988.

[18] Lehmann, E.L. *Testing Statistical Hypotheses*, J. Wiley, 1959.

[19] Lozanov-Crvenković, Z., Rajter, D. *Zbirka rešenih zadataka iz verovatnoće i statistike*, Prirodno-matematički fakultet, Novi sad, 1999.

[20] Mališić, J. *Zbirka zadataka iz teorije verovatnoće sa primenama*, Građevinska knjiga, Beograd, 1990.

[21] Mittelhammer, R.C. *Mathematical Statistics for Economics and Bussines*, Springer, New York, 1996.

[22] Pauše, Ž. *Uvod u matematičku statistiku*, Školska knjiga, Zagreb, 1993.

[23] Pauše, Ž. *Vjerojatnost, informacija, stohastički procesi*, Školska knjiga, Zagreb, 1988.

[24] Pavlić, I. *Statistička teorija i primjena*, Tehnička knjiga, Zagreb, 1985.

[25] Pogány, T. *Teorija vjerojatnosti, zbirka riješenih ispitnih zadataka*, Odjel za pomorstvo Sveučilišta u Rijeci, Rijeka, 1999.

[26] Sarapa, N. *Teorija vjerojatnosti*, Školska knjiga, Zagreb, 1988.

[27] Sarapa, N. *Vjerojatnost i statistika I. dio: osnove vjerojatnosti - kombinatorika*, Školska knjiga, Zagreb, 1995.

[28] Sarapa, N. *Vjerojatnost i statistika II. dio: osnove statistike - slučajne varijable*, Školska knjiga, Zagreb, 1996.

[29] Seber G.A.F, Lee A.J. *Linear Regression Analysis*, Wiley, Hoboken-New Jersey, 2003.

[30] Triola, M.F. *Elementary Statistics*, The Benjamin/Cummings Publishing company, Inc. 1989.

[31] Vranić, V. *Vjerojatnost i statistika*, Tehnička knjiga, Zagreb, 1971.

[32] Vranjković, P. *Zbirka zadataka iz vjerojatnosti i statistike*, Školska knjiga, Zagreb, 1990.

Indeks