Vera Čuljak

VJEROJATNOST I STATISTIKA

Građevinski fakultet
Sveučilište u Zagrebu

Predgovor

Poštovani čitatelji,

nadam se da ćete naći korisne informacije u ovom nastavnom tekstu.

Ruski matematičar P.L. Čebišev (1821-1894) koji je unio nove ideje u teoriju vjerojatnosti izrekao je tezu da je povijest matematike moguće podijeliti na tri perioda:
$*$ prvi: kada su zadatke zadavali grčki bogovi npr. delski problemi: nerješivi antički geometrijski zadatak da se pomoću šestara i ravnala konstruira kocka s obujmom dvostruko većim od zadanog; problem kvadrature kruga, problem trisekcije kuta;
$*$ drugi: kada su zadatke zadavali polubogovi npr. B. Pacal (1623-1662) i P. Fermat (1601-1665);
$*$ treći: kada zadatke zadaje praksa.

Veliki interes za teoriju vjerojatnosti i matematičku statistiku počeo u drugoj polovici 20. stoljeća potaknut eksperimentalnim istraživanjima u industriji, ekonomiji, psihologiji, biologiji, fizici.

Nova oblast teorije vjerojatnosti je teorija stohastičkih procesa koju je zasnovao A. Kolmogorov (1903 -1987). Stohastički procesi su terorijski modeli za npr. dinamičke sisteme u slučajevima kad se stanja sistema određuju s vjerojatnostima. Zahvaljujući razvoju teorije vjerojatnosti razvile su se i znanstvene discipline: teorija masovnog opsluživanja, teorija informacija, teorija pouzdanosti tehničkih sistema, teorija zaliha.

Matematička statistika se razvija na temeljima teorije vjerojatnosti u smjeru teorije procjena, teorije provjera statističkih hipoteza i teorije planiranja eksperimenata.

Možda u ovim uvodnim rečenicama čitatelj nađe motiv za pregled ovog matematičkog sadržaja.

Ovaj nastavni materijal (skripta) je nastao kao radni materijal za predavanja iz kolegija *Vjerojatnost i statistika* koji predajem od 2005. godine na preddiplomskom studiju Građvinskog fakulteta Sveučilišta u Zagrebu.

Cilj kolegija je da se studenti upoznaju s osnovnim elmentima matematičke statistike kako bi u praksi mogli obraditi i interpertirati statističke podatke koje su dobili mjerenjem, testom ili anketom i kako bi uspješno primijenili elemente statističkog zaključivanja u stručnim kolegijima npr. *Poznavanje materijala, Hidrologija, Pouzdanost konstrukcija i dr.*

Teorijski temelj matematičke statistike je teorija vjerojatnosti. To znanje je podloga i za kolegij *Stohastički procesi* koji se predaje na diplomskom studiju Građevinskog fakulteta.

Metodički razlozi uvjetovali su organizaciju skripte u tri cjeline: Vjerojatnost, Slučajne varijable i Matematička statistika. Skripta, preko 15 poglavlja, prati 15 predavanja (po 2 sata tjedno).
Strogi matematički sadržaj ublažen je imenovanjem definicija i teorema, izborom tema *tko želi znati više* i shematiziranom organizacijom svakog poglavlja:

motivacijski primjer + definicija + primjer + teorem + primjer +
napomene + "šalabahter " za ponavljanje gradiva.

Zbog opsežnosti gradiva broj primjera i njihova težina primjereni su predavanjima.

Pokušala sam pripremiti skriptu u kojoj će studenti naći :
* osnovne informacije i tehnike rješavanja pojedinih tipova zadataka - "kuharice";
* svoju bilježnicu, tako da dopisuju komentare na isprintani primjerak poglavlja;
* materijal u kojem će, ako *žele znati više*, naći odgovore na pitanje ZAŠTO:
* zašto je lozinka varijacija bez ili s ponavljanjem;
* zašto kontolor kvalitete proizvoda u skladištu mora znati hipergeometrijsku razdiobu;
* zašto je važan Gaussov šešir za odredjivanje očekivane mase opeke ili određivanje maksimalnog opterećenja da je bi očekivani broj slomljenih ploča bio najviše 5%;
* zašto su deformacija i Brinellova tvrdoća za neku vrstu čelika jako korelirane;
* zašto učimo o slučajnim varijablama da bi odredili interval unutar kojeg

će se kretati postotak glasova za nekog kandidata na izborima ili testirali hipotezu da dva tipa betona imaju jednake tlačne čvrstoće uz razinu značajnosti 1%.

Ovo je radni materijal i bit ću zahvalna ako sve uočene greške i prijedloge kolegice i kolege studenti i čitatelji pošalju na e-mail adresu vera@grad.hr.

Dodatni nastavni materijali na web stranici predmeta koji prate ovu web skriptu odnose se na primjenu programskog paketa Mathematika u statistici. Baza primjera kolokvija s teorijskim pitanjima obogaćuje se svake godine. Kolgice dr. sc. Tajana Slijepčević-Manger, Martina Benković i kolege Boško Kojundžić i Nikola Sandrić pripremali su zadatke za kolokvije i auditorne vježbe.

Zahvaljujem se kolegici Martini Benković na pomoći pri grafičkoj obradi teksta i izradi slika.
Zahvaljujem se kolegi Vladimiru Beniću koji je napravio web format skripte.

Zagreb, 1.7.2011. Vera Čuljak

Sadržaj

Dio I

Vjerojatnost

Poglavlje 1

ELEMENTI KOMBINATORIKE

KOMBINATORIKA je grana grana matematike koja se bavi osnovnim svojstvima konačnih skupova i metodama prebrojavanja. Poznati kombinatorni problemi su: problem izbora upravnog odbora, problem mostova u Konigsbergu, problem magičnih kvadrata, problem 4 boje. Zadatak ovog poglavlja je da se upozanju metode prebrojavanja koje će se iskoristiti za računanje vjerojatnosti slučajnih događaja.

1.1 UVOD

MOTIV 1.1 *Izračunajte koeficijent uz* x^6 *u razvoju* $(1 + x^3)^{14}$.

Definicija 1.1 *(FAKTORIJEL)*

Za prirodan broj $n \in \mathbb{N}$*, definiramo prirodan broj "n faktorijela" u oznaci* $n!$ *sa:*

$$n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n,$$

a dogovorno $0! = 1$.

(BINOMNI KOEFICIJENT)

Neka su $n, k \in \mathbb{N},\ k \leq n$*. Binomni koeficijent je prirodni broj u oznaci* $\binom{n}{k}$ *("n povrh k") definiran na sljedeći način:*

$$\binom{n}{k} = \frac{n(n-1) \cdot \ldots \cdot (n-k+1)}{1 \cdot 2 \cdot \ldots \cdot k} = \frac{n!}{k!(n-k)!},$$

a dogovorno: $\binom{n}{0} = 1$, $\binom{0}{0} = 1$.

PRIMJER 1.1 ***T: Svojstva binomnih koeficijenata:***

(i) $\binom{n}{k} = \binom{n}{n-k}$.

(ii) $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$.

Rješenje:
D: *tko želi znati više*

(i) $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n!}{(n-k)!k!} = \frac{n!}{(n-k)!(n-(n-k))!} = \binom{n}{n-k}$.

(ii)

$$\begin{aligned}
\binom{n}{k} + \binom{n}{k+1} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k+1)!(n-k-1)!} \\
&= \frac{n!}{k!(n-k-1)!(n-k)} + \frac{n!}{k!(k+1)(n-k-1)!} \\
&= \frac{n!(k+1) + n!(n-k)}{k!(k+1)(n-k-1)!(n-k)} \\
&= \frac{n!(n+1)}{(k+1)!(n-k)!} \\
&= \frac{(n+1)!}{(k+1)!((n+1)-(k+1))!} \\
&= \binom{n+1}{k+1}.
\end{aligned}$$

NAPOMENA 1.1 *Stirlingova formula*

Za približno izračunavanje faktorijela velikih brojeva možemo primijeniti aproksimativnu Stirlingovu formulu:

$$n! \approx n^n \cdot e^{-n} \sqrt{2\pi n}.$$

PRIMJER 1.2 *Izračunajte* 5! *pomoću Stirlingove formule. Kolika je greška aproksimacije?*

Rješenje:
Pomoću Stirlingove formule računamo $5! \approx 5^5 \cdot e^{-5}\sqrt{2\pi 5} = 118.019$. Kako je po definiciji $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$ greška aproksimacije je 1.981.

Peti Peanov aksiom prirodnih brojeva je aksiom matematičke indukcije koji ćemo koristiti u dokazima formula za broj permutacija, varijacija i kombinacija.

TEOREM 1.1 *AKSIOM MATEMATIČKE INDUKCIJE (AMI)*
Neka je $\mathbb{N}$ skup prirodnih brojeva. Ako je $M \subseteq \mathbb{N}$, i vrijedi:

(i) $1 \in M$,

(ii) $(\forall m \in \mathbb{N}),\ m \in M \Rightarrow m + 1 \in M$

onda je $M = \mathbb{N}$.

AMI primijenjujemo pri dokazivanju općenite valjanosti formula koje sadrže varijablu $n \in \mathbb{N}$.

TEOREM 1.2 *(BINOMNI TEOREM)*
Neka su $a, b \in \mathbb{R}$, $n \in \mathbb{N}$. Binomni teorem izražava razvoj n-te potencije binoma formulom:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}, \qquad \forall n \in \mathbb{N}$$

.

Dokaz: *tko želi znati više*
Tvrdnja se dokazuje pomoću aksioma matematičke indukcije (AMI). Neka je

$$M = \left\{ n \in \mathbb{N} : (a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \right\}.$$

Treba dokazati da je $M = \mathbb{N}$.
(B.I) $1 \in M$:

$$\begin{aligned} (a+b)^1 &= \sum_{k=0}^{1} \binom{1}{k} a^1 b^{1-k} \\ a + b &= b + a. \end{aligned}$$

(P.I.) $m \in M$:

$$(a+b)^m = \sum_{k=0}^{m} \binom{m}{k} a^k b^{m-k}$$

(K.I.) Treba pokazati koristeći (P.I.) da je $m+1 \in M$ tj.

$$(a+b)^{m+1} = \sum_{k=0}^{m+1} \binom{m+1}{k} a^k b^{m+1-k}.$$

Računamo

$$\begin{aligned}
(a+b)^{m+1} &= (a+b)^m(a+b) \\
&= P.I. \\
&= \sum_{k=0}^{m} \binom{m}{k} a^k b^{m-k} * (a+b) \\
&= \sum_{k=0}^{m} \binom{m}{k} a^{k+1} b^{m-k} + \sum_{k=0}^{m} \binom{m}{k} a^k b^{m+1-k}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{k=0}^{m-1} \binom{m}{k} a^{k+1} b^{m-k} + \binom{m}{m} a^{m+1} b^0 + \binom{m}{0} a^0 b^{m+1} + \sum_{k=1}^{m} \binom{m}{k} a^k b^{m+1-k} \\
&= \sum_{k=1}^{m} \binom{m}{k-1} a^k b^{m+1-k} + \binom{m}{m} a^{m+1} b^0 \binom{m}{0} a^0 b^{m+1} + \sum_{k=1}^{m} \binom{m}{k} a^k b^{m+1-k} \\
&= \binom{m}{0} a^0 b^{m+1} + \sum_{k=1}^{m} [\binom{m}{k-1} + \binom{m}{k}] a^k b^{m+1-k} + \binom{m}{m} a^{m+1} b^0 \\
&= \binom{m+1}{0} a^0 b^{m+1} + \sum_{k=1}^{m} [\binom{m}{k-1} + \binom{m}{k}] a^k b^{m+1-k} + \binom{m+1}{m+1} a^{m+1} b^0 \\
&= \binom{m+1}{0} a^0 b^{m+1} + \sum_{k=1}^{m} \binom{m+1}{k} a^k b^{m+1-k} + \binom{m+1}{m+1} a^{m+1} b^0 \\
&= \sum_{k=0}^{m+1} \binom{m+1}{k} a^k b^{m+1-k}.
\end{aligned}$$

Zaključujemo prema AMI da je $M = \mathbb{N}$.

Prisjetimo se motivacijskog primjera:

PRIMJER 1.3 *motiv*

Izračunajte koeficijent uz x^6 *u razvoju* $(1+x^3)^{14}$.

Rješenje:
Prema binomnom teoremu računamo $(1+x^3)^{14} = \sum_{k=0}^{n}\binom{n}{k}1^k(x^3)^{n-k}$. U razvoju ćemo odrediti k iz uvjeta da je $(x^3)^{n-k} = x^6$. Za $n = 14$ dobili smo da je $k = 12$. Dvanaesti član u razvoju je x^6, pa je njegov koeficijent $\binom{n}{k} = \binom{14}{12} = 91$. Traženi koeficijent uz potenciju x^6 je 91.

1.2 PRINCIPI PREBROJAVANJA

MOTIV 1.2 *U gradu ima 8 studentskih restorana ravnomjerno raspoređenih u 4 gradske četvrti. U okolini svakog restorana nalaze se dvije sportske dvorane. Student želi unajmiti stan. Na koliko načina može odabrati četvrt, studentski restoran i sportsku dvoranu ako: a) nije bitno ni da studentski restoran bude u istoj četvrti niti dvorana; b) nije bitno da studentski restoran bude u istoj četvrti ali dvorana treba biti; c) sve bude u najbližoj okolini.*

TEOREM 1.3 *(PRINCIP SUME)*

Neka konačni skupovi imaju n_i elemenata, $n_i \in \mathbb{N}$, $i = 1, 2, .., k$, $|S_i| = n_i$ i neka su disjunktni za svaki izbor $i \neq j$, $S_i \cap S_j = \emptyset$.
Ako je $S = S_1 \cup S_2 \cup ... \cup S_k$, onda skup S ima n=$n_1 + n_2 + ... + n_k$ elemenata:

$$|S_1 \cup S_2 \cup ... \cup S_k| = |S_1| + |S_2| + ... + |S_k|.$$

TEOREM 1.4 *(PRINCIP PRODUKTA)*

Neka konačni skupovi S_i imaju n_i elemenata, $n_i \in \mathbb{N}$, $i = 1, 2, .., k$,
$|S_i| = n_i$. Ako je $S = S_1 \times S_2 \times ... \times S_k$ (kartezijev produkt skupova), onda skup S ima n=$n_1 \cdot n_2 \cdot ... \cdot n_k$ elemenata (uređenih k-torki s=(s_1,s_2, ...,s_k)):

$$|S_1 \times S_2 \times ... \times S_k| = |S_1| \cdot |S_2| \cdot ... \cdot |S_k|.$$

PRIMJER 1.4 *Bacamo dvije igraée kocke različite boje.*
(a) Koliko različitih ishoda ima ako bacamo jednu za drugom?
(b) Na koliko različitih načina mogu pasti ako ih bacamo zajedno?

Rješenje:
S_1 ima $n_1 = 6$ elemenata, S_2 ima $n_2 = 6$ elemenata.
(a) $S = S_1 \cup S_2$, $|S| = |S_1| + |S_2| = n_1 + n_2 = 12$.
(b) $S = S_1 \times S_2$, $|S| = |S_1| \cdot |S_2| = n_1 n_2 = 36$.

Princip produkta ili osnovni princip kombinatorike ima drugu interpretaciju u teoremu o uzastopnom prebrojavanju.

TEOREM 1.5 *(TEOREM O UZASTOPNOM PREBROJAVANJU)*

Proučavamo uređene k-torke. Neka prvi element uređene k-torke možemo izabrati na n_1 načina, a za već izabrani prvi element, drugi možemo izabrati na n_2 načina i tako dalje do k-tog koji možemo izabrati na n_k načina. Tada uređenu k-torku možemo izabrati na n=$n_1 \cdot n_2 \cdot ... \cdot n_k$ načina.

Prisjetimo se motivacijskog primjera:

PRIMJER 1.5 *motiv*

U gradu ima 8 studentskih restorana ravnomjerno raspoređenih u 4 gradske četvrti. U okolini svakog restorana nalaze se dvije sportske dvorane. Student želi unajmiti stan. Na koliko načina može odabrati četvrt, studentski restoran i sportsku dvoranu ako: a) nije bitno ni da studentski restoran bude u istoj četvrti niti dvorana; b) nije bitno da studentski restoran bude u istoj četvrti ali dvorana treba biti; c) sve bude u najbližoj okolini.

Rješenje:

a) n = 4·8·16=512,

b) n = 4·8·4=128

c) n = 4·2·2=16.

1.3 PERMUTACIJE BEZ PONAVLJANJA

MOTIV 1.3 *Koliko se lozinki duljine 4 znaka može formirati od slova v i c (mala slova) i brojeva 1 i 5 ako nije dozvoljeno ponavljanje znakova?*

Definicija 1.2 *Neka skup S ima n različitih elemenata. Svaka uređena n-torka elemenata skupa S zove se permutacija skupa S.*

T: Broj svih n-članih permutacija je

$$P(n) = n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1;$$

D: Prema teoremu o uzastopnom prebrojavanju, prvi element možemo izabrati na n načina, drugi možemo izabrati na $(n-1)$ načina, treći na $(n-2)$ načina itd.

D: *tko želi znati više*

Pomoću AMI po n : Neka je $M = \{n \in \mathbb{N} : P_n = n!\}$.
(B.I) Provjeravamo je li $1 \in M$. Za $n = 1$ broj uređenih jednorki iz skupa koji ima samo 1 element jednak $P_1 = 1$. Prema formuli $n! = 1! = 1$ za $n = 1$; Tako je $1 \in M$.
(P.I) Pretpostavimo $m \in M$,
tj. vrijedi formula $V_m = m!$.
(K.I) U koraku indukcije trebamo pokazati da je $m + 1 \in M$,
tj. treba pokazati da vrijedi $P_{m+1} = (m + 1)!$.
Skup svih permutacija od $m + 1$ element možemo podijeliti na $m + 1$ podskupova u kojima su permutacije s fiksnim prvim elementom npr. a_1, sve permutacije s fiksnim prvim elementom a_2 itd. Tako je

$$P_{m+1} = (m + 1) \cdot P_m = (P.I.) = (m + 1) \cdot m! = (m + 1)!,$$

pa smo pokazali da je $m + 1 \in M$.
Prema AMI zaključujemo da je $M = \mathbb{N}$.

Prisjetimo se motivacijskog primjera:

PRIMJER 1.6 *motiv*

Koliko se lozinki duljine 4 znaka može formirati od slova v i c (mala slova) i brojeva 1 i 5, ako nije dozvoljeno ponavljanje znakova.

Rješenje:
Svaka lozinka je jedna permutacija bez ponavljanja $n = 4$-članog skupa $S = \{v, c, 1, 5\}$. Ukupan broj permutacija bez ponavljanja je $P(4) = 4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$. Možemo formirati 24 lozinke.

PRIMJER 1.7 *Koliko ima svih četveroznamenkastih brojeva sastavljenih od znamenki skupa S={1,2,3,4} takvih da se znamenke ne ponavljaju?*

Rješenje:
Svaki 4-znamenkasti broj je permutacija $n = 4$-članog skupa S. Broj permutacija je $P(4) = 4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$.

1.3.1 PERMUTACIJE S PONAVLJANJEM

MOTIV 1.4 *Koliko se lozinki duljine 4 znaka može formirati od slova v i c (mala slova) i brojeva 1 i 5, ako sa slovo v ponovi 2 puta?*

MOTIV 1.5 *Eksperiment koji ima k=3 ishoda ponavljamo n=7 puta. Koliko je mogućih nizova eksperimenata takvih da se prvi ishod dogodi 1 put, drugi ishod 2 puta, a treć ishod 4 puta?*

Definicija 1.3 *Neka skup S ima n elemenata od kojih je n_1 jedne vrste, n_2 druge vrste, ... , n_k k-te vrste, $n = n_1 + n_2 + ... + n_k$. Uređena n-torka elemenata skupa S zove se n-člana permutacija s ponavljanjem.*

T: Broj n-članih permutacija s ponavljanjem je

$$\overline{P}_n(n_1, n_2, ..., n_k) = \frac{n!}{n_1!n_2!...n_k!}.$$

D: *tko želi znati više*

Pretpostavimo da su svi elementi u permutaciji sa ponavljanjem različiti i da imamo permutaciju bez ponavljanja od $n = n_1 + n_2 + ... + n_k$ elemenata. Ukupan broj tih permutacija je $(n_1 + n_2 + ... + n_k)!$. U permutaciji s ponavljanjem možemo zamijeniti mjesta na kojima su elementi prve vrste na $n_1!$ načina i pri tom se permutacija neće promijeniti. Slično zaključujemo i za elemnte druge, ..., k-te vrste. Za svaku permutaciju s ponavljanjem postoji $n_1!n_2!...n_k!$ permutacija bez ponavljanja u kojima se ne mijenja poredak različitih elemnata skupa S. Mijenjajući poredak različitih elemenata dobili bismo ukupan broj permutacija bez ponavljanja od $n = n_1 + n_2 + ... + n_k$ elemenata. Stoga je ukupan broj permutacija bez ponavljanja jednak $(n_1 + n_2 + ... + n_k)! = n_1!n_2!...n_k!\overline{P}_n(n_1, n_2, ..., n_k)$. Otuda slijedi da je ukupan broj permutacija s ponavanjem dan formulom

$$\overline{P}_n(n_1, n_2, ..., n_k) = \frac{n!}{n_1!n_2!...n_k!}.$$

Prisjetimo se motivacijskog primjera:

PRIMJER 1.8 *motiv*

Koliko se lozinki duljine 4 znaka može formirati od slova v (malo slovo) i brojeva 1 i 5, ako se slovo v ponovi 2 puta?

Rješenje:
Lozinka je je 4-člana permutacija s ponavljanjem elemenata iz skupa $S = \{v, v, 1, 5\}$, tako da je $n = n_1 + n_2 + n_3 = 2 + 1 + 1 = 4$. Broj permutacija s ponavljanjem je $\overline{P}_n(n_1, n_2, ..., n_k) = \overline{P}_4(2, 1, 1) = \frac{4!}{2!1!1!} = 12$. Možemo formirati 12 lozinki.

PRIMJER 1.9 *Koliko ima peteroznamenkastih brojeva sastavljenih od znamenki iz skupa* $S = \{1, 2, 3, 4\}$, *takvih da se znamenke ponavljaju i to znamenka 1 dva puta?*

Rješenje:
Svaki 5-znamenkasti broj je 5-člana permutacija s ponavljanjem elemenata tako da je $n = n_1 + n_2 + n_3 + n_4 = 2 + 1 + 1 + 1 = 5$. Broj permutacija je $\overline{P}_n(n_1, n_2, ..., n_k) = \overline{P}_5(2, 1, 1, 1) = \frac{5!}{2!1!1!1!} = 60$.

PRIMJER 1.10 *U kutiji je 10 kuglica: 4 crne, 2 bijele, 2 crvene, 1 zelene i 1 plava. Izvlačim ih jednu po jednu i slažem po redu u niz. (zapišemo i vratimo). Koliko mogućih nizova možemo napraviti?*

Rješenje:
Svaki niz duljine 10 je permutacija s ponavljanjem, tako da je $n = n_1 + n_2 + n_3 + n_4 + n_5 = 4 + 2 + 2 + 1 + 1 = 10$. Broj permutacija je $\overline{P}_n(n_1, n_2, ..., n_k) = \overline{P}_{10}(4, 2, 2, 1, 1) = \frac{10!}{4!2!2!1!1!} = 604800$.

PRIMJER 1.11 *Koliko različitih riječi se može napraviti od slova u riječi VIVAMARIA?*

Rješenje: $\overline{P}_9(2,2,3)$.

PRIMJER 1.12 *Uzorak s vraćanjem*

Skup S ima n elemenata od kojih je n_1 jedne vrste, $n_2$druge vrste, ... , n_k k-te vrste, n=$n_1 + n_2 + ... + n_k$. Uređena r-torka ima r_1 elemenata prve skupine, r_2 elemenata druge skupine, ...,r_k elemenata k - te skupine r=r_1+r_2+...+r_k. Broj uređenih r-torki s vraćanjem je:

$$\overline{P}_r(r_1, r_2, ..., r_k) \cdot (n_1)^{r_1}(n_2)^{r_2} \cdot ... \cdot (n_k)^{r_k}.$$

Ako je n veliki a r mali u odnosu na n možemo odrediti broj uređenih r-torki s vraćanjem:

$$\overline{P}_r(r_1, r_2, ..., r_k).$$

PRIMJER 1.13 *U velikoj kutiji su crvene, bijele i plave olovke. Na koliko načina možemo izabrati uzorak od 9 olovaka takav da su 2 crvene, 3 bijele i 4 plave olovke.*

Rješenje:
Koristimo formulu za uzorak s vraćanjem (veliki n u odnosu na r).
Osnovni skup je velik n(nepoznat)). Uzorak je uređena r-torka sastavljena od $r = r_1 + r_2 + r_3$ =2+3+4=9 olovaka.
Broj uzoraka je broj uzoraka s vraćanjem je $\overline{P}_9(4,2,3)$.

Prisjetimo se motivacijskog primjera:

PRIMJER 1.14 *motiv*

Ponavljamo eksperiment koji ima k=3 ishoda n=7 puta a da se uvjeti eksperimenta ne mijenjaju. Broj mogućih nizova eksperimenata takvih da se prvi ishod dogodi $n_1 = 1$ puta, $n_2 = 2$ puta,...,k-ti ishod $n_k = 4$ puta je:

$$\overline{P}_n(n_1, n_2, ..., n_k),$$
$$\overline{P}_7(1,2,4) = \frac{7!}{1! \cdot 2! \cdot 4!}.$$

jer je broj različitih nizova eksperimenata jednak broju uzoraka s vraćanjem.

1.4 VARIJACIJE BEZ PONAVLJANJA

MOTIV 1.6 *(a) Koliko se lozinki duljine 4 znaka može formirati od slova v i c (mala slova) i brojeva 1 i 5, ako nije dozvoljeno ponavljanje znakova*
(b) Koliko se lozinki duljine 3 znaka može formirati od slova v i c (mala slova) i brojeva 1 i 5, ako nije dozvoljeno ponavljanje znakova

Definicija 1.4 *Neka skup S ima n različitih elemenata. Uređena r-torka (r ≤ n) elemenata skupa S zove se varijacija r-tog razreda od n elemenata.*

T: Broj svih varijacija r-tog razreda od n elemenata je

$$V_n^{(r)} = n(n-1)(n-2)\cdot \ldots \cdot (n-r+1) = \frac{n!}{(n-r)!};$$

$$V_n^{(n)} = P(n) = n!$$

D: Prema teoremu o uzastopnom prebrojavanju, prvi element možemo izabrati na n načina, drugi možemo izabrati na $(n-1)$ načina, treći na $(n-2)$ načina, r-ti na $(n-r+1)$ način.

D: *tko želi znati više*

Pomoću AMI po n : Neka je $M = \{n \in \mathbb{N} : V_n^{(r)} = \frac{n!}{(n-r)!},\ 1 \leq r \leq n\}$.
(B.I) Provjeravamo je li $1 \in M$. Za $n = 1$ parametar $1 \leq r \leq n$ može poprimiti samo vrijednost $r = 1$ pa je broj uređenih jednoorki iz skupa koji ima samo 1 element jednak $V_1^{(1)} = 1$. Prema formuli $\frac{n!}{(n-r)!} = \frac{1!}{(1-1)!} = 1$ za $n = 1$; Tako je $1 \in M$.
(P.I) Pretpostavimo $m \in M$,
pa tada za svaki dani $1 \leq r \leq m$ vrijedi formula $V_m^{(r)} = \frac{m!}{(m-r)!}$.
(K.I) U koraku indukcije trebamo pokazati da je $m + 1 \in M$.
(i) Ako je $r = 1$ onda je očito broj svih uređenih jeednorki iz skupa koji ima $m + 1$ element jednak $V_{m+1}^{(1)} = m+1$. Prema formuli $\frac{n!}{(n-r)!} = \frac{(m+1)!}{(m+1-1)!} = (m+1)$ za $n = m+1$ i $r = 1$.
Tako je $m + 1 \in M$ za $r = 1$.
(ii) Ako je $2 \leq r \leq m + 1$ treba pokazati da vrijedi $V_{m+1}^{(r)} = \frac{(m+1)!}{(m+1-r)!}$.
Za svaki fiksni r možemo skup svih varijacija podijeliti na $m + 1$ podskupove u kojima su varijacije s fiksnim prvim elementom npr. a_1, sve varijacije s fiksnim prvim elementom a_2 itd. Tako je

$$V_{m+1}^{(r)} = (m+1)\cdot V_m^{(r-1)} = (P.I.) = (m+1)\cdot \frac{m!}{(m-r+1)!} = \frac{(m+1)!}{(m+1-r)!}.$$

Tako smo pokazali da je $m + 1 \in M$ za $2 \leq r \leq m + 1$.
Prema AMI zaključujemo da je $M = \mathbb{N}$.

Prisjetimo se motivacijskog primjera:

PRIMJER 1.15 *motiv*
(a) Koliko se lozinki duljine 4 znaka može formirati od slova v i c (mala slova) i brojeva 1 i 5, ako nije dozvoljeno ponavljanje znakova
(b) Koliko se lozinki duljine 3 znaka može formirati od slova v i c (mala slova) i brojeva 1 i 5, ako nije dozvoljeno ponavljanje znakova

Rješenje:
(a) Lozinka duljine 4 znaka je jedna varijacija 4.tog razreda od $n = 4$-članog skupa$S = \{v, c, 1, 5\}$ tj. permutacija bez ponavljanja od 4 elementa. Broj varijacija je $V_4^{(4)} = P(4) = 4! = 24$. Možemo formirati 24 lozinke. (b) Lozinka duljine 3 znaka je jedna varijacija 3-eg razreda od $n = 4$-članog skupa$S = \{v, c, 1, 5\}$. Broj varijacija je $V_4^{(3)} = \frac{4!}{(4-3)!} = 24$. Možemo formirati 24 lozinke.

PRIMJER 1.16 *Koliko ima dvoznamenkastih brojeva sastavljenih od znamenki iz skupa* S=$\{1, 2, 3, 4\}$ *, takvih da se znamenke ne ponavljaju?*

Rješenje:
Svaki 2-znamenkasti broj je varijacija 2-og razreda od $n = 4$-članog skupa S. Broj varijacija je $V_4^{(2)} = \frac{4!}{(4-2)!} = 12$..

PRIMJER 1.17 *Koliko ima četveroznamenkastih brojeva sastavljenih od znamenki iz skupa* S=$\{0, 1, 2, 5, 7\}$ *, takvih da se znamenke ne ponavljaju?*

Rješenje:
Svaki 4-znamenkasti broj je varijacija 4-og razreda od $n = 5$ elemeata. Broj varijacija je $V_5^{(4)} = \frac{5!}{(5-4)!} = 120$.
Nula ne može biti na prvom mjestu pa od $V_5^{(4)}$ moramo oduzeti sve varijacije 3-eg razreda od 4 elementa $V_4^{(3)} = \frac{4!}{(4-3)!} = 24$. Traženih brojeva ima $V_5^{(4)} - V_4^{(3)} = 120 - 24 = 96$.

PRIMJER 1.18 *Koliko ima četveroznamenkastih brojeva djeljivih s 5 sastavljenih od znamenki iz skupa* $S = \{0, 1, 2, 5, 7\}$ *, takvih da se znamenke ne ponavljaju?*

Rješenje:
Broj je djeljiv s 5 ako mu je zadnja znamenka 0 ili 5.
Prvo, fiksiramo zadnju znamenku 5.
Svaki takav troznamenkasti broj je varijacija 3-eg razreda od $n = 4$-članog skupa $S = \{0, 1, 2, 7\}$. Broj varijacija je $V_4^{(3)} = \frac{4!}{(4-3)!} = 24$.
Nula ne može biti na prvom mjestu pa od $V_4^{(3)}$ moramo oduzeti sve varijacije 2-og razreda od 3 elementa iz skupa $\{1, 2, 7\}$ $V_3^{(2)} = \frac{3!}{(3-2)!} = 6$.
Zatim, fiksiramo zadnju znamenku 0. Svaki takav 3-znamenkasti broj je varijacija 3-eg razreda od $n = 4$-članog skupa
$S = \{1, 2, 5, 7\}$, pa ih ukupno ima $V_4^{(3)} = \frac{4!}{(4-3)!} = 24$.
Ukupan broj traženih brojeva je $V_4^{(3)} - V_3^{(2)} + V_4^{(3)} = 24 - 6 + 24 = 42$.

PRIMJER 1.19 *Na koliko se načina u razredu u kojem je 30 učenika može odabrati glumačka družina za "Crvenkapicu"? Likovi su Crvenkapica, vuk, baka i lovac.*

Rješenje: $V_{30}^{(4)}$.

1.4.1 VARIJACIJE S PONAVLJANJEM

MOTIV 1.7 *Koliko se lozinki duljine 4 znaka može formirati od slova v i c (mala slova) i brojeva 1 i 5, ako je dozvoljeno ponavljanje znakova.*

Definicija 1.5 *Neka skup S ima n različitih elemenata. Uređena r-torka elemenata n-članog skupa S ali tako da se elementi mogu i ponavljati (r može biti i veće od n) zove se varijacija s ponavljanjem r-tog razreda od n elemenata.*

T: Broj varijacija s ponavljanjem r-tog razreda od n elemenata

$$\overline{V}_n^{(r)} = n^r.$$

D: Prema teoremu o uzastopnom prebrojavanju, prvi element možemo izabrati na n načina, drugi možemo izabrati na n načina, treći na n načina, itd., r-ti na n načina jer je dozvoljeno ponavljanje elemenata iz skupa S.

D: *tko želi znati više*

Pomoću AMI po r : Neka je $M = \{r \in \mathbb{N} : \overline{V}_n^{(r)} = n^r,\ n \in \mathbb{N}\}$.
(B.I) Provjeravamo je li $1 \in M$. Za $r = 1$ broj uređenih jednorki iz skupa koji ima samo n element jednak $\overline{V}_n^{(1)} = n$. Prema formuli $n^r = n^1 = n$ za $r = 1$; pa je $1 \in M$.
(P.I) Pretpostavimo da je $m \in M$,
tj. vrijedi formula $\overline{V}_m^{(r)} = n^m$.
(K.I) U koraku indukcije trebamo pokazati da je $m + 1 \in M$.
tj. treba pokazati da vrijedi $\overline{V}_n^{(m+1)} = n^{m+1}$.
Skup svih varijacija s ponavljanjem možemo podijeliti na n podskupova u kojima su varijacije s ponavljanjem s fiksnim prvim elementom npr. a_1, sve varijacije s fiksnim prvim elementom a_2 itd. Tako je

$$\overline{V}_n^{(m+1)} = n \cdot \overline{V}_n^{(m)} = (P.I.) = n \cdot n^m = n^{m+1}.$$

Pokazali da je $m + 1 \in M$ za $n \in \mathbb{N}$.
Prema AMI zaključujemo da je $M = \mathbb{N}$.

Prisjetimo se motivacijskog primjera:

PRIMJER 1.20 *motiv*

(a) Koliko se lozinki duljine 4 znaka može formirati od slova v i c (mala slova) i brojeva 1 i 5, ako je dozvoljeno ponavljanje znakova. (b) Koliko se lozinki duljine 3 znaka može formirati od slova v i c (mala slova) i brojeva 1 i 5, ako je dozvoljeno ponavljanje znakova.

Rješenje:
(a) Lozinka je varijacija s ponavljanjem 4-tog razreda od $n = 4$-članog skupa $S = v, c, 1, 5$. Broj varijacija je $\overline{V}_4^{(4)} = 4^4 = 256$. Možemo formirati 256 lozinki.
(b) Lozinka je varijacija s ponavljanjem 3-tog razreda od $n = 4$-članog skupa $S = v, c, 1, 5$. Broj varijacija je $\overline{V}_4^{(3)} = 4^3 = 64$. Možemo formirati 256 lozinki.

PRIMJER 1.21 *Koliko ima dvoznamenkastih brojeva sastavljenih od znamenki iz skupa $S = \{1, 2, 3, 4\}$, takvih da se znamenke ponavljaju?*

Rješenje:
Svaki takav dvoznamenkasti broj je varijacija s ponavljanjem 2-og razreda od $n = 4$-članog skupa S. Broj varijacija je $\overline{V}_4^{(2)} = 4^2 = 16$.

PRIMJER 1.22 *Razdioba različitih predmeta*

Svaka razdioba r različitih predmeta na n različitih mjesta je varijacija s ponavljanjem r-tog razreda od n elemenata. Broj razdioba je: $\overline{V}_n^{(r)}$.

PRIMJER 1.23 *3 kuglice različitih boja raspoređujemo u 6 kutija. Koliko razdioba možemo dobiti?*

Rješenje: $\overline{V}_6^{(3)}$

PRIMJER 1.24 *Uočimo da je izbor r kuglica iz jedne kutije u kojoj je n različitih kuglica s vraćanjem u kutiju (a poredak je važan), analogan rasporedu r različitih kuglica u n različitih kutija.*

PRIMJER 1.25 *Iz kutije u kojoj je šest kuglica različitih boja izvlačimo tri kuglice jednu po jednu s vraćanjem ponovo u kutiju. Koliko različitih uzoraka možemo dobiti tim postupkom?*

Rješenje: $\overline{V}_6^{(3)}$.

PRIMJER 1.26 *Listić sportske prognoze ima 10 parova. Svaki par može dobiti oznaku 0,1 ili 2 (poraz, neriješeno,pobjeda domaćina). Koliko listića treba ispuniti da bi sigurno jedan listić bio dobitni?*

Rješenje: $S = \{0, 1, 2\}$, $n = 3$.
Listić je varijacija s ponavljanjem $r = 10$-og razreda od $n = 3$ elementa. Broj varijacija je $\overline{V}_3^{(10)} = 3^{10} = 59049$.

1.5 KOMBINACIJE BEZ PONAVLJANJA

MOTIV 1.8 *U skladištu je 100 proizvoda, 70 proizvoda prve klase, 20 proizvoda druge klase i 10 proizvoda treće klase. Kontrolor testira tri proizvoda i daje pozitvnu ocjenu ako su svi proizvodi prve klase. Na osnovu koliko posto svih uzoraka će kontrolor dati pozitivnu ocjenu proizvoda?*

Definicija 1.6 *Neka skup S ima n različitih elemenata. Svaki r-člani podskup ($r \leq n$) (redoslijed elemenata u skupu nije bitan) n-članog skupa S zove se kombinacija r-tog razreda od n elemenata.*

T: Broj svih kombinacija r-tog razreda je od n elemenata je

(i) $C_n^{(r)} = \binom{n}{r} = \frac{n!}{r! \cdot (n-r)!}$;

(ii) $C_n^{(r)} = \frac{V_n^{(r)}}{r!}$.

D: (i) Budući da u r-članom skupu redoslijed nije bitan onda broj uređenih r-torki od n elemenata moramo podijeliti s brojem permutacija r-članog skupa.
(ii) Prema teoremu o uzastopnom prebrojavanju, prvi element možemo izabrati na n načina, drugi možemo izabrati na $(n-1)$ načina, treći na $(n-2)$ načina, r-ti na $(n-r+1)$ način.

D: *tko želi znati više*

Pomoću AMI po n :
Neka je $M = \{n \in \mathbb{N} : C_n^{(r)} = \frac{n!}{(n-r)! \cdot r!},\ 1 \leq r \leq n\}$.
(B.I) Provjeravamo je li $1 \in M$. Za $n = 1$ parametar $1 \leq r \leq n$ može poprimiti samo vrijednost $r = 1$ pa je broj jednočlanih podskupova iz skupa koji ima samo 1 element jednak $C_1^{(1)} = 1$. Prema formuli $\frac{n!}{(n-r)! \cdot r!} = \frac{1!}{(1-1)! \cdot 1!} = 1$ za $n = 1$; Tako je $1 \in M$.
(P.I) Pretpostavimo $m \in M$,
pa tada za svaki dani $1 \leq r \leq m$ vrijedi formula $C_m^{(r)} = \frac{m!}{(m-r)! \cdot r!}$.
(K.I) U koraku indukcije trebamo pokazati da je $m + 1 \in M$.
($*$) Ako je $r = 1$ onda je očito broj svih jednočlanih podskupova iz skupa koji ima

$m+1$ element jednak $C_{m+1}^{(1)} = m+1$. Prema formuli $\frac{n!}{(n-r)!\cdot r!} = \frac{(m+1)!}{(m+1-1)!\cdot r!} = (m+1)$ za $n = m+1$ i $r = 1$.
Pokazali smo da je $m+1 \in M$ za $r = 1$.
($**$) Ako je $2 \leq r < m+1$ treba pokazati da vrijedi $C_{m+1}^{(r)} = \frac{(m+1)!}{(m+1-r)!\cdot r!}$.
Za svaki fiksni r možemo skup svih kombinacija podijeliti na dva podskupa u kojima su kombinacije bez ponavljanja koje sadrže fiksni element npr. a_1, i podskup u kojem su kombinacije bez ponavljanja koje ne sadrže fiksni element a_1. Tako je

$$C_{m+1}^{(r)} = \cdot C_m^{(r-1)} + C_m^{(r)} = (P.I.) = \frac{m!}{(m-r+1)\cdot(r-1)!} + \frac{m!}{(m-r)\cdot r!} = \frac{m+1!}{(m+1-r)!\cdot r!}.$$

Tako smo pokazali da je $m+1 \in M$ za $2 \leq r < m+1$.
($***$) Ako je $r = m+1$ onda je očito broj svih $m+1$ -članih podskupova iz skupa koji ima $m+1$ element jednak $C_{m+1}^{(m+1)} = 1$. Prema formuli $\frac{n!}{(n-r)!\cdot r!} = \frac{(m+1)!}{(m+1-(m+1))!\cdot(m+1)!} = 1$ za $n = m+1$ i $r = m+1$.
Vrijedi $m+1 \in M$ za $r = m+1$.

Prema AMI zaključujemo da je $M = \mathbb{N}$.

PRIMJER 1.27 *Loto ima 39 brojeva. Izvlači se slučajno 7 brojeva. Koliko različitih listića s kombinacijama 7 brojeva treba ispuniti da se dobije jedan siguran pogodak?*

Rješenje: $S = \{1, 2, 3, ..., 39\}$, $n = 39$.
Listić je kombinacija 7-og razreda ($r = 7$) od 39 elemenata. Broj kombinacija je $C_{39}^{(7)} = \binom{39}{7} = \frac{39\cdot38\cdot37\cdot36\cdot35\cdot34\cdot33}{7!} = 15380937$. Treba ispuniti 15380937 listića da bi bili sigurni da ćemo imati jedan dobitak.

PRIMJER 1.28 *U ravnini je 5 točaka od kojih 3 nikada ne leže na istom pravcu.*
a) Koliko pravaca određuju te točke?
b) Koliko trokuta određuju te točke?

Rješenje: $S = \{1, 2, 3, 4, 5\}$, $n = 5$.
a) Pravac je kombinacija 2-og razreda ($r = 2$) od 5 elemenata. Broj kombinacija je $C_5^{(2)} = \binom{5}{2} = \frac{5\cdot4}{2!} = 10$. Točke određuju 10 pravaca.
b) Trokut je kombinacija 3-eg razreda ($r = 3$) od 5 elemenata. Broj kombinacija je $C_5^{(3)} = \binom{5}{3} = \frac{5\cdot4\cdot3}{3!} = 10$. Točke određuju 10 trokuta.

PRIMJER 1.29 *Uzorak bez vraćanjem*

Skup S ima n elemenata od kojih je n_1 jedne vrste, $n_2$druge vrste, ... , n_k k-te vrste, n=$n_1 + n_2 + ... + n_k$. Uređena r-torka ima r_1 elemenata prve skupine, r_2 elemenata druge skupine, ...,r_k elemenata k - te skupine r=r_1+r_2+...+r_k

Broj uređenih r-torki bez vraćanja je: $C_{n_1}^{r_1} \cdot C_{n_2}^{r_2} \cdot ... \cdot C_{n_k}^{r_k}$.

PRIMJER 1.30 *Studenti dva turnusa biraju po tri predstavnika u Klub studenata prve godine GF. Prvi turnus ima 20 studenata, a drugi 30 studenata. Koliko moguće je različitih sastava Kluba studenata?*

Rješenje:

Koristimo formulu za uzorak bez vraćanja (koristimo teorem o uzastopnom prebrojavanju i broj kombinacija r_i-razreda od n_i elemenata).

$n = n_1 + n_2$ =20+30,$r = r_1 + r_2$ =3+3.

Ukupan broj načina da se dobije 6-člani Klub je

$C_{20}^{(3)} \cdot C_{30}^{(3)} = \binom{20}{3} \cdot \binom{30}{3} = 4628400$.

PRIMJER 1.31 *U skupu od 27 proizvoda 7 je neispravnih. Na koliko načina se može dobiti uzorak koji se sastoji od 5 dobrih i 3 neispravna proizvoda?*

Rješenje:

Koristimo formulu za uzorak bez vraćanja (koristimo teorem o uzastopnom prebrojavanju i broj kombinacija r_i-razreda od n_i elemenata).

$S = 20T + 7D$, $n = n_T + n_D$; $n_T = 20$, $n_D = 7$.

Uzorak $r = 5T + 3D$, $r = r_T + r_D$; $r_T = 5$, $r_D = 3$.

Broj traženih uzoraka je $C_{20}^{(5)} \cdot C_7^{(3)} = \binom{20}{5} \cdot \binom{7}{3} = 542640$.

PRIMJER 1.32 *Ako želimo ispitati kvalitetu 10 proizvoda od kojih je 6 ispravnih i 4 neispravna uzimamo uzorak od tri proizvoda. Koliko uzoraka ima u kojima*

a) nema neispravnih proizvoda

b) ima jedan neispravan proizvod

c) ima barem dva ispravna proizvoda?

Rješenje: Koristimo formulu za uzorak bez vraćanja.
a) Osnovni skup ima $n = 10$ elemenata; $n = n_T + n_D = 6+4$. Uzorak bez neispravnih proizvoda ima $r = r_T + r_D = 3 + 0$.
Broj uzoraka koji nemaju neispravnih proizvoda je: $C_6^{(3)} \cdot C_4^{(0)}$ =20.
b) Osnovni skup ima $n = 10$ elemenata; $n = n_T + n_D = 6+4$. Uzorak s 1 neispravnim proizvodom ima $r = r_T + r_D = 2 + 1$.
Broj uzoraka koji imaju 1 neispravni proizvod je: $C_6^{(2)} \cdot C_4^{(1)} = 60$.
c) Osnovni skup ima $n = 10$ elemenata; $n = n_T + n_D = 6 + 4$. Uzorak s bar dva ispravna proizvoda može biti ako ima 1 ili 0 neispravna proizvoda.
Broj uzoraka s bar 2 ispravna proizvoda je:
$c) = a) + b) = C_6^{(3)} \cdot C_4^{(0)} + C_6^{(2)} \cdot C_4^{(1)} = 20 + 60 = 80$.

Prisjetimo se motivacijskog primjera:

PRIMJER 1.33 *motiv*

U skladištu je 100 proizvoda, 70 proizvoda prve klase, 20 proizvoda druge klase i 10 proizvoda treće klase. Kontrolor testira tri proizvoda i daje pozitvnu ocjenu ako su svi proizvodi prve klase. Na osnovu koliko posto svih uzoraka će kontrolor dati pozitivnu ocjenu proizvoda?

Rješenje:
Broj uzoraka veličine $r = 3$ od $n = 100$ elemenata je $C_{100}^{(3)}$ =161700.
Koristimo formulu za uzorak bez vraćanja.
Osnovni skup ima $n = 100 = n_1 + n_2 + n_3 = 70 + 20 + 10$ proizvoda
Uzorak prve klase ima $r = r_1 + r_2 + r_3 = 3 + 0 + 0 = 3$ proizvoda.
Broj uzoraka prve klase je: $C_{70}^{(3)} \cdot C_{20}^{(0)} \cdot C_{10}^{(0)}$ =54740.
Kontrolor će dati pozitivnu ocjenu u $\frac{54740}{161700} = 0.3385$ ili u $33,8\%$ slučajeva.

1.5.1 KOMBINACIJE S PONAVLJANJEM *tko želi znati više*

MOTIV 1.9 *Na koliko načina možmo rasporediti 3 bagera (jednaki) na 6 gradilišta?*

Definicija 1.7 *Neka skup S ima n različitih elemenata. Svaki r-člani podskup ($r \in \mathbb{N}$), n-članog skupa S gdje se elementi mogu i ponavljati (redoslijed elemenata u r-torci nije bitan) zove se kombinacija s ponavljanjem r-tog razreda od n elemenata.*

T: Broj svih kombinacija s ponavljanjem r-tog razreda od n elemenata je

(i) $\overline{C}_n^{(r)} = \binom{n+r-1}{r}$;

(ii) $\overline{C}_n^{(r)} = C_{n+r-1}^{(r)}$;

(iii) $\overline{C}_n^{(r)} = \overline{P}_{n+r-1}(n-1, r)$.

D: *tko želi znati više*

(i) Pomoću AMI po $s = n + r - 1$:
Neka je $M = \{s = n + r - 1 \in \mathbb{N} : \overline{C}_n^{(r)} = \binom{n+r-1}{r}, n, r \in \mathbb{N}\}$.
(B.I) Provjeravamo je li $1 \in M$.
Za $s = 1$ tj. $n + r = 1$, parametari su $r = 1$, $n = 1$. Broj 1-čl. podskupova iz skupa koji ima samo 1 element jednak je $\overline{C}_1^{(1)} = 1$. Prema formuli $\binom{n+r-1}{r} = \frac{1!}{(1-1)!\cdot 1!} = 1$ za $s = 1$; pa je $s = 1 \in M$.
(P.I) Pretpostavimo da $s = m \in M$,
tj. za $n + r - 1 = m$ vrijedi formula $\overline{C}_n^{(r)} = \binom{n+r-1}{r}$.
(K.I) U koraku indukcije trebamo pokazati da je $s = m + 1 \in M$
tj. da za $n + r - 1 = m + 1$ vrijedi formula $\overline{C}_n^{(r)} = \binom{n+r-1}{r}$.

($*$) Ako je $r = 1$ onda je očito broj svih 1-članih podskupova s ponavljanjem iz skupa koji ima $n = m + 1$ element jednak $\overline{C}_{m+1}^{(1)} = m + 1$. Prema formuli $\binom{n+r-1}{r} = \binom{m+1}{1} = (m + 1)$ za $n = m + 1$ i $r = 1$.
Pokazali smo da je $s = m + 1 \in M$ za $r = 1$.
($**$) Ako je $n = 1$ onda je očito broj svih $m + 1$-članih podskupova s ponavljanjem iz skupa koji ima $n = 1$ element jednak $\overline{C}_1^{(m+1)} = 1$. Prema formuli $\binom{n+r-1}{r} = \binom{m+1}{m+1} = 1$ za $n = 1$ i $r = m + 1$.
Tako je $s = m + 1 \in M$ za $n = 1$.
($* * *$) U ostalim slučajevima, skup svih kombinacija s ponavljanjem možemo podijeliti na dva podskupa u kojima su kombinacije sa ponavljanjem koje sadrže fiksni element npr. a_1, i kombinacije sa ponavljanjem koje ne sadrže fiksni element a_1.

Zato je

$$\overline{C}_n^{(r)} = \overline{C}_n^{(r-1)} + \overline{C}_{n-1}^{(r)} = \binom{n+r-2}{r-1} + \binom{n+r-2}{r} = (P.I.) = \binom{n+r-1}{r}.$$

Pokazali smo da je $s = m + 1 \in M$ za $2 \leq r$ i $2 \leq n$.

Prema AMI zaključujemo da je $M = \mathbb{N}$.

(ii) Pokazat ćemo na primjeru $n = 3$, $r = 2$ da formula vrijedi.
Neka je skup $S = \{1, 2, 3\}$. Sve kombinacije s ponavljanjem 2-og razreda od 3 elementa su (elemente smo poredali po uzlaznim vrijednostima):
{1,1} {1,2} {1,3} {2,2} {2,3} {3,3}.
Tih kombinacija ima $\overline{C}_3^{(2)}$ =6.
Definirajmo preslikavanje koje će ovim podskupovima pridijeliti podskupove dobivene tako da prvom članu podskupa dodamo 0, a drugom članu 1.

{1,1}→{1,2}
{1,2}→{1,3}
{1,3}→{1,4}
{2,2}→{2,3}
{2,3}→{2,4}
{3,3}→{3,4}.

Slike su kombinacije bez ponavljanja 2-og razreda od 4-članog skupa {1,2,3,4}. Preslikavanje je bijekcija između skupa svih kombinacija s ponavljanjem 2-razreda od 3 elementa i kombinacija 2-razreda od 4 elementa. Budući da postoji bijekcija, ti su skupovi jednakobrojni pa vrijedi $\overline{C}_3^{(2)} = C_4^{(2)} = 6$.

(ii) Uočimo da je izbor r kuglica iz jedne kutije u kojoj je n različitih kuglica s vraćanjem u kutiju, analogan rasporedu r jednakih kuglica u n različitih kutija. Svaki raspored kako se r kuglica može rasporediti u n kutija je kombinaciju s ponavljanjem r-tog razreda od n elemenata. Zamislimo da imamo n kutija poredanih u niz i da u njih bacamo k kuglica.
Problem je sada kao da slažemo kuglice i pregrade između njih. Jedan raspored možemo gledati kao uređenu $n - 1 + r$ -torku gdje imamo $n - 1$ izbor pregrada i r izbora za kuglice. Svaki takav raspored je permutacija s ponavljanjem od $n - 1 + r$ elementata od kojih jedne vrste ima $n - 1$, a druge r. Ukupan broj takvih permutacija s ponavljanjem ima $\overline{P}_{n-1+r}(n - 1, r)$.

PRIMJER 1.34 *Zamislimo da imamo n=5 kutija poredanih u niz i da u njih bacamo r=9 kuglica.*
Problem možemo razmatrati kao da slažemo kuglice i pregrade među kutijama. Imamo n-1=4 jednake pregrade i r=9 jednakih kuglica. Dakle, imamo permutacije n+r-1=13 elemenata od kojih su 4 jedne vrste i 9 druge vrste pa je njihov broj: $\overline{P}_{13}(4,9)$.

PRIMJER 1.35 *Razdioba jednakih predmeta*

Svaka razdioba r jednakih predmeta na n različitih mjesta je kombinacija s ponavljanjem r -tog razreda od n elemenata.

Broj razdioba je $\overline{C}_n^{(r)}$.

Prisjetimo se motivacijskog primjera:

PRIMJER 1.36 *motiv*

Na koliko načina možmo rasporediti 3 bagera (jednaki) na 6 gradilišta?

Rješenje: Bagere možemo rasporediti na $\overline{C}_6^{(3)} = \binom{n+r-1}{r} = \binom{6+3-1}{3} = \binom{8}{3} = 56$ načina.

PRIMJER 1.37 *3 kuglice iste boje raspoređujemo u 6 kutija. Koliko razdioba možemo dobiti?*

Rješenje: $\overline{C}_6^{(3)}$.

PRIMJER 1.38 *Uočimo da je izbor r kuglica iz jedne kutije u kojoj je n različitih kuglica s vraćanjem u kutiju (redoslijed nije važan), analogan rasporedu r jednakih kuglica u n različitih kutija.*

PRIMJER 1.39 *Iz kutije u kojoj je 6 kuglica različite boje izvlačimo tri kuglice jednu po jednu s vraćanjem. Koliko uzoraka možemo dobiti ako redoslijed nije važan?*

Rješenje: $\overline{C}_6^{(3)}$.

PRIMJER 1.40 *U prodavaonici se može kupiti 5 vrsta čarapa. Koliko različitih poklona može napraviti prodavač ako je pakirao po 9 pari?*

Rješenje: $S = \{1, 2, 3, 4, 5\}$, $n = 5$ vrsta čarapa.
Poklon je kombinacija s ponavljanjem 9-razreda ($r = 9$) od 5 elemenata.
Broj poklona je $\overline{C}_5^{(9)} = \binom{n+r-1}{r} = \binom{5+9-1}{9} = \binom{13}{9} = 715$.

PRIMJER 1.41 *Iz kutije u kojoj je šest kuglica različitih boja izvlačimo tri odjednom (bez vraćanja). Koliko takvih kombinacija možemo dobiti ako redoslijed nije važan?*

Rješenje: $C_6^{(3)}$.

PRIMJER 1.42 *Iz kutije u kojoj je 6 kuglica različitih boja izvlačimo tri kuglice jednu po jednu s vraćanjem. Koliko uzoraka možemo dobiti ako redoslijed nije važan?*

Rješenje: $\overline{C}_6^{(3)}$.

1.6 Ponovimo

BEZ PONAVLJANJA

broj permutacija od n elemenata	$P(n) = n!$
broj varijacija $r-$tog razreda od n elemenata	$V_n^{(r)} = \frac{n!}{(n-r)!}$
broj kombinacija $r-$tog razreda od n elemenata	$C_n^{(r)} = \binom{n}{r} = \frac{n!}{r!\cdot(n-r)!}$

S PONAVLJANJEM

br. permutacija s pon. od n el.	$\overline{P}_n(n_1, n_2, ..., n_k) = \frac{n!}{n_1!n_2!...n_k!}$
br. varijacija s pon. $r-$tog raz. od n el.	$\overline{V}_n^{(r)} = n^r$
br. kombinacija s pon. $r-$tog raz. od n el.	$\overline{C}_n^{(r)} = \binom{n+r-1}{r}$

IZBOR - s vraćanjem

IZBOR: r-čl. uzorka iz n-čl. skupa različitih elemenata	
nije važan poredak	$\overline{C}_n^{(r)}$
važan poredak	$\overline{V}_n^{(r)}$

IZBOR - bez vraćanja

IZBOR: r-čl. uzoraka iz n-čl. skupa različitih elemenata	
nije važan poredak	$C_n^{(r)}$
važan poredak	$V_n^{(r)}$

RAZDIOBE - proizvoljno predmeta u kutijama

RAZDIOBE: r predmeta u n različitih kutija	
jednakih predmeta	$\overline{C}_n^{(r)}$
različitih predmeta	$\overline{V}_n^{(r)}$

RAZDIOBE - najviše po jedan predmet u kutiji

RAZDIOBE: r predmeta u n različitih kutija	
jednakih predmeta	$C_n^{(r)}$
različitih predmeta	$V_n^{(r)}$

UZORCI

UZORCI: veličine $r = r_1 + r_2 + .. + r_k$	iz $n = n_1 + n_2 + .. + n_k$-čl. skupa
bez vraćanja	$C_{n_1}^{(r_1)} \cdot C_{n_2}^{(r_2)} \cdot \cdot C_{n_k}^{(r_k)}$
s vraćanjem	$\overline{P}_r(r_1, r_2, ..., r_k) \cdot (n_1)^{r_1}(n_2)^{r_2} \cdot ... \cdot (n_k)^{r_k}$
s vraćanjem $n \gg r$	$\overline{P}_r(r_1, r_2, ..., r_k)$

Poglavlje 2

KLASIČNA DEFINICIJA VJEROJATNOSTI

Teorija vjerojatnosti je matematička disciplina čiji je zadatak formirati i proučavati matematički model slučajnog pokusa. U 17. stoljeću, inspirirani kockarskim igramam matematičari B. Pascal i P. Fermat su se prvi bavili vjerojatnosnim problemima.

2.1 PROSTOR ELEMNTARNIH DOGAĐAJA

Pokus (eksperiment) je definiran odnosom uzroka i posljedica. Pretpostavke za realizaciju pokusa su: ponavljanje pokusa proizvoljno konačno mnogo puta i poznavanje moguć ih ishoda.
Ishodi pokusa su jedini objekti za izgradnju matematičkog modela pokusa.
U determinističkom pokusu ishod je jednoznačno određen uvjetima pokusa, a u slučajnom pokusu ishod nije jednoznačno odeđen uvjetima pokusa.

Osnovna pretpostavka slučajnog pokusa je da svako vršenje pokusa mora dati ishod (događaj) koji odgovara jednom i samo jednom elementarnom događaju.

MOTIV 2.1 *(a) Bacamo igraću kocku. Koji su ishodi slučajnog pokusa?*
(b) Bacamo istovremeno dvije igraće kocke. Koji su ishodi slučajnog pokusa?

Definicija 2.1 *(ELEMENTARNI DOGAĐAJ.PROSTOR ELEMENTARNIH DOGAĐAJA)*
Slučajni pokus je definiran svojim osnovnim ishodima koji se međusobno isključuju i zovu se elementarni događaji. Označavaju se malim grčkim slovima $\omega_1, \omega_2, ..$
Skup $\Omega = \{\omega_i : \omega_i$ *=elementarni događaji,* $i = 1, .., n, ...\}$ *je neprazan skup i zove se prostor elementarnih događaja.*

PRIMJER 2.1 *Slučajni pokus= bacanje novčića;*
elementarni događaji: ω_1, ω_2*;* $\omega_1 =$ *"palo pismo",* $\omega_2 =$ *"pala glava";*
Prostor elementarnih događaja $\Omega = \{\omega_1, \omega_2\}$.

Prisjetimo se motivacijskog primjera:

PRIMJER 2.2 *motiv*
(a) Slučajni pokus= bacanje kocke;
elementarni događaji: $\omega_1, ..., \omega_6$*;* $\omega_1 =$ *"pao broj 1", ...,omega*$_6$ $=$ *"pao broj 6";*
Prostor elementarnih događaja $\Omega = \{\omega_1, ..., \omega_6\}$
(b) Slučajni pokus= bacanje istovremeno dvije kocke;
elementarni događaji: $\omega_1, ..., \omega_{36}$*;* $\omega_1 = (1, 1)$,... ,$\omega_{36} = (6, 6)$.
Prostor elementarnih događaja $\Omega = \{\omega_1, ..., \omega_{36}\}$.

PRIMJER 2.3 *Slučajni pokus (eksperiment)= vrijeme trajanja tri sijalice;*
elementarni događaj: $\omega = (x, y, z) \in \mathbb{R}^3 : x > 0, y > 0, z > 0$*; Prostor elementarnih događaja* $\Omega = \{\omega = (x, y, z) \in \mathbb{R}^3 : x > 0, y > 0, z > 0\}$

PRIMJER 2.4 *Slučajni pokus (eksperiment)= dnevna količina padalina i maksimalna dnevna temperatura u Zagrebu;*
elementarni događaj: $\omega = (x, y) : x \leq 0, -30 < y \leq 50$*; Prostor elementarnih događaja* $\Omega = \{\omega = (x, y) \in \mathbb{R}^2 : x \leq 0, -30 < y \leq 50\}$

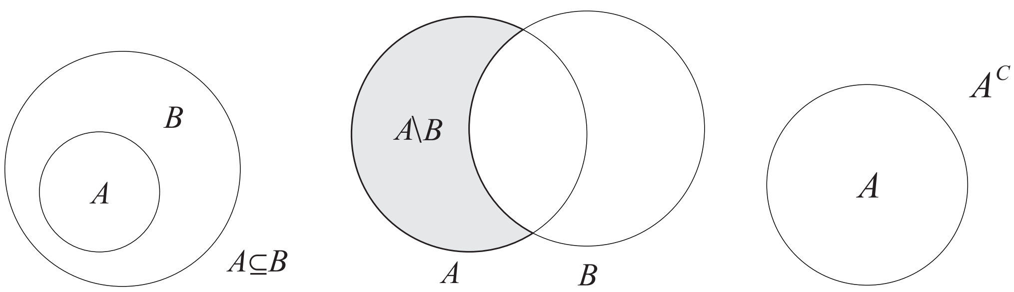
Definicija 2.2 *(SLUČAJNI DOGAĐAJ)*
Slučajni događa je podskup prostora elementarnih događaja . Slučajni događaji označavaju se velikim tiskanim slovima latinice A,B,...; $A \subseteq \Omega$.
(SIGURAN DOGAĐAJ)
Cijeli prostor elementarnih događaja Ω *je siguran događaj koji se mora dogoditi u svakom vršenju pokusa.*
(NEMOGUĆ DOGAĐAJ)

Prazan skup $\emptyset$ je nemoguć događaj koji se nikad neće dogoditi.
(POVOLJAN DOGAĐAJ)

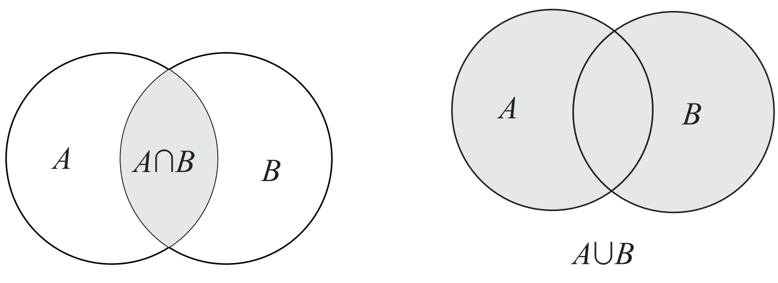
Elementarni događaj koji pripada događaju A zove se povoljan za događaj A ako pojavljivanje tog elementarnog događaja u pokusu povlačda se dogodio događaj A.

PRIMJER 2.5 *Slučajni pokus=bacanje igraće kocke;*
elementarni događaji: ω_1 = *"pala 1",...,* ω_6 = *"pala 6";*
Prostor elementarnih događaja: $\Omega = \{\omega_1, \omega_2, ..., \omega_6\} = \{1, 2, 3, 4, 5, 6\}$;
A="pao je paran broj"; A=$\{\omega_2, \omega_4, \omega_6\} = \{2, 4, 6\}$.
Elementarni događaji $\omega_2, \omega_4, \omega_6$ *su povoljni za događaj A.*

OPERACIJE SA SKUPOVIMA
Slučajni događaji su podskupovi od Ω. Operacije sa događajima definiramo pomoću operacija sa skupovima.
Podskup događaja: $A \subseteq B$: (dogodi se A $\Rightarrow$ dogodi se B);
Razlika događaja: $A \backslash B$: (događaj $A \backslash B$ se dogodi ako se dogodi A i ne dogodi B);
Suprotan događaj: $A^c = \Omega \backslash A$: (A^c se dogodi $\Leftrightarrow$ A ne dogodi);



Presjek događaja : $A \cap B$: (događaj $A \cap B$ se dogodi $\Leftrightarrow$ dogode se i A i B);
Unija događaja:$A \cup B$: (događaj $A \cup B$ se dogodi $\Leftrightarrow$ dogodi se ili A ili B);



T: Vrijede de Morganova pravila:

$(\bigcup_k A_k)^c = \bigcap_k A_k^c;$
$(\bigcap_k A_k)^c = \bigcup_k A_k^c.$

Definicija 2.3 *(DOGAĐAJI SE ISKLJUČUJU)*

Za događaje A i B kažemo da se međusobno isključuju ako je njihov presjek jednak $\emptyset$.

Definicija 2.4 *(POTPUN SISTEM DOGAĐAJA)*

Skupovi $A_1, A_2, ..., A_n$ *čine potpun sistem događaja ako se svi međusobno isključuju i ako im je unija cijeli prostor elementarnih događaja:*

$$A_{i,} \cap A_j = \emptyset,\ \forall i \neq j,\ i = 1, ..., n;\ \bigcup_{i=1}^{n} A_i = \Omega.$$

PRIMJER 2.6 *Iz skupa jednoznamenkastih brojeva izabiremo jedan broj. Neka je događaj A="broj je djeljiv s 2" , događaj B="broj je djeljiv s 3".*
Naći izraze za događaje C:
(a) C="broj djeljiv i s 2 i s 3";
(b) C="broj djeljiv ili s 2 ili s 3";
(c) C="broj paran a nije djeljiv s 3";
(d) C="broj nije paran a djeljiv s 3";
(e) C="broj nije ni paran ni djeljiv s 3".

Rješenje: $A = \{2, 4, 6, 8\}$, $B = \{3, 6, 9\}$
(a) C=$A \cap B = \{6\}$;
(b) C=$A \cup B = \{2, 3, 4, 6, 8, 9\}$;
(c) C=$A \backslash B = \{2, 4, 8\}$;
(d) C=$B \backslash A = \{3, 9\}$;
(e) C=$(A \cup B)^c = A^c \cap B^c = \{1, 5, 7\}$.

PRIMJER 2.7 *Iz tablice slučajnih brojeva izabran je jedan broj. Događaj A="broj je djeljiv s dva", događaj B="zadnja znamenka je 0". Što označava događaj C:*
(a) C=$A \cap B$;
(b) C=$A^c \cap B$;
(c)C=$A \cup B$;

Rješenje:
(a) C="broj je paran i zadnja znamenka je 0";
(b) C=$\emptyset$ nemoguć događaj;
(c) C=A ="broj je paran".

PRIMJER 2.8 *Slučajni pokus je gađanje u metu. Pokus se ponavlja 3 puta. Promatraju se događaji A, B, C koji znače pogađanje mete u prvom, drugom i trećem pokušaju. Pomoću tih događaja opisati slijedeće događaje:*
(a) "sva tri pogotka";
(b) "tri promašaja";
(c) "bar jedan pogodak";
(d) "bar jedan promašaj";
(e) "najviše dva pogotka";
(f) "najviše jedan pogodak";
(g) "bar dva pogotka";
(h) "do trećeg gađanja nije bilo pogodaka".

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

PRIMJER 2.9 *Za koje događaje vrijedi:*
(a) $A \cup B = A$;
(b) $A \cap B = A$;
(c) $A \cap B = A \cup B$;
(d) $A \cap B = A^c$
(e) $A \cup B = A^c$.

Rješenje:
(a) $B \subset A$;
(b) $A \subset B$;
(c) $A = B$;
(d) $A = \Omega, B = \emptyset$;
(e) $A = \emptyset, B = \Omega$.

2.2 KLASIČNA DEFINICIJA VJEROJATNOSTI (A PRIORI)

Francuski matematičar P. S. Laplace (1749.-1827.) objavio je 1812. djelo "Theorie analitique des probabilities" u kojem uvodi pojam elemntarnih događaja i pojam vjerojatnosti a priori i a posteriori. Vjerojatnost događaja a priori određuje preko omjera koji ovisi o prirodi događaja. Za homogeni novčić prije bacanja može se odrediti da je vjerojatnost (a priori) da će pasti pismo jednaka $\frac{1}{2}$. Vjerojatnost događaja a posteriori određuje nakon velikog broja ponavljanja eksperimenta kao omjer broja pojavljivanja događaja i ukupnog broja ponavljanja.

MOTIV 2.2 *U skupu od 27 proizvoda 7 je neispravnih. Kolika je vjerojatnost da izaberemo uzorak koji se sastoji od 5 dobrih i 3 neispravna proizvoda?*

Definicija 2.5 *(KLASIČNA DEFINICIJA VJEROJATNOSTI A PRIORI)*
Neka je prostor elementarnih događaja konačan skup $|\Omega| = n$, $\Omega = \{\omega_1, ..., \omega_n\}$. Neka su svi elementarni događaji jednako mogući. Neka događaj A ima m povoljnih elementarnih događaja, $A \subset \Omega, |A| = m$.
Vjerojatnost svakog elementarnog događaja je $P(\omega) = \frac{1}{|\Omega|}$.
Vjerojatnost događaja A definira se kao broj:

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

T: SVOJSTVA:
(1) $P(\Omega) = 1$,
(2) $P(A^c) = 1 - P(A)$,
(3) $P(\emptyset) = 0$,
(4) $0 \leq P(A) \leq 1$,
(5) $A \subset B \Rightarrow P(A) \leq P(B)$,
(6) $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$,
(7) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

D:
(1) $P(\Omega) = \frac{n}{n} = 1$,
(2) $P(A^c) = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - P(A)$,
(3) $P(\emptyset) = P(\Omega^c) = (2) = 1 - P(\Omega) = (1) = 1 - 1 = 0$,

(4) $A \subset \Omega, 0 < |A| < n \Rightarrow 0 < P(A) < 1$, prema (1) i (3) $\Rightarrow 0 \leq P(A) \leq 1$.
(5) $|A| \leq |B| \Rightarrow \frac{|A|}{n} \leq \frac{|B|}{n} \Rightarrow P(A) \leq P(B)$,
(6) $|A \cup B| = |A| + |B| \Rightarrow$
$P(A \cup B) = \frac{|A \cup B|}{n} = \frac{|A|+|B|}{n} = \frac{|A|}{n} + \frac{|B|}{n} = P(A) + P(B)$,
(7) $|A \cup B| = |A| + |B| - |A \cap B| \Rightarrow$
$P(A \cup B) = \frac{|A \cup B|}{n} = \frac{|A|+|B|-|A \cap B|}{n} = \frac{|A|}{n} + \frac{|B|}{n} - \frac{|A \cap B|}{n} = P(A) + P(B) - P(A \cap B)$.

SLABOSTI:

-restriktivna jer se može primijeniti samo na slučajne pokuse s konačno mnogo elementarnih događaja;

-kružna jer se u definiciji vjerojatnosti koristi formulacija "jednako mogući" tj. "jednako vjerojatni".

PRIMJER 2.10 *Slučajni pokus: bacanje igraće kocke. Kolika je vjerojatnost (klasična apriori) događaja A="pao broj veći ili jednak 3"?*

Rješenje:
$\Omega = \{1, 2, 3, 4, 5, 6\},\ n = |\Omega| = 6$.
$A = \{3, 4, 5, 6\},\ |A| = 4$.
$P(A) = \frac{|A|}{n} = \frac{4}{6} = 0.666$.

PRIMJER 2.11 *U kutiji je pet kuglica: dvije bijele i 3 crne. Iz kutije slučajno izvučemo jednu kuglicu. Kolika je vjerojatnost da će izvučena kuglica biti crna?*

Rješenje:
Slučajni pokus: izvlačimo kuglicu iz kutije u kojoj su 2 b i 3 c kuglice; $\Omega = \{b, b, c, c, c\},\ n = |\Omega| = 5$.
A="kuglica je crna"; $A = \{c, c, c\},\ |A| = 3$,
$P(A) = \frac{|A|}{n} = \frac{3}{5}$.

PRIMJER 2.12 *U kutiji je n olovaka jedne vrste od kojih je n_T ispravnih, a $n_D = n - n_T$ neispravnih. Uzmemo slučajni uzorak od r olovaka. Kolika je vjerojatnost da je među njima r_T ispravnih $(0 \leq r_T \leq r)$i r_D neispravnih?*

Rješenje:
Slučajni pokus: izbor od r elemenata (bez vraćanja) iz skupa koji ima n elemenata $n = n_T + n_D, r = r_T + r_D$.
Ω = {svi uzorci veličine r iz n-čl. skupa}, $|\Omega| = \binom{n}{r}$;
Događaj A="uzorak ima r_T ispravnih i r_D neispravnih";
$|A|$ = Broj svih uzoraka veličine r iz n-čl. skupa bez vraćanja koji imaju r_T ispravnih i r_D neispravnih:
Koristimo formulu za uzorak bez vraćanja (koristimo teorem o uzastopnom prebrojavanju i broj kombinacija r_i-razreda od n_i elemenata).
$|A| = C_{n_T}^{(r_T)} \cdot C_{n_D}^{(r_D)} = \binom{n_T}{r_T} \cdot \binom{n_D}{r_D} = \binom{n_T}{r_T} \cdot \binom{n-n_T}{r-r_T}$.
$P(A) = \frac{|A|}{|\Omega|} = \frac{\binom{n_T}{r_T} \cdot \binom{n-n_T}{r-r_T}}{\binom{n}{r}}$.

Prisjetimo se motivacijskog primjera:

PRIMJER 2.13 *motiv*

U skupu od 27 proizvoda 7 je neispravnih. Kolika je vjerojatnost da izaberemo uzorak koji se sastoji od 5 dobrih i 3 neispravna proizvoda?

Rješenje:
Slučajni pokus: izbor od r elemenata (bez vraćanja) iz skupa koji ima n elemenata $n = n_T + n_D, r = r_T + r_D$.
$n = 27,\ r = 8, n_T = 20,\ n_D = 7,\ r_T = 5,\ r_D = 3$.
Ω ={svi uzorci veličine $r = 8$ iz $n = 27$-čl. skupa};
$|\Omega| = \binom{n}{r} = \binom{27}{8} = 2220\,075$;
Događaj A="uzorak ima $r_T = 5$ ispravnih i r_D =3 neispravnih";
$|A|$ =Broj svih uzoraka veličine r iz n -čl. skupa bez vraćanja koji imaju r_T ispravnih od n_T i r_D neispravnih od n_D:
$|A| = \binom{n_T}{r_T} \cdot \binom{n_D}{r_D} = \binom{n_T}{r_T} \cdot \binom{n-n_T}{r-r_T} = \binom{20}{5} \cdot \binom{7}{3} = 542640$
P(A)=$\frac{|A|}{|\Omega|} = \frac{542640}{2220\,075} = 0.244\,42$.

2.3 KLASIČNA DEFINICIJA VJEROJATNOSTI A POSTERIORI

MOTIV 2.3 *Kolika je vjerojatnost šestice kod bacanja igraće kocke? Bacili smo kocku 1000 puta i 170 puta je pala šestica.*

Definicija 2.6 *(KLASIČNA DEFINICIJA VJEROJATNOSTI A POSTERIORI)*

Neka se slučajni pokus ponavlja n puta, $n \in N$, i neka događaj A nastupi n_A puta. Ako slučajni pokus zadovoljava uvjet statističke stabilnosti relativnih frekvencija, tj. $\lim\limits_{n\to\infty} \frac{n_A}{n} = p$, onda se vjerojatnost a posteriori događaja A definira kao broj:

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n} = p, \quad 0 \leq P(A) \leq 1$$

Često se P(A) zove STATISTIČKA DEFINICIJA VJEROJATNOSTI.

T: SVOJSTVA:
$\omega_1 = 0 =$ "nije se dogodio A";
$\omega_2 = 1 =$ "dogodio se A";
$\Omega = \{0,1\} \times ... \times \{0,1\}, n$ puta;
(1) $P(\Omega) = \lim\limits_{n\to\infty} \frac{n_\Omega}{n} = \lim\limits_{n\to\infty} \frac{n}{n} = 1$,
(2) $P(A^c) = 1 - P(A)$,
(3) $P(\emptyset) = 0$,
(4) $0 \leq P(A) \leq 1$,
(5) $A \subset B \Rightarrow P(A) \leq P(B)$,
(6) $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$,
(7) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

D: analogno kao klasična definicija vjerojatnosti a priori.

SLABOSTI:
P(A)=$\lim\limits_{n\to\infty} \frac{n_A}{n}$,
$\forall \epsilon > 0, \exists n_0 \in N, \forall n > n_0 \Rightarrow |P(A) - \frac{n_A}{n}| < \epsilon$, ali n_0 ovisi o realizaciji slučajnog pokusa.
Za konkretan slučajni pokus teško provjeriti ima li svojstvo stabilnosti relativnih frekvencija!

Prisjetimo se motivacijskog primjera:

PRIMJER 2.14 *motiv*

Kolika je vjerojatnost šestice kod bacanja igraće kocke? Bacili smo kocku 1000 puta i 170 puta je pala šestica.

Rješenje: Slučajni pokus: bacanje kocke.
$\omega_1 = 1 =$ " pala broj 1"; $\omega_6 = 6 =$ "pao broj 6";
$\Omega = \{1, 2, .., 6\} \times ... \times \{1, 2, .., 6\}, n = 1000$ puta;
Pokus zadovoljava uvjet statističke stabilnosti relativnih frekvencija ako jekocka homogena.
$P(\omega_1) = \lim\limits_{n\to\infty} \frac{n_{\omega_1}}{n} = p = \frac{1}{6}$ (iskustvo),...,
$P(\omega_6) = \lim\limits_{n\to\infty} \frac{n_{\omega_6}}{n} = p = \frac{1}{6}$ (iskustvo).
$A = \omega_6,\ n_{\omega_6} = 170$
$P(A) = \frac{n_{\omega_6}}{n} = \frac{170}{1000} = 0.1705.$

PRIMJER 2.15 *Bacamo novčić n=24000 puta. Kolika je vjerojatnost a posteriori događaja A="palo pismo" ako je pismo palo 12012 puta.*

Rješenje: Slučajni pokus: bacanje novčića.
$\omega_1 = 0 =$ "nije palo pismo"; $\omega_2 = 1 =$ "palo pismo";
$\Omega = \{0, 1\} \times ... \times \{0, 1\}, n = 24000$ puta;
Pokus zadovoljava uvjet statističke stabilnosti relativnih frekvencija ako je homogen.
$P(\omega_1) = \lim\limits_{n\to\infty} \frac{n_{\omega_1}}{n} = p = \frac{1}{2}$ (iskustvo),
$P(\omega_2) = \lim\limits_{n\to\infty} \frac{n_{\omega_2}}{n} = p = \frac{1}{2}$ (iskustvo).
$A = \omega_2,\ n_{\omega_2} = 12012$
$P(A) = \frac{n_{\omega_2}}{n} = \frac{12012}{24000} = 0.5005.$

2.4 GEOMETRIJSKA VJEROJATNOST

MOTIV 2.4 *Mladić i djevojka su dogovorili sastanak na trgu u 12 sati. Čekat će se najdulje 20 minuta nakon dolaska. Kolika je vjerojatnost da se susretnu ako su oboje došli na trg od 12 do 13 sati?*

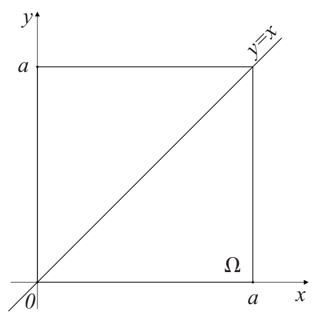
Definicija 2.7 *Neka se slučajni pokus sastoji u slučajnom izboru točke u skupu* $\Omega \subseteq R^n$ *za koji vrijedi* $\mu(\Omega) > 0$*, gdje je* μ *geometrijska mjera skupa:*

- $n = 1$ *duljina;*
- $n = 2$ *površina;*
- $n = 3$ *obujam.*

Neka je događaj $A \subset \Omega$ *izbor točke iz skupa* A*. Geometrijska vjerojatnost događaja* A *tj. vjerojatnost da je izabrana točka iz skupa* A *je broj*

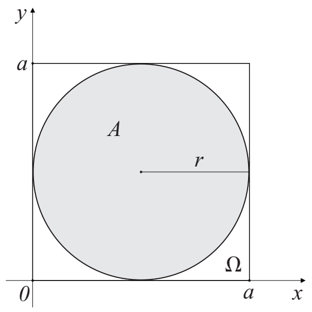
$$P(A) = \frac{\mu(A)}{\mu(\Omega)}.$$

PRIMJER 2.16 *U kvadratu stranice a slučajno je izabrana jedna točka. Kolika je vjerojatnost da je točka s dijagonale kvadrata?*



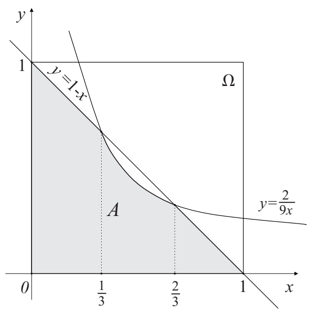
Rješenje:
Slučajni pokus: slučajni izbor točke u skupu $\Omega = [0, a] \times [0, a]$.
A=dijagonala kvadrata, $A \subset \Omega$, $A = \{(x, y) \in \Omega : x = y\}$.
$\mu(\Omega) = a^2$, $\mu(A) = 0$, $P(A) = \frac{\mu(A)}{\mu(\Omega)} = 0$.

PRIMJER 2.17 *U kvadratu stranice a slučajno je izabrana jedna točka. Kolika je vjerojatnost da je točka unutar upisanog kruga?*



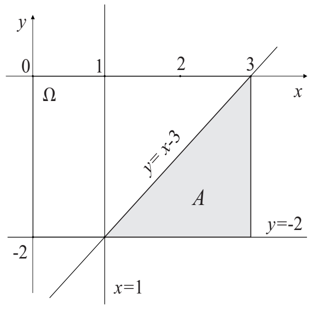
Rješenje:
Slučajni pokus: slučajni izbor točke u skupu $\Omega = [0, a] \times [0, a]$.
A=upisani krug u kvadrat, $A \subset \Omega$, $A = \{(x, y) \in \Omega : x^2 + y^2 = (\frac{a}{2})^2\}$.
$\mu(\Omega) = a^2$, $\mu(A) = (\frac{a}{2})^2\pi$, $P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{(\frac{a}{2})^2\pi}{a^2} = \frac{\pi}{4}$.

PRIMJER 2.18 *Kolika je vjerojatnost da zbir dva slučajno izabrana broja unutar segmenta [0,1] bude manji od 1, a da njihov produkt bude manji od 2/9?*



Rješenje: $\Omega = [0, 1] \times [0, 1]; A \subset \Omega$,
$A = \{(x, y) \in \Omega : 0 \le x \le 1/3, 0 \le y \le 1 - x\} \cup$
$\{(x, y) \in \Omega : \dfrac{1}{3} \le x \le 2/3, 0 \le y \le \dfrac{2}{9x}\} \cup$
$\{(x, y) \in \Omega : \dfrac{2}{3} \le x \le 1, 0 \le y \le 1 - x\}$.
$\mu(\Omega) = 1$, $\mu(A) = \frac{1}{3} + \frac{2}{9}\ln 2$, $P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{1}{3} + \frac{2}{9}\ln 2$.

PRIMJER 2.19 *Na brojevnom pravcu odaberem slučajno točke a i b tako da je a* $\in [0, 3], b \in [-2, 0]$. *Odredite vjerojatnost da je udaljenost točaka a i b veća od 3?*



Rješenje: Ω= točka je unutar pravokutnika $[0, 3] \times [-2, 0]$.
A = točka je unutar trokuta $\{(x, y) \in \Omega : x - y > 3,\ x > 1,\ y > -2\}$.
$\mu(\Omega) = 6,\ \ \mu(A) = 2, P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{6}{2} = \frac{1}{3}$.

Prisjetimo se motivacijskog primjera:

PRIMJER 2.20 *motiv Mladić i djevojka su se dogovorili sastanak na trgu u 12 sati i čekat će se najdulje 20 minuta nakon dolaska. Kolika je vjerojatnost da se susretnu ako su oboje došli na trg od 12 do 13 sati?*

Rješenje: Susret je unutar vremenskog razmaka od 60 minuta pa je $\Omega = [0, 60] \times [0, 60]$. Treba odrediti vjerojatnost događaja A=par se susreo=$\{(x, y) \in \Omega : |x - y| < 20\}$.
$P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{5}{9}$.

2.5 Ponovimo

Prostor elemntarnih događaja	$\Omega = \{\omega_1, ..., \omega_n\}$
Događaj $A \subset \Omega$	$\|A\|$ broj povoljnih elem. dog. za A
Klasična vjerojatnost a priori	$P(A) = \frac{\|A\|}{\|\Omega\|}$
Klasična vjerojatnost a posteriori	$P(A) = \lim\limits_{n\to\infty} \frac{n_A}{n}$
Geometrijska vjerojatnost	$P(A) = \frac{\mu(A)}{\mu(\Omega)}$

Poglavlje 3

AKSIOMATSKA DEFINICIJA VJEROJATNOSTI

Klasična definicija a priori i klasična definicija a posteriri imaju nedostatke. Ruski matemtičar A. Kolmogorov(1903.-1987.) smatra se ocem moderne terorije vjerojatnosti i on uvodi aksiomatsku definiciju vjerojatnosti u monografiji *Osnovni pojmovi teorije vjerojantosti*(1933.).

MOTIV 3.1 *Vjerojatnost da jedna vrsta gume ima rok trajanja veći od 10000 km je 0.95. Kolika je vjerojatnost da će na autu*
(a) sve gume trajati duže od 10000 km?
(b) barem jedna guma puknuti prije pređenih 10000 km?

MOTIV 3.2 *(a) Tri bagerista su slučajno uzimali ključeve svojih strojeva iz kutije. Kolika je vjerojatnost da je bar jedan izabrao ključeve svog vozila?*

tko želi znati više
(b)Ako je bilo n bagerista kolika je vjerojatnost da je bar jedan sjeo u svoj bager?

Definicija 3.1 *(SKUP SVIH MOGUĆIH DOGAĐAJA SLUČAJNOG POKUSA)*
Neka je Ω prostor elementarnih događaja. Partitivni skup ili skup svih podskupova od Ω, $\mathcal{P}(\Omega)$ zovemo skup svih mogućih događaja slučajnog pokusa.
Podskup $\mathcal{A}\subset\mathcal{P}(\Omega)$ zovemo familija događaja iz Ω.

Definicija 3.2 *(SIGMA ALGEBRA DOGAĐAJA)*

Neka familija događaja $\mathcal{F} \subset \mathcal{P}(\Omega)$ ima svojstva:

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

(ii) $A \subset \mathcal{F} \Rightarrow A^c \subset \mathcal{F}$

(iii) Ako je $A_i \in \mathcal{F}, i \in \mathbb{N} \Rightarrow \bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{F}$

Takvu familiju skupova $\mathcal{F}$ zovemo sigma algebra događaja (σ-algebra).
Ako je Ω konačan skup onda je i svaka σ-algebra $\mathcal{A} \subset \mathcal{P}(\Omega)$ konačna i naziva se algebra događaja.

PRIMJER 3.1 *Familija skupova $\mathcal{F}_0 = \{\emptyset, \Omega\}$ je σ-algebra (minimalna).*
Familija skupova $\mathcal{F} = \mathcal{P}(\Omega)$ je σ-algebra (maksimalna).

T: SVOJSTVA:

(a) $\Omega \in \mathcal{F}$,

(b) Ako je $A_i \in \mathcal{F}, i \in \mathbb{N} \Rightarrow \bigcap\limits_{i=1}^{\infty} A_i \in \mathcal{F}$.

D:

(a) Prema definiciji (i), (ii) $\Omega = \emptyset^c \in \mathcal{F}$.

(b) Prema definiciji (iii) $\bigcap\limits_{i=1}^{\infty} A_i = (\bigcup\limits_{i=1}^{\infty} A_i^c)^c \in \mathcal{F}$.

Definicija 3.3 *(AKSIOMATSKA DEFINICIJA VJEROJATNOSTI)*

Neka je Ω prostor elementarnih događaja slučajnog pokusa. neka je $\mathcal{F}$ σ-algebra skupova na Ω . Funkcija $P{:}\mathcal{F} \rightarrow \mathbb{R}$ zove se vjerojatnost na $\mathcal{F}$ ako vrijedi:

(P1) $P(A) \geq 0, A \subset \mathcal{F}$ (svojstvo nenegativnosti);

(P2) $P(\Omega) = 1$ (svojstvo normiranosti);

(P3) $A_i \in \mathcal{F}, i \in \mathbb{N}, A_i \cap A_j = \emptyset, i \neq j \Rightarrow P(\bigcup\limits_{i=1}^{\infty} A_i) = \sum\limits_{i=1}^{\infty} P(A_i)$ (svojstvo prebrojive aditivnosti).

Definicija 3.4 *(VJEROJATNOSNI PROSTOR)*

Vjerojatnosni prostor zovemo uređenu trojku $(\Omega, \mathcal{F}, P)$, gdje je $\mathcal{F}$ σ-algebra na Ω, a P vjerojatnost na $\mathcal{F}$.

Ako je Ω prebrojiv ili konačan skup elementarnih događaja onda $(\Omega, \mathcal{F}, P)$ zovemo diskretni vjerojatnosni prostor.

Ako je Ω konačan skup elementarnih događaja, $|\Omega| = n$, onda $(\Omega, \mathcal{P}(\Omega), P)$ zovemo n dimenzionalni diskretni vjerojatnosni prostor.

Definicija 3.5 *(DOGAĐAJ)*

Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor. Događaj A je element od $\mathcal{F}$.

TEOREM 3.1 *(svojstva funkcije P)*

Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor. Tada za funkciju vjerojatnosti P vrijedi:
(a) $P(\emptyset) = 0$;
(b) $A_i \in \mathcal{F}, i \in \{1, ..., n\}, A_i \cap A_j = \emptyset, i \neq j \Rightarrow P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$
(svojstvo konačne aditivnosti);
(c) $A, B \in \mathcal{F}, A \subset B \Rightarrow P(A) \leq P(B)$ *(svojstvo monotonosti);*
(d) $A \in \mathcal{F} \Rightarrow P(A^c) = 1 - P(A)$;
(e) $A, B \in \mathcal{F} \Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Dokaz:
(a) Neka je $A_1 = \Omega, A_i = \emptyset, i \geq 2$. Prema definiciji vjerojatnosti (P2)

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \Rightarrow P(\Omega) = P(A_1) + \sum_{i=2}^{\infty} P(\emptyset) \Rightarrow 1 = 1 + \sum_{i=2}^{\infty} P(\emptyset) \Rightarrow P(\emptyset) = 0.$$

(b) Neka su $A_i = \emptyset, i > n$. Prema svojstvu (a) $P(A_i) = \emptyset$. Koristeći definiciju vjerojatnosti (P2)

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \Rightarrow P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i).$$

(c) $A \subset B \Rightarrow B = A \cup (B \backslash A)$. Prema svojstvu (b)

$$P(A \cup (B \backslash A)) = P(A) + P(B \backslash A) \Rightarrow P(B) = P(A) + P(B \backslash A).$$

Prema definiciji vjerojatnosti (P1) $P(B \backslash A) \geq 0 \Rightarrow P(B) \geq P(A)$.
(d) $\Omega = A \cup A^c, \Rightarrow 1 = P(A \cup A^c) = P(A) + P(A^c)$.
(e) Uočimo slijedeće relacije $A \cup B = A \cup (B \backslash A)$, $B = (A \cap B) \cup (B \backslash A)$.
Prema svojstvu (b) računamo:
$P(A \cup B) = P(A \cup (B \backslash A)) = P(A) + P(B \backslash A)$, i
$P(B) = P((A \cap B) \cup (B \backslash A)) = P(A \cap B) + P(B \backslash A)$.
Zaključujemo da je $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

NAPOMENA 3.1 *Vjerojatnost je funkcija* $P : \mathcal{F} \to [0, 1]$.

NAPOMENA 3.2 *važno*
T: Ako je Ω konačan i ako su svi elementarni događaji jednako vjerojatni onda je vjerojatnost događaja A prema aksiomatskoj definiciji jednaka vjerojatnosti a priori:

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

D:

Neka je $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$ *konačan skup elementarnih događaja koji su svi jednako vjerojatni, a* $(\Omega, \mathcal{P}(\Omega), P)$ *vjerojatnosni prostor. Prema svojstvu normiranosti i zahtjevu da je vjerojatnost svakog elementarnog događaja jednaka:*

$$P(\Omega) = 1 \Rightarrow n \cdot P(\{\omega_i\}) = 1 \Rightarrow P(\{\omega_i\}) = \frac{1}{n}.$$

Neka je događaj $A = \{\omega_{i1}, \omega_{i2}, ..., \omega_{ik}\}$*. Koristeći svojstvo konačne aditivnosti vjerojatnost događaja* $A \subset P(\Omega)$ *je onda jednaka*

$$P(A) = \sum_{\omega_{ik} \in A} P(\{\omega_{ik}\}) = k \cdot P(\{\omega_{ik}\}) = k \cdot \frac{1}{n} = \frac{|A|}{|\Omega|}.$$

Uočimo da je P(A)= klasičnoj definiciji vjerojatnosti a priori.

PRIMJER 3.2 *Bacamo dvije kocke. Kolika je vjerojatnost da će pasti jednaki brojevi ili paran produkt?*

Rješenje: Prema svojstvu (e) $A, B \in \mathcal{F} \Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$, za događaj A =pali jednaki brojevi i događaj B = produkt brojeva koji su pali je paran broj računamo vjerojatnost $P(A \cup B) = \frac{6}{36} + \frac{27}{36} - \frac{3}{36} = \frac{30}{36}$.

PRIMJER 3.3 *Pokažite da vrijedi*

$$A_i \in \mathcal{F}, i \in \{1,2,3\}, \Rightarrow P(\bigcup_{i=1}^{3} A_i) = \sum_{i=1}^{3} P(A_i) - P(A_1 \cup A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3).$$

Rješenje:

tko želi znati više

Koristimo svojstvo vjerojatnosti (e) $A, B \in \mathcal{F} \Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$ za $A = A_1$, $B = A_2 \cup A_3$.

Prisjetimo se motivacijskog primjera:

PRIMJER 3.4 *motiv*

(a) Tri bageristа su slučajno uzimali ključeve svojih strojeva iz kutije. Kolika je vjerojatnost da je bar jedan izabrao ključeve svog vozila?

tko želi znati više

(b)Ako je bilo n bagerista kolika je vjerojatnost da je bar jedan sjeo u svoj bager?

Rješenje: (a) Događaj A_i =i-ti bagerist je izabrao ključeve svog vozila, $i = 1, 2, 3$.
Događaj B =bar jedan bagerist je izabrao ključeve svog vozila; tj. $B = A_1 \cup A_2 \cup A_3$.
$P(B) = P(A_1 \cup A_2 \cup A_3) = \sum\limits_{i=1}^{3} P(A_i) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$.
$P(B) = 3 \cdot \frac{1}{3} - \frac{1}{2*3} - \frac{1}{2\cdot 3} - \frac{1}{2\cdot 3} + \frac{1}{3!} = \frac{2}{3}$

tko želi znati više

(b) $P(A_1 \cup ... \cup A_n) = 1 - \frac{\binom{n}{2}}{n\cdot(n-1)} + \frac{\binom{n}{3}}{n\cdot(n-1)\cdot(n-2)} - ... + (-1)^n \frac{\binom{n}{n}}{n!}$
$P(B) = 1 - e^{-1} \approx 0.68$, za $n > 10$.
Vjerojatnost je ista 0.68 ako je 11 bagerista ili ako je npr. 100 bagerista.

Prisjetimo se motivacijskog primjera:

PRIMJER 3.5 *motiv*

Ako je vjerojatnost da jedne vrsta gume ima rok trajanja veći od 10000 km jednaka 0.95, kolika je vjerojatnost da će na autu
(a) sve gume trajati duže od 10000 km?
(b) barem jedna guma puknuti prije prđenih 10 000 km?

Rješenje:
Događaj A = jedna guma je prešla 10000 km; događaj B = sve četiri su prešle 10000 km, događaj C = bar jedna guma je pukla prije 10000 km; tj. $C = B^c$.
(a) $P(A) = 0.95 P(B) = P(A)^4$;
(b) $P(C) = 1 - P(B) = 1 - 0.95^4$.

NAPOMENA 3.3 *tko želi znati više*

T: Svojstva funkcije vjerojatnosti P :
(f) $A_n \in \mathcal{F}, n \in \mathbb{N},\ A_1 \subset A_2 \subset ... \subset A_n, A_n = \bigcup\limits_{i=1}^{n} A_i$,

$\Rightarrow P(\lim\limits_{n\to\infty} A_n) = P(\bigcup\limits_{i=1}^{\infty} A_i) = \lim\limits_{n\to\infty} P(A_n)$;
(svojstvo neprekidnost vjerojatnosti u odnosu na rastući niz);

(g) $A_n \in \mathcal{F}, n \in \mathbb{N}, A_n \subset A_{n-1} \subset ... \subset A_1, A_n = \bigcap\limits_{i=1}^{n} A_i$

$\Rightarrow P(\lim\limits_{n\to\infty} A_n) = P(\bigcap\limits_{i=1}^{n} A_i) = \lim\limits_{n\to\infty} P(A_n)$,
(svojstvo neprekidnosti vjerojatnosti u odnosu na padajući niz);

D: tko želi znati više

(f) *Za niz događaja* $A_1 \subset A_2 \subset ... \subset A_n, A_n = \bigcup\limits_{i=1}^{n} A_i$*, definiramo pomoćni niz događaja* $B_n \in \mathcal{F}, n \in \mathbf{N}$ *na slijedći način:*
$B_1 = A_1, B_2 = A_2 \backslash A_1, ..., B_n = A_n \backslash A_{n-1}$.
Lako se provjeri da se događaji isključuju $B_i \cap B_j = \emptyset, i \neq j$*, a* $A_n = \bigcup\limits_{i=1}^{n} B_i$.
Definirajmo $A = \lim\limits_{n\to\infty} A_n = \bigcup\limits_{n=1}^{\infty} B_n$.
Prema definiciji vjerojatnosti (P2) $P(\bigcup\limits_{n=1}^{\infty} B_n) = \sum\limits_{n=1}^{\infty} P(B_n)$,

$$\Rightarrow P(\lim_{n\to\infty} A_n) = \sum_{n=1}^{\infty} P(B_n) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = (b) = \lim_{n\to\infty} P(\bigcup_{i=1}^{n} B_i) = \lim_{n\to\infty} P(A_n).$$

(g) *Za niz događaja* $A_n \subset A_{n-1} \subset ... \subset A_1, A_n = \bigcap\limits_{i=1}^{n} A_i$*, definiramo pomoćni niz događaja* $C_n \in \mathcal{F}, n \in \mathbf{N}$ *na slijedeći način:*
$C_1 = A_1 \backslash A_1 = \emptyset, C_2 = A_1 \backslash A_2, ..., C_n = A_1 \backslash A_n$.
Lako se provjeri da je dobiven monotono rastući niz $C_1 \subset C_2 \subset ... \subset C_n \subset C_{n+1} ...$ *i vrijedi* $\bigcup\limits_{i=1}^{n} C_i = A_1 \backslash A_n$*. i za* $A_n = \bigcap\limits_{i=1}^{n} A_i, A = \lim\limits_{n\to\infty} A_n$
$\Rightarrow \bigcup\limits_{n=1}^{\infty} C_n = A_1 \backslash A$.
Prema definiciji (P2) i svojstvu (d) vrijedi: $P(A_1 \backslash A) = P(\bigcup\limits_{n=1}^{\infty} C_n)$,
$P(A_1) - P(A) = \lim\limits_{n\to\infty} P(C_n) = \lim\limits_{n\to\infty} P(A_1 \backslash A_n) = P(A_1) - \lim\limits_{n\to\infty} P(A_n)$,
$\Rightarrow P(A) = \lim\limits_{n\to\infty} P(A_n)$.

3.1 Ponovimo

ALGEBRA DOGAĐAJA

Skup svih mogućih događaja	$\mathcal{P}(\Omega)$
Familija događaja	$\mathcal{F} \subset \mathcal{P}(\Omega)$
sigma algebra događaja	$\mathcal{F}$ sa svojsvima
	$\emptyset \in \mathcal{F}$
	$A \subset \mathcal{F} \Rightarrow A^c \subset \mathcal{F}$
	sadrži prebrojive unije događaja iz $\mathcal{F}$
algebra događaja	$\mathcal{F}$ ako je Ω konačan

AKSIOMATSKA definicija VJEROJATNOSTI

aksiomi vjerojatnosti	
	$P(A) \geq 0, A \subset \mathcal{F}$
	$P(\Omega) = 1$
	svojstvo prebrojive aditivnosti
vjerojatnost	$P : \mathcal{F} \to [0, 1]$
vjerojatnosni prostor	$(\Omega, \mathcal{F}, P)$

Poglavlje 4

UVJETNA VJEROJATNOST

Thomas Bayes (1702 – 1762) uvodi pojam uvjetne vjerojatnosti: vjerojatnost da se dogodi događaj B ako se dogodio događaj A jednaka je kvocijentu vjerojatnosti da se dogode događaji i A i B i vjerojatnosti događaja A.

MOTIV 4.1 *Kontolor u tvornici daje nakon vizualnog pregleda tri tipa odluke: 1) proizvod je defektan i šalje se na daljnje pretrage 2) sumnja se da je proizvod defektan i šalje se na daljnje pretrage 3) proizvod je ispravan. Pokazalo se dosada da je kontolor bio u pravu kad je odlučio: o neispravnosti u* 80% *slučajeva, o sumnji na neispravnost u* 50% *slučajeva, a o ispravnosti proizvoda a u* 90% *slučajeva.*

U toku jednog dana kontrolor donosi prvu odluku kod 50%*, drugu kod* 20% *i treću dijagnozu kod* 30% *proizvoda. (a) Odredite vjerojatnost neispravnosti proizvoda. (b) Odredite vjerojatnost pogrešne odluke tj. odluke kontolora da je proizvod ispravan a on je zaista neispravan. (c) Odredite vjerojatnost nepotrebnih troškova ili slanja na daljnje pretrage proizvoda ako je zaista ispravan.*

Definicija 4.1 *(UVJETNA VJEROJATNOST)*

Neka je $(\Omega, \mathcal{F}, P)$ *vjerojatnosni prostor i neka je* $A \in \mathcal{F}$ *takav da je* $P(A) > 0$*. Tada funkciju* $P_A : \mathcal{F} \to [0,1]$ *zovemo uvjetna vjerojatnost a definiramo* $\forall\ B \in \mathcal{F}$ *kao*

$$P_A(B) = P(B/A) = \frac{P(B \cap A)}{P(A)},$$

vjerojatnost od B uz uvjet da se dogodio A.

$$P(A) = 0 \Rightarrow P_A(B) = P(B).$$

tko želi znati više

TEOREM 4.1 *Uvjetna vjerojatnost je vjerojatnost. Zadovoljava uvjete*
(UP1) $P_A(B) \geq 0, B \subset \mathcal{F}$ *(svojstvo nenegativnosti);*
(UP2) $P_A(\Omega) = 1$ *(svojstvo normiranosti);*
(UP3) $A_i \in \mathcal{F},\ i \in \mathbb{N},\ A_i \cap A_j = \emptyset,\ i \neq j \ \Rightarrow\ P_A(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P_A(A_i)$
(svojstvo prebrojive aditivnosti).

Dokaz: (UP3) $A_i \in \mathcal{F},\ i \in \mathbf{N},\ A_i \cap A_j = \emptyset,\ i \neq j$
$\Rightarrow A_i \cap A \in \mathcal{F},\ i \in \mathbf{N},\ (A_i \cap A) \cap (A_j \cap A) = \emptyset,\ i \neq j$

$$\begin{aligned} P_A(\bigcup_{i=1}^{\infty} A_i) &= \frac{P(\bigcup_{i=1}^{\infty} A_i \cap A)}{P(A)} = \frac{P(\bigcup_{i=1}^{\infty}(A_i \cap A))}{P(A)} = \frac{\sum_{i=1}^{\infty} P(A_i \cap A)}{P(A)} \\ &= \sum_{i=1}^{\infty} \frac{P(A_i \cap A)}{P(A)} = \sum_{i=1}^{\infty} P_A(A_i)\,. \end{aligned}$$

NAPOMENA 4.1 *važno*
T: Ako je Ω *konačan i ako su svi elementarni događaji jednako vjerojatni onda je uvjetna vjerojatnost događaja B uz uvjet da se dogodio događaj A prema aksiomatskoj definiciji jednaka vjerojatnosti a priori:*

$$P_A(B) = \frac{|A \cap B|}{|A|}.$$

D:

Neka je $(\Omega, \mathcal{P}(\Omega), P)$ *diskretni vjerojatnosni prostor, gdje je* Ω *konačan prostor elementarnih događaja i* $P(\{\omega_i\}) = \frac{1}{n}$, $i = \{1, \ldots, n\}$. *Neka je* $A \in \mathcal{P}(\Omega)$, $|A| = m > 0$ *i* $P(A) = \frac{m}{n}$. *Za* $B \in \mathcal{P}(\Omega)$ *tako da je* $|A \cap B| = r$, *uvjetna vjerojatnost* $P_A : \mathcal{P}(\Omega) \to [0,1]$ *je definirana na slijedeći način*

$$P_A(B) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{r}{n}}{\frac{m}{n}} = \frac{r}{m}.$$

PRIMJER 4.1 *Bacamo kocku. Kolika je vjerojatnost da će "pasti" paran broj pod uvjetom da je je "pao" broj manji od 4?*

Rješenje: $\Omega = \{\omega_1, ..., \omega_6\} = \{1, 2, 3, 4, 5, 6\}$, $P(\{\omega_i\}) = \frac{1}{6}$.
$A = \{1, 2, 3\}$, $P(A) > 0$, $P(A) = \frac{3}{6}$ $B = \{2, 4, 6\}$, $P_A(B) =?$
$P_A(B) = \dfrac{P(B \cap A)}{P(A)} = \dfrac{P(\{2\})}{P(A)} = \dfrac{\frac{1}{6}}{\frac{3}{6}} = \dfrac{1}{3}$.

PRIMJER 4.2 *Izabiremo slučajno dva broja između brojeva od 1 do 9. Ako je njihov zbroj paran broj kolika je vjerojatnost da su oba neparna?*

Rješenje:
$\Omega = \{\{\omega_i, \omega_j\} : \omega_i \neq \omega_j, \omega_1, ..., \omega_9 \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}\}$, $P(\{\omega_i\}) = \frac{1}{9}$.
$A \subset \Omega$, $A = \{\omega_{i_1}, \omega_{i_2}\}$, $\omega_{i_1} + \omega_{i_2} =$ paran,

$$P(A) = \frac{C_4^{(2)} + C_5^{(2)}}{C_9^{(2)}} = \frac{\binom{4}{2} + \binom{5}{2}}{\binom{9}{2}} = \frac{4 \cdot 3 + 5 \cdot 4}{9 \cdot 8} = \frac{32}{72} = \frac{4}{9}$$

$B \subset \Omega$, $B = \{\omega_{i_1}, \omega_{i_2}\}$, $\omega_{i_1}, \omega_{i_2} =$ neparni, $P_A(B) =?$

$$P(B \cap A) = \frac{C_5^{(2)}}{C_9^{(2)}} = \frac{20}{72}, P_A(B) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{20}{72}}{\frac{4}{9}} = \frac{5}{8}.$$

TEOREM 4.2 *(FORMULA PRODUKTA VJEROJATNOSTI)*
Vjerojatnost produkta (presjeka) dva događaja ($A \cap B$) jednaka je produktu vjerojatnosti jednog od njih i uvjetne vjerojatnosti drugog, pod uvjetom da se prvi dogodio.

$$P(A \cap B) = P(A) \cdot P_A(B), \quad P(A \cap B) = P(B) \cdot P_B(A).$$

Dokaz:
Ako je $P(A) > 0$, prema definiciji uvjetne vjerojatnosti P_A :
$P_A(B) = \dfrac{P(A \cap B)}{P(A)} \Rightarrow P(A \cap B) = P(A) \cdot P_A(B)$.
Ako je $P(A) = 0 \Rightarrow P_A(B) = P(B) \Rightarrow P(A \cap B) = P(A) \cdot P(B) = 0$.

TEOREM 4.3 *(FORMULA PRODUKTA VJEROJATNOSTI*)*
Vjerojatnost produkta (presjeka) n događaja ($A_1 \cap A_2 \cap ... \cap A_n$) jednaka je

$$\begin{aligned} P(\bigcap_{i=1}^{n} A_i) &= P(A_1) \cdot P_{A_1}(A_2) \cdot P_{(A_1 \cap A_2)}(A_3) \cdot ... \cdot P_{(\bigcap_{i=1}^{n-1} A_i)}(A_n) \\ &= P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1 \cap A_2) \cdot ... \cdot P(A_n/\bigcap_{i=1}^{n-1} A_i). \end{aligned}$$

Dokaz: Pomoću AMI.

PRIMJER 4.3 *Iz špila karata (52 karte) izvlačimo jednu za drugom dvije karte. Kolika je vjerojatnost da obe karte budu pik?*

Rješenje: A="obje karte su pik"
A_1="prva je pik", $P(A_1) = \frac{13}{52}$
A_2="druga je pik", $P(A_2/A_1) = \frac{12}{51}$
$A = A_1 \cap A_2$.
Prema formuli produkta (presjeka) vjerojatnosti:
$P(A) = P(A_1 \cap A_2) = P(A_1) \cdot P(A_2/A_1) = \frac{13}{52} \cdot \frac{12}{51} = \frac{1}{17}$.
Ili direktno (pomoću formule za uzorak bez vraćanja):
$$P(A_1 \cap A_2) = \frac{C_{13}^{(2)} \cdot C_{13}^{(0)} \cdot C_{13}^{(0)} \cdot C_{13}^{(0)}}{C_{52}^{(2)}} = \frac{13 \cdot 12}{52 \cdot 51} = \frac{1}{17}.$$

PRIMJER 4.4 *U kutiji se nalazi 10 kuglica: 6 bijelih i 4 crne. Izvlačimo 3 kuglice jednu za drugom. Kolika je vjerojatnost da će bar jedna od njih biti bijela?*

Rješenje: A="izvučena bar jedna bijela", A^c="izvučene sve crne",
A_1="prva izvučena crna", $P(A_1) = \frac{4}{10}$
A_2="druga izvučena crna", $P(A_1/A_2) = \frac{3}{9}$,
A_3="treća izvučena crna", $P(A_3/A_1 \cap A_2) = \frac{2}{8}$.
Prema formuli produkta (presjeka) vjerojatnosti
$P(A^c) = P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1 \cap A_2) = \frac{4}{10} \cdot \frac{3}{9} \cdot \frac{2}{8} = \frac{1}{30}$.
Ili direktno (pomoću formule za uzorak bez vraćanja):
$P(A^c) = P(A_1 \cap A_2 \cap A_3) = \frac{C_4^{(3)} \cdot C_6^{(0)}}{C_{10}^{(3)}} = \frac{\binom{4}{3} \cdot \binom{6}{0}}{\binom{10}{3}} = \frac{4 \cdot 3 \cdot 2}{10 \cdot 9 \cdot 8} = \frac{1}{30}$.
$P(A) = 1 - P(A^c) = \frac{29}{30}$.

PRIMJER 4.5 *U kutiji se nalazi 50 proizvoda: 20% neispravnih. Kontrolor izvlači 5 proizvoda sukcesivno (bez vraćanja). Kolika je vjerojatnost da će ocjena kontrolora biti pozitivna (svi proizvodi u uzorku ispravni)?*

Rješenje: A="izvučeni svi ispravni predmeti",
A_1="prvi izvučeni ispravan", $P(A_1) = \frac{40}{50}$
A_2="drugi izvučeni ispravan", $P(A_2/A_1) = \frac{39}{49}$,
A_3="treći izvučeni ispravan", $P(A_3/A_1 \cap A_2) = \frac{38}{48}$.

A_4="četvrti izvučeni ispravan", $P(A_4/A_1 \cap A_2 \cap A_3) = \frac{37}{47}$,
A_5="peti izvučeni ispravan", $P(A_5/A_1 \cap A_2 \cap A_3 \cap A_4) = \frac{36}{46}$
Prema formuli produkta (presjeka) vjerojatnosti

$$\begin{aligned} P(A) &= P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1 \cap A_2) \\ &\quad \cdot P(A_4/A_1 \cap A_2 \cap A_3) \cdot P(A_5/A_1 \cap A_2 \cap A_3 \cap A_4) \\ &= \frac{40}{50} \cdot \frac{39}{49} \cdot \frac{38}{48} \cdot \frac{37}{47} \cdot \frac{36}{46} = 0.31 \end{aligned}$$

Ili direktno (pomoću formule za uzorak bez vraćanja):

$$\begin{aligned} P(A) &= P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) = \frac{C_{40}^{(5)} \cdot C_{20}^{(0)}}{C_{50}^{(5)}} = \frac{\binom{40}{5} \cdot \binom{20}{0}}{\binom{50}{5}} \\ &= \frac{40 \cdot 39 \cdot 38 \cdot 37 \cdot 36}{50 \cdot 49 \cdot 48 \cdot 47 \cdot 46} = 0.31 \end{aligned}$$

Definicija 4.2 *(NEZAVISNI DOGAĐAJI)*

Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor i neka su $A, B \in \mathcal{F}$. Za događaje A i B kažemo da su nezavisni ako vrijedi

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

Definicija 4.3 *(FAMILIJA NEZAVISNIH DOGAĐAJA)*

Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor i neka je $A_i \in \mathcal{F}, i \in I$ familija događaja. Kažemo da je to nezavisna familija događaja ako za svaki konačni podskup različitih indeksa $\{i_1, ..., i_k\} \in I$ vrijedi

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

PRIMJER 4.6 *Ako su A i B nezavisni događaji, takvi da $P(A) > 0, P(B) > 0$ onda vrijedi*

$$P(B/A) = P(B) = P_A(B), \quad P(A/B) = P(A) = P_B(A).$$

$$P_A(B) = P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A) \cdot P(B)}{P(A)} = P(B).$$

PRIMJER 4.7 *(a) Proizvoljan događaj A i siguran događaj su nezavisni.*
(b) Proizvoljan događaj A i nemoguć događaj uvijek su nezavisni.

Rješenje:
(a) $\Omega \cap A = A,\;\; P(\Omega) = 1,\;\; P(\Omega \cap A) = P(A) = P(A) \cdot 1 = P(A) \cdot P(\Omega)$.
(b) $\emptyset \cap A = \emptyset,\;\; P(\emptyset) = 0,\;\; P(\emptyset \cap A) = P(\emptyset) = P(\emptyset) \cdot P(A) = 0$.

PRIMJER 4.8 *Ako su A i B nezavisni događaji onda su nezavisni događaji:*
(a) A^c i B,
(b) A i B^c,
(c) A^c i B^c.

Rješenje:
(a) $(A^c \cap B) \cap (A \cap B) = \emptyset,\;\; (A^c \cap B) \cup (A \cap B) = B$
$\Rightarrow P(B) = P(A^c \cap B) + P(A \cap B)$

$$\begin{aligned} \Rightarrow \;\; P(A^c \cap B) &= P(B) - P(A \cap B) = P(B) - P(A) \cdot P(B) \\ &= P(B) \cdot (1 - P(A)) = P(A^c) \cdot P(B). \end{aligned}$$

(b) $(A \cap B^c) \cap (A \cap B) = \emptyset,\;\; (A \cap B^c) \cup (A \cap B) = A,$
$\Rightarrow P(A) = P(A \cap B^c) + P(A \cap B),$

$$\begin{aligned} \Rightarrow \;\; P(A \cap B^c) &= P(A) - P(A \cap B) = P(A) - P(A) \cdot P(B) \\ &= P(A) \cdot (1 - P(B)) = P(A) \cdot P(B^c). \end{aligned}$$

(c) $(A^c \cap B^c) = (A \cup B)^c,$

$$\begin{aligned} \Rightarrow P(A^c \cap B^c) &= P((A \cup B)^c) = 1 - P(A \cup B) \\ &= 1 - (P(A) + P(B) - P(A \cap B)) \\ &= 1 - (P(A) + P(B) - P(A) \cdot P(B)) \\ &= (1 - P(A)) \cdot (1 - P(B)) = P(A^c) \cdot P(B^c). \end{aligned}$$

PRIMJER 4.9 *Motor pokreće električni generator. Vjerojatnost otkazivanja motora u roku jednog mjeseca je 0.08, a generatora 0.04. Kolika je vjerojatnost da ćemo morati popravljati cijeli uređaj tijekom tog mjeseca?*

Rješenje: Prema svojstvu (e) vjerojatnosti

$$A, B \in \mathcal{F} \Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

za događaj A =pokvario se motor i događaj B = pokvario se generator, računamo vjerojatnost $P(A \cup B)$. Budući su događaji A i B nezavisni onda je

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

Vjerojatnost otkazivanja cijelog uređaja je

$$P(A \cup B) = 0.08 + 0.04 - 0.08 \cdot 0.04 = 0.1168.$$

Vjerojatnost da će trebati popravak je 11.68

PRIMJER 4.10 *Ako su A i B nezavisni događaji pozitivnih vjerojatnosti onda se događaji ne isključuju.*

Rješenje: Pretpostavimo suprotno: $A \cap B = \emptyset$.
A i B su nezavisni i pozitivnih vjerojatnosti $\Rightarrow P(A \cap B) = P(A) \cdot P(B) > 0$, što je u kontradikciji s $P(A \cap B) = P(\emptyset) = 0$.
Zaključujemo da je pretpostavka kriva: $\Rightarrow A \cap B \neq \emptyset$, događaji se ne isključuju.

TEOREM 4.4 *(FORMULA POTPUNE VJEROJATNOSTI)*

Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor i neka skupovi $H_1, H_2, ..., H_n \in \mathcal{F}$ čine potpun sistem događaja. Tada

$$\forall A \in \mathcal{F} \quad \Rightarrow \quad P(A) = \sum_{i=1}^{n} P(H_i) \cdot P(A/H_i).$$

Dokaz:
Skupovi $H_1, H_2, ..., H_n \in \mathcal{F}$, čine potpun sistem (familiju) događaja:

$$H_i \cap H_j = \emptyset, \;\; i \neq j, \;\; \bigcup_{i=1}^{n} H_i = \Omega, \;\; \sum_{i=1}^{n} P(H_i) = 1.$$

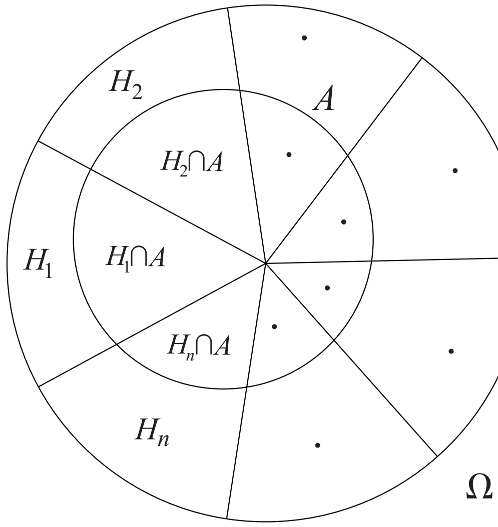
Prema teoremu o vjerojatnosti produkta (presjeka) događaja:
$P(A \cap H_i) = P(H_i) \cdot P(A/H_i)$.
$A \in \mathcal{F}$, $(A \cap H_i) \cap (A \cap H_j) = \emptyset$.

$$\begin{aligned} P(A) &= P(A \cap \Omega) = P(A \cap (\bigcup_{i=1}^{n} H_i)) = P(\bigcup_{i=1}^{n} (A \cap H_i)) \\ &= \sum_{i=1}^{n} P(A \cap H_i) = \sum_{i=1}^{n} P(H_i) \cdot P(A/H_i). \end{aligned}$$

TEOREM 4.5 *(BAYESOVA FORMULA)*

Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor i neka skupovi $H_1, H_2, ..., H_n \in \mathcal{F}$ čine potpun sistem događaja. Neka događaj $A \in \mathcal{F}$ ima pozitivnu vjerojatnost $P(A) > 0$. Tada je $\forall i$

$$P(H_i/A) = \frac{P(H_i) \cdot P(A/H_i)}{P(A)}, \quad P(H_i/A) = \frac{P(H_i) \cdot P(A/H_i)}{\sum_{j=1}^{n} P(H_j) \cdot P(A/H_j)}.$$



Dokaz:
Definicija uvjetne vjerojatnosti i formula produkta vjerojatnosti povlači da vrijedi:

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

Prema formuli potpune vjerojatnosti dobivamo:

$$P(H_i/A) = \frac{P(H_i) \cdot P(A/H_i)}{\sum\limits_{j=1}^{n} P(H_j) \cdot P(A/H_j)}.$$

NAPOMENA 4.2 *Bayesovu formulu koristimo kad želimo naći istinitu hipotezu iz skupa od n postavljenih hipoteza $H_i, i = 1, ..., n$, ako znamo da se dogodio događaj A. Za svako $i = 1, ..., n$, računamo $P(H_i/A)$. Hipoteza H_{i0} za koju je $P(H_{i0}/A) \approx 1$, uzima se da je ispravna.*

PRIMJER 4.11 *Na našem fakultetu je 4% studenata i 1% studentica koji nisu državljani RH. Omjer studenata i studentica upisanih na fakultet je 40:60. Ako je slučajno izabrana jedna osoba upisana na naš fakultet koja je strani državljanin kolika je vjerojatnost da je to studentica?*

Rješenje:
Postavljamo hipoteze-potpun sistem događaja:
H_1 ="izabrana osoba je studentica", $P(H_1) = \frac{60}{100}$.
H_2 ="izabrana osoba je student", $P(H_2) = \frac{40}{100}$.
$H_1 \cap H_2 = \emptyset, \quad H_1 \cup H_2 = \Omega.$

Događaj A koji se dogodio: A="izabrana osoba je strani državljanin".
Zadane su uvjetne vjerojatnosti događaja A uz uvjet jedne i druge hipoteze:
$P(A/H_1) = \frac{1}{100}$, $P(A/H_2) = \frac{4}{100}$. Trebamo odrediti $P(H_1/A) =?$
Koristimo Bayesovu formulu:

$$\begin{aligned} P(H_1/A) &= \frac{P(H_1)\cdot P(A/H_1)}{\sum\limits_{j=1}^{2} P(H_j)\cdot P(A/H_j)} = \frac{\frac{60}{100}\cdot\frac{1}{100}}{P(H_1)\cdot P(A/H_1) + P(H_2)\cdot P(A/H_2)} \\ &= \frac{\frac{60}{100}\cdot\frac{1}{100}}{\frac{60}{100}\cdot\frac{1}{100} + \frac{40}{100}\cdot\frac{4}{100}} = \frac{60}{220} = \frac{3}{11} = 0.27. \end{aligned}$$

PRIMJER 4.12 *Tri stroja S1, S2 i S3 učestvuju u ukupnoj proizvodnji u omjeru* 60 : 30 : 10. *Stroj S1 proizvodi* 2%, *stroj S2* 3% *i stroj S3* 4% *neispravnih proizvoda. Ako se slučajno izabere jedan proizvod koji je neispravan, kolika je vjerojatnost da je bio napravljen na stroju S3?*

Rješenje:
Postavljamo hipoteze-potpun sistem događaja:
H_1 ="izabrani predmet je sa stroja S1", $P(H_1) = \frac{60}{100}$.
H_2 ="izabrani predmet je sa stroja S2", $P(H_2) = \frac{30}{100}$.
H_3 ="izabrani predmet je sa stroja S3", $P(H_3) = \frac{10}{100}$
$H_i \cap H_j = \emptyset, i \neq j, \quad H_1 \cup H_2 \cup H_3 = \Omega$.
Događaj A koji se dogodio: A="izabrani proizvod je neispravan".
Zadane su uvjetne vjerojatnosti događaja A uz uvjet pojedine hipoteze:
$P(A/H_1) = \frac{2}{100}$, $P(A/H_2) = \frac{3}{100}$, $P(A/H_3) = \frac{4}{100}$.
Trbamo odrediti $P(H_3/A) =?$
Koristimo Bayesovu formulu:

$$\begin{aligned} P(H_3/A) &= \frac{P(H_3)\cdot P(A/H_3)}{\sum\limits_{j=1}^{3} P(H_j)\cdot P(A/H_j)} \\ &= \frac{\frac{10}{100}\cdot\frac{4}{100}}{P(H_1)\cdot P(A/H_1) + P(H_2)\cdot P(A/H_2) + P(H_3)\cdot P(A/H_3)} \\ &= \frac{\frac{10}{100}\cdot\frac{4}{100}}{\frac{60}{100}\cdot\frac{2}{100} + \frac{30}{100}\cdot\frac{3}{100} + \frac{10}{100}\cdot\frac{4}{100}} = \frac{4}{25} = 0.16. \end{aligned}$$

Prisjetimo se motivacijskog primjera:

PRIMJER 4.13 *motiv*

Kontolor u tvornici daje nakon vizualnog pregleda tri tipa odluke: 1) proizvod je defektan i šalje se na daljnje pretrage 2) sumnja se da je proizvod defektan i šalje se na daljnje pretrage 3) proizvod je ispravan. Pokazalo se dosada da je kontolor bio u pravu kad je odlučio: o neispravnosti u 80% *slučajeva, o sumnji na neispravnost u* 50% *slučajeva, a o ispravnosti proizvoda a u* 90% *sluČajeva.*

U toku jednog dana kontrolor donosi prvu odluku kod 50%*, drugu kod* 20% *i treću dijagnozu kod* 30% *proizvoda. (a) Odredite vjerojatnost neispravnosti proizvoda. (b) Odredite vjerojatnost pogrešne odluke tj. odluke kontolora da je proizvod ispravan a on je zaista neispravan. (c) Odredite vjerojatnost nepotrebnih troškova ili slanja na daljnje pretrage proizvoda ako je zaista ispravan.*

Rješenje:
Postavljamo hipoteze-potpun sistem događaja:
H_1 ="prva odluka -proizvod je neispravan", $P(H_1) = 0.5$;
H_2 ="druga odluka - sumnja se da je proizvod neispravan ", $P(H_2) = 0.2$;
H_3 ="treća odluka- proizvod je ispravan", $P(H_3) = 0.3$.
$H_i \cap H_j = \emptyset, i \neq j, \quad H_1 \cup H_2 \cup H_3 = \Omega$.
Događaj A = "proizvod je neispravan".
Događaj A^c ="proizvod je ispravan".

Zadane su uvjetne vjerojatnosti događaja A uz uvjet pojedine hipoteze:
$P(A/H_1) = 0.8$, $P(A/H_2) = 0.5$ $P(A/H_3) = 0.1$.
(a) Trebamo odrediti $P(A)$
(b) Trbamo odrediti $P(H_3/A)$ =?
(c) Trebamo odrediti $P(H_1 \cup H_2/A^c)$

(a) Koristimo formulu potpune vjerojatnosti:

$$P(A) = \sum_{j=1}^{3} P(H_j) \cdot P(A/H_j)$$

$$P(A) = 0.5 \cdot 0.8 + 0.2 \cdot 0.5 + 0.3 \cdot 0.1 = 0.53$$

$$P(A^c) = 0.47.$$

(b)Koristimo Bayesovu formulu:

$$
\begin{aligned}
P(H_3/A) &= \frac{P(H_3)\cdot P(A/H_3)}{\sum\limits_{j=1}^{3} P(H_j)\cdot P(A/H_j)} \\
&= \frac{0.3\cdot 0.1}{0.53} \\
&= 0.056.
\end{aligned}
$$

(c) Koristimo Bayesovu formulu:

$$
\begin{aligned}
P(H_1\cup H_2/A^c) &= \frac{P(H_1)\cdot P(A^c/H_1)+P(H_2)\cdot P(A^c/H_2)}{P(A^c)} \\
&= \frac{0.5\cdot 0.2+0.2\cdot 0.50}{0.47} \\
&= 0.425.
\end{aligned}
$$

4.1 Ponovimo

UVJETNA VJEROJATNOST

Uvjetna vjerojatnost od B uz uvjet da se dogodio A	$P_A(B) \equiv P(B/A)$
ako $P(A) > 0$	$P_A(B) = \frac{P(B \cap A)}{P(A)}$
ako $P(A) = 0$	$P_A(B) = P(B)$
formula produkta vjerojatnosti	$P(A \cap B) = P(A) \cdot P_A(B)$
nezavisni događaji A i B	$P(A \cap B) = P(A) \cdot P(B)$
formula potpune vjerojatnosti	$P(A) = \sum\limits_{j=1}^{n} P(H_j) \cdot P(A/H_j)$
Bayesova formula	$P(H_i/A) = \frac{P(H_i) \cdot P(A/H_i)}{\sum\limits_{j=1}^{n} P(H_j) \cdot P(A/H_j)}$

Dio II

Slučajne varijable

Poglavlje 5

DISKRETNA SLUČAJNA VARIJABLA

MOTIV 5.1 *U agenciji za nekretnine prema pokazateljima prodaje u prošlom razdoblju zarada po prodanom jednosobnom stanu je 3000 eura. Kolika je očekivana mjesečna zarada na jednosobnim stanovima ako je prema zadnjim pokazateljima vjerojatnost prodaje nijednog stana 0.1, 1 stana 0.3, 2 stana 0.4, a vjerojatnost prodaje 3 stana 0.2 .*

Opisna definicija pojama slučajne varijable: to je funkcija koja elementarnim događajima nekog slučajnog pokusa pridružuje realan broj.

Ako je pokus takav da slučajna varijabla predstavlja prebrajanje onda ćemo govoriti o diskretnoj slučajnoj varijabli.

Ako je pokus takav da slučajna varijabla uključuje mjerenje (temperatura, tlak, vrijeme, postotak) onda ćemo govoriti o kontinuiranoj slučajnoj varijabli jer se vrijednosti mjerenja prikazuju intervalima.

Diskretna slučajna varijabla se zadaje vrijednostima koje su pridružene ishodima pokusa i vjerojatnostima događaja da slučajna varijabla poprimi te vrijednosti.

Kontinuirana slučajna varijabla zadaje se vjerojatnostima da poprimi vrijednosti unutar pojedinog intervala. Vjerojatnost da kontinuirana slučajna varijabla poprimi pojedinu vrijednost unutar intervala je nula.

PRIMJER 5.1 *Neka je slučajni pokus bacanje dvije kocke istovremeno. Elemntarnim događajima možemo pridružiti realan broj koji odgovara sumi palih brojeva. To pridruživanje je funkcija X sa vrijednostima u skupu* $\{2, 3, 4, ..., 12\}$ *koje ovise o ishodu slučajnog pokusa,* *pa to opravdava naziv slučajna varijabla X =suma brojeva koji su pali. To je diskretna slučajna varijabla.*

PRIMJER 5.2 *Neka je slučajni pokus godišnje poslovanje jednog poduzeća. Kontinuirana slučajna varijabla X se može definirati kao omjer prodaje i profita u jednoj godini.*

5.1 DISKRETNA SLUČAJNA VARIJABLA

MOTIV 5.2 *Promatramo slučajan pokus bacanje 2 igraće kocke i slučajnu varijablu $X =$ suma brojeva koji su pali. Izračunajte vjerojatnost da je zbroj brojeva koji su pali veći od 3 a manji ili jednak 6, tj. $P(3 < X \leq 6)$?*

Definicija 5.1 *(DiSKRETNA SLUČAJNA VARIJABLA*

Neka je $(\Omega, \mathcal{P}(\Omega), P)$ diskretni vjerojatnosni prostor. Funkciju $X : \Omega \to \mathbb{R}$ zovemo diskretna slučajna varijabla ako je slika $\mathcal{R}(X) = \{X(\omega_1), X(\omega_2), ..., X(\omega_n), ...\}$ diskretan (konačan ili prebrojiv) skup.

PRIMJER 5.3 *Neka je $(\Omega, \mathcal{P}(\Omega), P)$ diskretni vjerojatnosni prostor gdje je $\Omega = \{\omega_1, \omega_2\}$. $X : \Omega \to \mathbb{R}$ je diskretna slučajna varijabla definirana na sljedeći način:*

$$X(\omega_1) = 0,\ X(\omega_2) = 1.$$

Definicija 5.2 *Za diskretnu slučajnu varijablu X na vjerojatnosnom prostoru $(\Omega, \mathcal{P}, P)$ definiramo vjerojatnost da poprimi vrijednost $a \in \mathbb{R}$*

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

PRIMJER 5.4 *Neka je $(\Omega, \mathcal{P}, P)$ diskretni vjerojatnosni prostor gdje je $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, a vjerojatnost $P(\{\omega_i\}) = \frac{1}{4}, i = 1, .., 4$. Neka je $X : \Omega \to \mathbb{R}$ slučajna varijabla definirana na sljedeći način: $X(\omega_1) = X(\omega_2) = 0$, $X(\omega_3) = X(\omega_4) = 1$.*
Izračunajte vjerojatnost da slučajna varijabla X poprimi vrijednost 1 tj. izračunajte $P(X = 1)$.

Rješenje:
Vjerojatnost da slučajna varijabla poprimi vrijednost 1 iznosi:

$$\begin{aligned} P(X = 1) &= P(\{\omega \in \Omega : X(\omega) = 1\} \\ &= P(\{\omega_3\}) + P(\{\omega_4\}) = \frac{1}{2}. \end{aligned}$$

Definicija 5.3 *(FUNKCIJA VJEROJATNOSTI SLUČAJNE VARIJABLE, engl. probability function of X)*

Neka je $X : \Omega \to \mathbb{R}$ diskretna slučajna varijabla sa slikom $\mathcal{R}(X) = \{x_1, x_2, ...\}$. Funkciju $f : \mathbb{R} \to \mathbb{R}$ definiranu na sljedeći način:

$$f(x) = \begin{cases} P(X = x_i), & x = x_i \in \mathcal{R}(X) \\ 0, & \text{inače.} \end{cases}$$

zovemo funkcija vjerojatnosti diskretne slučajne varijable X.

T: SVOJSTVA funkcije vjerojatnosti slučajne varijable X.

(i) $0 \leq f(x_i) \leq 1$,

(ii) $\sum_{i=1}^{\infty} f(x_i) = 1$.

D:
(i) Za $I = \{x_i\}$, $P(X = x_i) = P(\{\omega : X(\omega) = x_i\}) \Rightarrow 0 \leq f(x_i) \leq 1$.
(ii) Za $x_i \neq x_j \Rightarrow \{\omega : X(\omega) = x_i\} \bigcap \{\omega : X(\omega) = x_j\} = \emptyset$, svojstvo (ii) slijedi prema svojstvu (P3), prebrojive aditivnosti funkcije P :

$$\sum_{i=1}^{\infty} f(x_i) = \sum_{i=1}^{\infty} P(X = x_i) = P(\Omega) = 1.$$

NAPOMENA 5.1 *važno*

Diskretna slučajna varijabla je zadana sa svojom slikom $\mathcal{R}(X) = \{x_1, x_2, x_3, .., x_n, ..\}$ i funkcijom vjerojatnosti slučajne varijable f. Zapis slučajne varijable X kao uređene sheme:

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

gdje su $p_i = f(x_i)$, $i = 1, 2, ...$

PRIMJER 5.5 *Neka je $(\Omega, \mathcal{P}(\Omega), P)$ diskretni vjerojatnosni prostor gdje je $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $P(\{\omega_i\}) = \frac{1}{4}$, $i = 1, .., 4$. Neka je $X : \Omega \to \mathbb{R}$ slučajna varijabla definirana na sljedeći način: $X(\omega_1) = 7$, $X(\omega_2) = 8$, $X(\omega_3) = 9$, $X(\omega_4) = 10$. Odredite funkciju vjerojatnosti slučajne varijable X .*

Rješenje: Slika slučajne varijable X je $\mathcal{R}(X) = \{7, 8, 9, 10\}$.. Funkcija vjerojatnosti slučajne varijable X je $f : \mathbb{R} \to \mathbb{R}$

$$f(x) := \begin{cases} \frac{1}{4}, & x = x_i \in \{7, 8, 9, 10\} \\ 0, & \text{inače.} \end{cases}$$

Slučajnu varijablu X možemo zadati shemom $X \sim \begin{pmatrix} 7 & 8 & 9 & 10 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$.

PRIMJER 5.6 *Neka je $(\Omega, \mathcal{P}(\Omega), P)$ diskretni vjerojatnosni prostor slučajnog pokusa bacanja igraće kocke. $\Omega = \{\omega_1, \omega_2, \omega_3, ..., \omega_6\}$, a vjerojatnost $P(\{\omega_i\}) = \frac{1}{6}$, $i = 1, .., 6$.*

Neka je $X : \Omega \to \mathbb{R}$ slučajna varijabla definirana na sljedeći način: X = broj koji je pao.

Odredite funkciju vjerojatnosti slučajne varijable X.

Rješenje: Slika slučajne varijable je skup $\mathcal{R}(X) = \{1, 2, 3, 4, 5, 6\}$.

$$P(X = i) = P(\{\omega_i\}) = \frac{1}{6}, \quad i = 1, .., 6$$

Funkcija vjerojatnosti slučajne varijable X je $f : \mathbb{R} \to \mathbb{R}$

$$f(x) := \begin{cases} \frac{1}{6}, & x = x_i \in \{1, 2, 3, 4, 5, 6\} \\ 0, & \text{inače.} \end{cases}$$

Slučajnu varijablu X možemo zadati s $X \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$.

PRIMJER 5.7 *Promatramo slučajan pokus gađanja u metu 3 puta. U svakom gađanju vjerojatnost pogotka mete je $\frac{1}{2}$. Neka je slučajna varijabla X = broj pogodaka u metu. Nađite funkciju vjerojatnosti slučajne varijable X.*

Rješenje: Slučajnu varijablu X možemo zadati s $X \sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{pmatrix}$.

Definicija 5.4 *(FUNKCIJA DISTRIBUCIJE DIS. SL. VARIJABLE, engl. distribution function)*

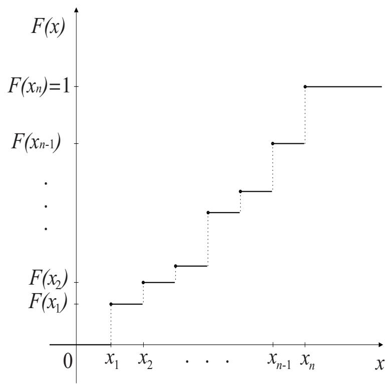
Neka je $X : \Omega \to \mathbb{R}$ diskretna slučajna varijabla sa slikom $\mathcal{R}(X) = \{x_1, x_2, ...\}$. Funkciju $F : \mathbb{R} \to \mathbb{R}$ definiranu na sljedeći način:

$$F(x) := P(X \leq x) = \sum_{i : x_i \leq x} f(x_i)$$

zovemo funkcija distribucije diskretne slučajne varijable X.
Veza funkcije vjerojatnosti i funkcije distribucije diskretne slučajne varijable je:

$$f(x_i) = F(x_i) - F(x_i-),$$

gdje je $F(x_i-) = \lim\limits_{x \to x_i-} F(x)$.



Slika 5.1: Graf funkcije distribucije $F(x)$ diskretne slučajne varijable X.

T: SVOJSTVA FUNKCIJE DISTRIBUCIJE diskretne slučajne varijable:

(F1) $\lim\limits_{x \to -\infty} F(x) = F(-\infty) = 0$

(F2) $\lim\limits_{x \to \infty} F(x) = F(\infty) = 1$

(F3) $0 \le F(x) \le 1$

(F4) F je neprekinuta zdesna, $F(x) = F(x+)$, $F(x-) = P(X < x)$

(F5) $P(a < X \le b) = F(b) - F(a)$, $a, b \in \mathbb{R}$, $a < b$

(F6) F je rastuća funkcija.

D:
(F1) Kako je događaj $X \le -\infty$ nemoguć događaj, slijedi
$F(-\infty) := P(X \le -\infty) = P(\emptyset) = 0$.
(F2) Kako je događaj $X \le \infty$ siguran događaj slijedi
$F(\infty) := P(X \le \infty) = P(\Omega) = 1$.
(F3) tvrdnja slijedi iz svojstava (F1) i (F2).
(F4) Iz definicije funkcije distribucije $F(x) = P(X \le x)$ u točkama prekida

x_i F je neprekinuta zdesna jer $F(x) = F(x+)$, $F(x-) = P(X < x)$
(F5) Neka su $a, b \in \mathbb{R}$, $a < b$. Računamo

$$\begin{aligned} F(b) &= P(X \le b) = P((X \le a) \cup (a < X \le b)) \\ &= P(X \le a) + P(a < X \le b)) = F(a) + P(a < X \le b), \end{aligned}$$

otkuda slijedi $P(a < X \le b) = F(b) - F(a)$
(F6) Neka su $a, b \in \mathbb{R}$, $a < b$. Iz svojstva (F5) slijedi da je $F(b) \le F(a)$, pa je funkcija F rastuća.

PRIMJER 5.8 *Za funkciju distribucije diskretne slučajne varijable* $F(x) = P(X \le x)$ *vrijede i slijedeće relacije:*

(a) $P(a < X < b) = F(b-) - F(a),\ a, b \in \mathbb{R},\ a < b,$

(b) $P(a \le X \le b) = F(b) - F(a-),\ a, b \in \mathbb{R},\ a < b,$

(c) $P(a \le X < b) = F(b-) - F(a-),\ a, b \in \mathbb{R},\ a < b.$

Rješenje:

$$\begin{aligned} (a)\quad F(b-) &= P(X < b) = P((X \le a) \cup (a < X < b)) \\ &= P(X \le a) + P(a < X < b) = F(a) + P(a < X < b). \end{aligned}$$

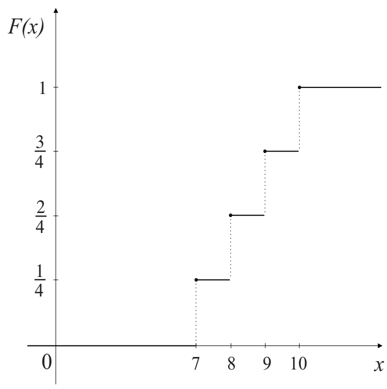
$$\begin{aligned} (b)\quad F(b) &= P(X \le b) = P((X < a) \cup (a \le X \le b)) \\ &= P(X < a) + P(a \le X \le b) = F(a-) + P(a \le X \le b). \end{aligned}$$

$$\begin{aligned} (c)\quad F(b-) &= P(X < b) = P((X < a) \cup (a \le X < b)) \\ &= P(X < a) + P(a \le X < b) = F(a-) + P(a \le X < b). \end{aligned}$$

PRIMJER 5.9 *Za slučajnu varijablu X iz primjera zadanu s*

$$X \sim \begin{pmatrix} 7 & 8 & 9 & 10 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

napišite funkciju distribucije. Izračunajte vjerojatnost da slučajna varijabla poprimi vrijednosti veće od 8 i manje ili jednako 10?



Slika 5.2: Graf funkcije distribucije $F(x)$ diskretne slučajne varijable X iz primjera (5.9).

Rješenje:

$$F(x) := P(X \leq x) = \sum_{i:x_i \leq x} f(x_i) = \begin{cases} 0, & x < 7 \\ \frac{1}{4}, & 7 \leq x < 8 \\ \frac{2}{4}, & 8 \leq x < 9 \\ \frac{3}{4}, & 9 \leq x < 10 \\ 1 & 10 \leq x. \end{cases}$$

Koristimo formulu

$$P(a < X \leq b) = F(b) - F(a) = \sum_{i:a<x_i \leq b} f(x_i),$$

$$P(8 < X \leq 10) = \sum_{i:8<x_i \leq 10} f(x_i) = f(9) + f(10) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2},$$

ili odmah uvrstimo vrijednosti $F(b), F(a)$:

$$P(8 < X \leq 10) = F(10) - F(8) = 1 - \frac{2}{4} = \frac{1}{2}.$$

PRIMJER 5.10 *Za slučajnu varijablu X iz primjera zadanu s*

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

napišite funkciju distribucije. Izračunajte

(a) $P(2 < X \leq 5)$

(b) $P(2 < X < 5)$

Rješenje:

$$F(x) := P(X \leq x) = \sum_{i:x_i \leq x} f(x_i) = \begin{cases} 0, & x < 1 \\ \frac{1}{6}, & 1 \leq x < 2 \\ \frac{2}{6}, & 2 \leq x < 3 \\ \frac{3}{6}, & 3 \leq x < 4 \\ \frac{4}{6}, & 4 \leq x < 5 \\ \frac{5}{6}, & 5 \leq x < 6 \\ 1, & 6 \leq x. \end{cases}$$

(a) Koristimo formulu

$$P(a < X \leq b) = F(b) - F(a) = \sum_{i:a<x_i \leq b} f(x_i),$$

$$P(2 < X \leq 5) = \sum_{i:2<x_i \leq 5} f(x_i) = f(3) + f(4) + f(5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6},$$

ili odmah uvrstimo vrijednosti $F(b), F(a)$:

$$P(2 < X \leq 5) = F(5) - F(2) = \frac{5}{6} - \frac{2}{6} = \frac{3}{6}.$$

(b) Koristimo formulu

$$P(a < X < b) = F(b-) - F(a) = \sum_{i:a<x_i<b} f(x_i),$$

$$P(2 < X < 5) = \sum_{i:2<x_i<5} f(x_i) = f(3) + f(4) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6},$$

ili odmah uvrstimo vrijednosti $F(b-), F(a)$:

$$P(2 < X < 5) = F(5-) - F(2) = \frac{4}{6} - \frac{2}{6} = \frac{2}{6}.$$

Prisjetimo se motivacijskog primjera:

PRIMJER 5.11 *motiv*

Promatramo slučajan pokus bacanje 2 igraće kocke i slučajnu varijablu X = suma brojeva koji su pali. Nađite funkciju vjerojatnosti i funkciju distribucije slučajne varijable X. Izračunajte vjerojatnost da je zbroj brojeva koji su pali veći od 3 a manji ili jednak 6?

Rješenje:

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

Funkcija distribucije slučajne varijable X je

$$F(x) := \begin{cases} 0, & x < 2 \\ \frac{1}{36}, & 2 \leq x < 3 \\ \frac{3}{36}, & 3 \leq x < 4 \\ \frac{6}{36}, & 4 \leq x < 5 \\ \frac{10}{36}, & 5 \leq x < 6 \\ \frac{15}{36}, & 6 \leq x < 7 \\ \frac{21}{36}, & 7 \leq x < 8 \\ \frac{26}{36}, & 8 \leq x < 9 \\ \frac{30}{36}, & 9 \leq x < 10 \\ \frac{33}{36}, & 10 \leq x < 11 \\ \frac{35}{36}, & 11 \leq x < 12 \\ 1, & 12 \leq x. \end{cases}$$

Koristimo formulu $P(a < X \leq b) = \sum\limits_{i:a<x_i\leq b} f(x_i)$,

$$P(3 < X \leq 6) = \sum_{i:3<x_i\leq 6} f(x_i) = f(4) + f(5) + f(6) = \frac{3}{36} + \frac{4}{36} + \frac{5}{36} = \frac{12}{36},$$

ili odmah uvrstimo vrijednosti $F(b), F(a)$:

$$P(3 < X \leq 6) = F(6) - F(3) = \frac{15}{36} - \frac{3}{36} = \frac{12}{36}.$$

PRIMJER 5.12 *Promatramo slučajan pokus gađanja u metu 3 puta. U svakom gađanju vjerojatnost pogotka mete je $\frac{1}{2}$. Neka je slučajna varijabla X = broj pogodaka u metu. Nađite funkciju vjerojatnosti i funkciju distribucije slučajne varijable X.*

Rješenje: Slučajnu varijablu X možemo zadati s $X \sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{pmatrix}$.

Funkcija distribucije slučajne varijable X je

$$F(x) := \begin{cases} 0, & x < 0 \\ \frac{1}{8}, & 0 \leq x < 1 \\ \frac{4}{8}, & 1 \leq x < 2 \\ \frac{7}{8}, & 2 \leq x < 3 \\ 1, & 3 \leq x. \end{cases}$$

Definicija 5.5 *(OČEKIVANJE DISKRETNE SLUČAJNE VARIJABLE, engl. mean ili mathematical expectation)*

Neka je $X : \Omega \to \mathbb{R}$ *diskretna slučajna varijabla sa slikom* $\mathcal{R}(X) = \{x_1, x_2, ...\}$ *i funkcijom vjerojatnosti* $f : \mathbb{R} \to [0,1]$

$$f(x) := \begin{cases} P(X = x_i), & x = x_i \in \mathcal{R}(X) \\ 0, & \text{inače}. \end{cases}$$

Kažemo da diskretna slučajna varijabla X ima očekivanje ako red $\sum_{i=1}^{\infty} x_i f(x_i)$ *konvergira i označavamo*

$$E(X) := \sum_{i=1}^{\infty} x_i f(x_i).$$

Ako je X diskretna slučajna varijabla s konačnom slikom $\mathcal{R}(X)$ *onda ona ima očekivanje*

$$E(X) := \sum_{i=1}^{n} x_i f(x_i).$$

Očekivanje slučajne varijable predstavlja očekivanu vrijednost slučajne varijable.

Prisjetimo se motivacijskog primjera:

PRIMJER 5.13 *motiv*

U agenciji za nekretnine prema pokazateljima prodaje u prošlom razdoblju zarada po prodanom jednosobnom stanu je 3000 eura. Kolika je očekivana tjedna zarada na jednosobnim stanovima ako je prema zadnjim pokazateljima vjerojatnost prodaje nijednog stana 0.1, 1 stana 0.3, 2 stana 0.4, a vjerojatnost prodaje 3 stana 0.2 .

Rješenje: Slučajnu varijablu X = broj prodanih jednosobnih stanova, možemo zadati s $X \sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{10} & \frac{3}{10} & \frac{4}{10} & \frac{2}{10} \end{pmatrix}$.

$$\begin{aligned} E(X) &= \sum_{i=1} x_i f(x_i) \\ &= 0 \cdot \frac{1}{10} + 1 \cdot \frac{3}{10} + 2 \cdot \frac{4}{10} + 3 \cdot \frac{2}{10} \\ &= 1.7. \end{aligned}$$

Očekivani broj prodanih jednosobnih stanova je 1.7. Očekivana tjedna zarada je 5100 eura.

NAPOMENA 5.2 *PRAVEDNA IGRA*

U teoriji igara matematičko očekivanje nekog klađenja je $E(X) = p \cdot a$, *gdje je* $a =$ *očekivani dobitak kad se dogodi događaj s vjerojatnošću* p. *Ako je* $E(X) < 0$ *onda je igrač na gubitku, ako* $E(X) > 0$ *onda je na dobitku, a ako je* $E(X) = 0$ *onda je igra pravedna. Ako je uložen novac u iznosu ulog, onda je ukupno matematičko očekivanje ulog* $\cdot\, E(X)$.

PRIMJER 5.14 *(a) U igri s novčićem prvi igrač se kladi u 1 kunu. Kad prvi igrač izgubi plati kunu protivniku, a kad protivnik izgubi plati 1 kunu prvom igraču. Izračunajte matematičko očekivanje? Ako je prvi igrač uložio 1000 kuna koliki je očekivani dobitak ili gubitak?*

(b) U igri s novčićem prvi igrač se kladi u 1 kunu. Kad prvi igrač izgubi plati kunu protivniku, a kad protivnik izgubi plati 90 lipa prvom igraču. Izračunajte matematičko očekivanje? Ako je prvi igrač uložio 1000 kuna koliki je očekivani dobitak ili gubitak?

Rješenje:

(a) Matematičko očekivanje je $E(X) = 1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0$. Igra je pravedna. Budući je prvi igrač uložio 1000 kn njegov dobitak (gubitak) je $1000 \cdot 0 = 0$ tj. nema niti gubitka niti dobitka jer je očekivani iznos jedank ulogu.

(b)Matematičko očekivanje je $E(X) = 0.90 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = -0.05$. Igrač je na gubitku jer je $E(X) < 0$. Budući je prvi igrač uložio 1000 kn njegov očekivani gubitak je $1000 \cdot (-0.05) = -50$ kn.

Definicija 5.6 *(VARIJANCA ILI DISPERZIJA DIS. SL. VARIJABLE, engl. variance)*

Neka $X : \Omega \to \mathbb{R}$ *diskretna slučajna varijabla sa slikom* $\mathcal{R}(X) = \{x_1, x_2, ...\}$ *i funkcijom vjerojatnosti* $f : \mathbb{R} \to [0, 1]$ *ima očekivanje* $E(X)$. *Kažemo da diskretna slučajna varijabla* X *ima varijancu ako red* $\sum_{i=1}^{\infty}(x_i - E(X))^2 f(x_i)$ *konvergira i označavamo*

$$Var(X) := \sum_{i=1}^{\infty}(x_i - E(X))^2 f(x_i).$$

Ako je X *diskretna slučajna varijabla s konačnom slikom* $\mathcal{R}(X)$ *onda ona ima varijancu*

$$Var(X) := \sum_{i=1}^{n}(x_i - E(X))^2 f(x_i).$$

Varijanca predstavlja srednje kvadratno odstupanje od očekivane vrijednosti slučajne varijable.

Varijancu možemo računati i pomoću relacije

$$Var(X) := \sum_{i=1}^{n} x_i^2 \cdot f(x_i) - (E(X))^2.$$

(Dokaz vidi kasnije - očekivanje funkcije slučajne varijable).

Definicija 5.7 *(STANDARDNA DEVIJACIJA, engl. standard deviation)*
Standardna devijacija slučajne varijable X definira se kao

$$\sigma(X) := \sqrt{Var(X)}.$$

PRIMJER 5.15 *Promatramo slučajan pokus bacanje 2 igraće kocke i slučajnu varijablu X = suma brojeva koji su pali. Nađite očekivanje i varijancu diskretne slučajne varijable X.*

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

$$\begin{aligned} E(X) &= \sum_{i=1} x_i f(x_i) \\ &= 2\cdot\frac{1}{36} + 3\cdot\frac{2}{36} + 4\cdot\frac{3}{36} + 5\cdot\frac{4}{36} + 6\cdot\frac{5}{36} + 7\cdot\frac{6}{36} + 8\cdot\frac{5}{36} + 9\cdot\frac{4}{36} \\ &\quad +10\cdot\frac{3}{36} + 11\cdot\frac{2}{36} + 12\cdot\frac{1}{36} \\ &= 7. \end{aligned}$$

$$\begin{aligned} Var(X) &= \sum_{i}^{n} x_i^2 \cdot f(x_i) - (E(X))^2 \\ &= 2^2\cdot\frac{1}{36} + 3^2\cdot\frac{2}{36} + 4^2\cdot\frac{3}{36} + 5^2\cdot\frac{4}{36} + 6^2\cdot\frac{5}{36} + 7^2\cdot\frac{6}{36} + 8^2\cdot\frac{5}{36} \\ &\quad +9^2\cdot\frac{4}{36} + 10^2\cdot\frac{3}{36} + 11^2\cdot\frac{2}{36} + 12^2\cdot\frac{1}{36} - 7^2 \\ &= 5.83. \end{aligned}$$

PRIMJER 5.16 *Promatramo slučajan pokus gađanja u metu 3 puta. U svakom gađanju vjerojatnost pogotka mete je $\frac{1}{2}$. Neka je slučajna varijabla X = broj pogodaka u metu. Izračunajte očekivanje i varijancu.*

Rješenje: Slučajna varijabla X je $X \sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{pmatrix}$.

$E(X) = \sum_{i=1}^{n} x_i f(x_i) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2}.$

$$\begin{aligned} Var(X) &= \sum_{i=1}^{n} x_i^2 \cdot f(x_i) - (E(X))^2 = 0^2 \cdot \frac{1}{8} + 1^2 \cdot \frac{3}{8} + 2^2 \cdot \frac{3}{8} + 3^2 \cdot \frac{1}{8} - (\frac{3}{2})^2 \\ &= \frac{3}{4}. \end{aligned}$$

5.2 Ponovimo

DISKRETNA SLUČAJNA VARIJABLA

diskretna slučajna varijabla	$X : \Omega \to \mathbb{R}$
slika diskretne slučajne varijable	$\mathcal{R}(X) = \{x_1, x_2, ..., x_n, ...\}$
vjerojatnost da sl. var. poprimi vrijednost x_i	$P(X = x_i) = P(\{\omega : X(\omega) = x_i\})$
funkcija vjerojatnosti dis. sl. var. X	$f(x) = P(X = x_i),\ x = x_i \in \mathcal{R}(X)$
funkcija distribucije dis. sl. var. X	$F(x) := P(X \leq x) = \sum_{i: x_i \leq x} f(x_i)$
očekivanje dis. sl. var. X	$E(X) := \sum_{i=1}^{\infty} x_i f(x_i)$
varijanca dis. sl. var. X	$Var(X) := \sum_{i=1}^{\infty} (x_i - E(X))^2 f(x_i)$

Poglavlje 6

PRIMJERI DISKRETNIH SLUČAJNIH VARIJABLI

U ovom poglavlju istaknut ćemo diskretne slučajne varijable koje se pojavljuju u određenim "scenarijima" i njihove funkcije vjerojatnosti definirat će se kao specifične distribucije (razdiobe): binomna, Poissonova, hipergeometrijska, geometrijska. Binomna distribucija vezana je uz scenario Bernoullijeve sheme, Poissonova je granični slučaj binomne, hipergeometrijska distribucija odgovara uzorku bez vraćanja, a scenario u kome se broje pokušaji dok se ne dogodi željeni događaj prati geometrijska distribucija.

6.1 Bernoullijeva shema. Binomna distribucija (razdioba)

MOTIV 6.1 *U velikoj kutiji nalazi se* $p = 10\%$ *neispravnih proizvoda. Ako uzmemo uzorak od* $r = 5$ *proizvoda iz kutije (s vraćanjem), slučajna varijabla* X = *broj neispravnih proizvoda u uzorku ima binomnu distribuciju* $X \sim B(r, p)$. *Izračunajte vjerojatnost*
(a) da se ne pojavi niti jedan neispravni proizvod,
(b) da se pojavi jedan neispravni proizvod u uzorku,
(c) da se pojavi bar jedan neispravni proizvod u uzorku.

Jacob Bernoulli (1654-1705) je bio jedan od najpoznatih članova švicarske matematičke obitelji Bernoulli.

Definicija 6.1 *(Bernoullijeva shema)*

U m nezavisnih pokusa, vjerojatnost da se dogodi događaj A u svakom od njih je jednaka i $P(A) = p$. Takvu shemu događaja zovemo Bernoullijeva shema. (Pretpostavljamo da su pokusi nezavisni, tj. da vjerojatnost pojave događaja A u svakom od pokusa ne ovisi od toga da li se on dogodio ili nije u nekom drugom pokusu.)

Definicija 6.2 *(Binomna slučajna varijabla)*

Slučajnu varijablu X = broj pojavljivanja događaja A u $m \in \mathbb{N}$ pokusa u Bernoullijevoj shemi s vjerojatnošću p, $p \in (0,1)$ zovemo Binomna slučajna varijabla. Specijalno, slučaju samo jednog pokusa $m = 1$ slučajnu varijablu zovemo Bernoullijeva. Slika Binomne slučajne varijable je $\mathcal{R}(X) = \{0, 1, ..., m\}$. Funkcija vjerojatnosti Binomne slučajne varijable je

$$f(x) := \begin{cases} P(X = k) = \binom{m}{k} p^k \cdot (1-p)^{m-k}, & x = k \in \mathcal{R}(X) \\ 0, & \textit{inače}. \end{cases}$$

Funkcija distribucije Binomne slučajne varijable je

$$F(x) := \begin{cases} 0, & x < 0 \\ \sum\limits_{k:k \le x} \binom{m}{k} p^k \cdot (1-p)^{m-k}, & 0 \le x < m \\ 1, & m \le x. \end{cases}$$

Definicija 6.3 *(BINOMNA DISTRIBUCIJA)*

Za sve slučajne varijable koje imaju sliku $\mathcal{R}(X) = \{0, 1, 2, ..., m\}$ i funkciju vjerojatnosti

$$f(x) = \binom{m}{x} p^x (1-p)^{m-x},\ x \in \mathcal{R}(X)$$

kažemo da imaju BINOMNU DISTRIBUCIJU (RAZDIOBU) s parametrima m i p i označavamo

$$X \sim B(m, p).$$

PRIMJER 6.1 *Funkcija $f(x) = \binom{m}{x} p^x (1-p)^{m-x}$ je funkcija vjerojatnosti. Prisjetimo se Binomnog teorema i iskoristimo za računanje*

$$\begin{aligned} \sum_{0 \le x \le m} f(x) &= \sum_{0 \le x \le m} \binom{m}{x} p^x \cdot (1-p)^{m-x} = \sum_{0 \le k \le m} \binom{m}{k} p^k \cdot (1-p)^{m-k} \\ &= (p + (1-p))^m = 1. \end{aligned}$$

PRIMJER 6.2 *Očekivanje diskretne slučajne varijable koja ima binomnu distribuciju $X \sim B(m, p)$ je*

$$E(X) = m \cdot p$$

Rješenje:

$$\begin{aligned} E(X) &= \sum_{0\le x\le m} x\cdot f(x) = \sum_{x=0}^{m} x\binom{m}{x}p^x\cdot(1-p)^{m-x} \\ &= \sum_{k=1}^{m} k\cdot\binom{m}{k}p^k\cdot(1-p)^{m-k} \\ &= mp\sum_{k=1}^{m}\binom{m-1}{k-1}p^{k-1}\cdot(1-p)^{(m-1)-(k-1)} \\ &= mp\cdot(1-(1-p))^{m-1} = mp\cdot 1 = mp. \end{aligned}$$

PRIMJER 6.3 *Varijanca diskretne slučajne varijable koja ima binomnu distribuciju $X\sim B(m,p)$ je*

$$Var(X) = m\cdot p\cdot(1-p)$$

Rješenje: (pokušajte sami)

NAPOMENA 6.1 *Ako je $p = 1/2$ binomna distribucija je simetrična.*

PRIMJER 6.4 *Promatramo slučajan pokus gađanja u metu 3 puta. U svakom gađanju vjerojatnost pogotka mete je $\frac{1}{2}$. Neka je slučajna varijabla X = broj pogodaka u metu. Nađite funkciju vjerojatnosti i funkciju distribucije slučajne varijable X, očekivanje i varijancu.*

Rješenje:
To je Bernoullijeva shema događaja s $m = 3$ nezavisna pokusa (ili ponavljamo isti pokus 3 puta nezavisno). Promatramo događ aj A = pogodak u metu.

$P(A) = \frac{1}{2} = p$ u svakom nezavisnom ponavljanju.

Slučajna varijabla
X = broj pojavljivanja događaja A (broj uspjeha) ima binomnu distribuciju $X\sim B(m,p) = B(3,\frac{1}{2})$.
Funkcija vjerojatnosti je

$$\begin{aligned} f(x) &:= \begin{cases} P(X=k) = \binom{3}{k}(\frac{1}{2})^k\cdot(1-\frac{1}{2})^{3-k}, & x=k\in\{0,1,2,3\} \\ 0, & \text{inače.} \end{cases} \\ &= \begin{cases} \binom{3}{k}(\frac{1}{2})^3, & x=k\in\{0,1,2,3\} \\ 0, & \text{inače.} \end{cases} \end{aligned}$$

Funkcija distribucije je:

$$F(x) \quad := \quad \begin{cases} 0, & x < 0 \\ \sum\limits_{k:k\leq x} \binom{3}{k}(\dfrac{1}{2})^k \cdot (1-\dfrac{1}{2})^{3-k}, & 0 \leq x < 3 \\ 1, & 3 \leq x. \end{cases}$$

$$= \quad \begin{cases} 0, & x < 0 \\ \sum\limits_{k:k\leq x} \binom{3}{k}(\dfrac{1}{2})^3, & 0 \leq x < 3 \\ 1, & 3 \leq x. \end{cases}$$

$$E(X) = mp = 3 \cdot \frac{1}{2} = \frac{3}{2}, \quad Var(X) = mp(1-p) = 3 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}.$$

(Do ovih rezultata došli smo i kad smo razmatrali taj primjer slučajne varijable u prethodnom poglavlju.)

PRIMJER 6.5 *Bacamo kocku 4 puta. Neka je slučajna varijabla X = broj šestica. Da li je to slučajna varijabla koja ima binomnu distribuciju? Nađite funkciju vjerojatnosti i funkciju distribucije slučajne varijable X, očekivanje i varijancu. Izračunajte vjerojatnost da su pale bar dvije šestice?*

Rješenje:
To je Bernoullijeva shema događaja s $m = 4$ nezavisna pokusa (ili ponavljamo isti pokus 4 puta nezavisno). Promatramo događaj A = "pala je 6". $P(A) = \frac{1}{6} = p$ u svakom nezavisnom ponavljanju. Slučajna varijabla X = broj šestica je broj pojavljivanja A (broj uspjeha) i ona ima binomnu distribuciju $X \sim B(m, p) = B(4, \frac{1}{6})$. Funkcija vjerojatnosti je

$$f(x) = \begin{cases} \binom{4}{k}(\frac{1}{6})^k(\frac{5}{6})^{4-k}, & x = k \in \{0, 1, 2, 3, 4\} \\ 0, & \text{inače.} \end{cases}$$

Funkcija distribucije je:

$$F(x) = \begin{cases} 0, & x < 0 \\ \sum\limits_{k:k\leq x} \binom{4}{k}(\dfrac{1}{6})^k(\dfrac{5}{6})^{4-k}, & 0 \leq x < 4 \\ 1, & 4 \leq x. \end{cases}$$

$$E(X) = mp = 4 \cdot \frac{1}{6} = \frac{4}{6}, \quad Var(X) = mp(1-p) = 4 \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{20}{36}.$$

Trebamo odrediti $P(X \geq 2)$.

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) = 1 - F(2-) = 1 - \sum_{i:x_i<2} f(x_i) \\ &= 1 - (f(0) + f(1)) = 1 - (\tbinom{4}{0}(\frac{1}{6})^0(\frac{5}{6})^4 + \tbinom{4}{1}(\frac{1}{6})^1(\frac{5}{6})^{4-1}) \\ &= \frac{171}{1296} \end{aligned}$$

PRIMJER 6.6 *tko želi znati više*

Označimo vrijednost funkcije vjerojatnosti f binomne distribucije s parametrima m i p s $f(m,p;k), k \in \{0,1,...,m\}$. Tada vrijedi rekurzivna relacija

$$f(m,p;k+1) = \frac{m-k}{k+1} \cdot \frac{p}{1-p} f(m,p;k)$$

Dokaz: Računamo

$$\frac{f(m,p;k+1)}{f(m,p;k)} = \frac{n! \cdot p^{k+1}(1-p)^{m-k-1}}{(k+1)!(m-k-1)!} : \frac{n! \cdot p^k(1-p)^{m-k}}{k!(m-k)!} = \frac{m-k}{k+1} \cdot \frac{p}{1-p}$$

NAPOMENA 6.2 *Prisjetimo se primjera za uzorak s vraćanjem.*
Slučajni pokus: izbor od r elemenata (s vraćanjem) iz skupa koji ima n elemenata $n = n_T + n_D$, $r = r_T + r_D$.
Ω *=svi uzorci veličine r iz n-čl. skupa, važan poredak*
$|\Omega| = \overline{V}_n^{(r)} = n^r$;
Događaj A="uzorak ima r_T ispravnih i r_D neispravnih";
$|A|$ *=Broj svih uzoraka veličine r iz n -čl. skupa s vraćanjem koji imaju r_T ispravnih i r_D neispravnih:*
Koristimo formulu za uzorak s vraćanjem $|A| = \binom{r}{r_T} \cdot (n_T)^{r_T} \cdot (n - n_T)^{(r-r_T)}$.

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\binom{r}{r_T} \cdot (n_T)^{r_T} \cdot (n-n_T)^{r-r_T}}{n^r} = \binom{r}{r_T} \cdot (\frac{n_T}{n})^{r_T} \cdot (1 - (\frac{n_T}{n}))^{(r-r_T)}.$$

Promatramo slučajnu varijablu
X = broj ispravnih predmeta u r članom uzorku iz skupa koji ima n elemenata, od toga n_T ispravnih
Ako s $p = \frac{n_T}{n}$ označimo postotak točnih (ispravnih) proizvoda ukupno, onda slučajna varijabla X = broj ispravnih predmeta u r članom uzorku iz skupa koji ima postotak p ispravnih elemenata ima binomnu distribuciju s parametrima r i p, $X \sim B(r,p)$. (Analogno za neispravne elemente)

Prisjetimo se motivacijskog primjera:

PRIMJER 6.7 *motiv*

U velikoj kutiji nalazi se $p = 10\%$ neispravnih proizvoda. Ako uzmemo uzorak od $r = 5$ proizvoda iz kutije (s vraćanjem), slučajna varijabla X = broj neispravnih proizvoda u uzorku ima binomnu distribuciju $X \sim B(r, p)$. Izračunajte vjerojatnost
(a) da se ne pojavi niti jedan neispravni proizvod,
(b) da se pojavi jedan neispravni proizvod u uzorku,
(c) da se pojavi bar jedan neispravni proizvod u uzorku.

Rješenje:
(a) $P(X = 0) = f(0) = f(5, \frac{1}{10}; 0) = \binom{5}{0}(\frac{1}{10})^0(1 - \frac{1}{10})^5 = 0.59049.$
(b) Možemo računati direktno

$$P(X = 1) = f(1) = f(5, \frac{1}{10}; 1) = \binom{5}{1}(\frac{1}{10})(1 - \frac{1}{10})^4 = \frac{5}{10}0.6561 = 0.32805.$$

Drugi način je da koristimo rekurzivnu formulu:*tko želi znati više*

$$f(r, p; k + 1) = \frac{r - k}{k + 1} \cdot \frac{p}{1 - p} f(r, p; k)$$

$$P(X = 1) = f(r, p; 1) = \frac{r - 0}{0 + 1} \cdot \frac{p}{1 - p} f(r, p; 0) = r\frac{p}{1 - p} f(r, p; 0).$$

Za $r = 5$ i $p = \frac{1}{10}$

$$P(X = 1) = f(5, \frac{1}{10}; 1) = 5\,\frac{1}{9}\,f(5, \frac{1}{10}; 0) = \frac{5}{9}0.59049 = 0.32805.$$

(c)

$$\begin{aligned} P(X \geq 1) &= 1 - P(X < 1) = 1 - P(X = 0) = 1 - f(0) = 1 - (1 - p)^m \\ &= 1 - (1 - \frac{10}{100})^5 = 1 - (\frac{9}{10})^5 = 1 - 0.59049 = 0.4095. \end{aligned}$$

NAPOMENA 6.3 *Ako $m \to \infty$ binomna distribucija teži normalnoj distribuciji (vidi kasnije - Moivre Laplace teorem).*

NAPOMENA 6.4 *Ako $p \to 0$ i $m \to \infty$ binomna distribucija teži Poissonovoj (vidi Poissonova distribucija).*

6.2 POISSONOVA DISTRIBUCIJA

MOTIV 6.2 *Kroz naplatnu kućicu prođu prosječno u minuti 2 automobila. Kolika je vjerojatnost da će u toku bilo koje minute proći*
(a) jedan automobil,
(b) barem 3 automobila?

Francuski matemati čar Simeon D. Poisson (1781-1840) je 1838 objavio rad u kojem je promatrao slučajne varijable koje broje dolaske na neko mjesto u vremenskom intervalu određene duljine.

Definicija 6.4 *(POISSONOVA DISTRIBUCIJA)*

Za diskretnu slučajnu varijablu X koja ima sliku $\mathcal{R}(X) = \{0, 1, 2, ..\}$ i funkciju vjerojatnosti

$$f(x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}, \;\; x \in \mathcal{R}(X)$$

kažemo da ima Poissonovu distribuciju s parametrom λ i označavamo

$$X \sim Po(\lambda).$$

PRIMJER 6.8 *Funkcija $f(x) = \dfrac{\lambda^x}{x!} \cdot e^{-\lambda}$ je funkcija vjerojatnosti. Prisjetimo se razvoja eksponencijalne funkcije u McLaurinov red $e^x = \sum\limits_{0 \le x \le \infty} \dfrac{x^k}{k!}$.*
Računamo

$$\sum_{0 \le k \le \infty} f(k) = \sum_{0 \le k \le \infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{\lambda} \cdot e^{-\lambda} = 1.$$

PRIMJER 6.9 *Očekivanje diskretne slučajne varijable koja ima Poissonovu distribuciju $X \sim Po(\lambda)$ je*

$$E(X) = \lambda.$$

Rješenje:

$$E(X) = \sum_{0 \le k \le \infty} k \cdot f(k) = \sum_{x=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda.$$

PRIMJER 6.10 *Varijanca diskretne slučajne varijable koja ima Poissonovu distribuciju $X \sim Po(\lambda)$ je*

$$Var(X) = \lambda.$$

Rješenje: (Pokušajte sami).

PRIMJER 6.11 *Ako u Bernoullijevoj shemi broj m nezavisnih pokusa teži ∞ (jako velik) a vjerojatnost događaja A kojeg promatramo u svakom pokusu teži 0 (jako mala) onda slučajna varijabla X = broj pojavljivanja događaja A ima binomnu distribuciju koju možemo dobro aproksimirati s Poissonovom distribucijom s parametrom $\lambda = mp$, tj. $X \sim Po(mp)$.*

Dokaz: *tko želi znati višr*

Neka je $X \sim B(m,p)$, s parametrima $m \to \infty,\ p \to 0,\ m \cdot p \to \lambda$.
Računamo graničnu vrijednost funkcije vjerojatnosti binomne distribucije:

$$\begin{aligned}
\lim f(k) &= \lim_{m\to\infty,\ p\to 0,\ m\cdot p\to\lambda} \binom{m}{k} p^k \cdot (1-p)^{m-k} \\
&= \lim_{p\to\frac{\lambda}{m}} \binom{m}{k} (\frac{\lambda}{m})^k \cdot (1-\frac{\lambda}{m})^{m-k} \\
&= \frac{(\lambda)^k}{k!} \lim_{m\to\infty} (m(m-1)\cdot(m-k+1))\frac{1}{m^k}(1-\frac{\lambda}{m})^{m-k} \\
&= \frac{(\lambda)^k}{k!} \cdot 1 \cdot e^{-\lambda} \\
&= \frac{(\lambda)^k}{k!} \cdot e^{-\lambda}.
\end{aligned}$$

Dobili smo funkciju vjerojatnosti Poissonove distribucije za $\lambda = mp$.

PRIMJER 6.12 *U velikoj kutiji nalazi se $p = 1\%$ neispravnih proizvoda. Ako uzmemo uzorak od $r = 100$ proizvoda iz kutije, slučajna varijabla X = broj neispravnih proizvoda u uzorku ima binomnu distribuciju $X \sim B(r,p)$. Budući je r veliki u odnosu na mali p i $r \cdot p = 1$ slučajnu varijablu X možemo aproksimirati Poissonovom distribucijom $X \sim Po(rp)$. Izračunajte vjerojatnost da se pojavi bar jedan neispravni proizvod u uzorku.*

Rješenje:
Računamo po Poissonovoj distribuciji: $X \sim Po(\lambda),\ \lambda = r \cdot p = 1$,

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - f(0) = 1 - \frac{\lambda^0}{0!} \cdot e^{-\lambda} = 0.6321$$

Ako bi računali po binomnoj distribuciji: $X \sim B(r,p),\ r = 100,\ p = \frac{1}{10}$,

$$\begin{aligned}
P(X \geq 1) &= 1 - P(X < 1) = 1 - P(X = 0) = 1 - f(0) \\
&= 1 - \binom{100}{0} (99/100)^{100} = 1 - 0.3660 = 0.63396
\end{aligned}$$

Usporedimo li rezultate dobivene vidimo da greška nije velika.

PRIMJER 6.13 *Promatramo broj automobila koji prođu kroz kontrolnu točku u intervalima od 1 minute. Neka je p konstantna vjerojatnost da će auto proći u svakom kratkom intervalu unutar 1 minute (i pretpostavimo nezavisnost o drugim događajima u tom kratkom vremenu). Slučajna varijabla* X = *broj automobila koji prođu kroz kontrolnu točku ima Poissonovu distribuciju s parametrom* $\lambda = 60 \cdot \frac{p}{60} = p$ = *prosječan broj automobila koji prođu kontrolnu točku u 1 minuti,* $X \sim Po(\lambda)$.

Prisjetimo se motivacijskog primjera:

PRIMJER 6.14 *motiv*

Kroz naplatnu kućicu prođu prosječno u minuti 2 automobila. Kolika je vjerojatnost da će u toku bilo koje minute proći
(a) jedan automobil,
(b) barem 3 automobila?

Rješenje: Broj automobila koji prođu naplatnu kućicu u jednoj minuti ima Poissonovu distribuciju $X \sim Po(\lambda)$, $\lambda = 2$
(a) $P(X = 1) = f(1) = \frac{\lambda^1}{1!} \cdot e^{-\lambda} = 2 \cdot e^{-2} = 0.27067$
(b)

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) = 1 - (f(0) + f(1) + f(2)) \\ &= 1 - (e^{-2}(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!})) = 1 - (0.6766) = 0.3232. \end{aligned}$$

NAPOMENA 6.5 *Poissonova distribucija primjenjuje se u demografiji, biologiji, fizici, prometu i dr.*

6.3 HIPERGEOMETRIJSKA DISTRIBUCIJA

MOTIV 6.3 *U velikom pakovanju od 100 cigli distributer garantira da je najviše $p = 10\%$ oštećenih. Kontrolor prihvaća pošiljku samo ako u uzorku od 10 proizvoda iz pakovanja (bez vraćanja) nema oštećenih. Kolika je vjerojatnost da će kontrolor odbiti pošiljku?*

NAPOMENA 6.6 *Prisjetimo se primjera za uzorak bez vraćanja. Slučajni pokus: izbor od r elemenata (bez vraćanja) iz skupa koji ima n elemenata $n = n_T + n_D$, $r = r_T + r_D$.*
Ω ={svi uzorci veličine r iz n-čl. skupa}, $|\Omega| = \binom{n}{r}$;
Događaj A="uzorak ima r_T ispravnih i r_D neispravnih";
$|A|$ =Broj svih uzoraka veličine r iz n -čl. skupa bez vraćanja koji imaju r_T ispravnih i r_D neispravnih.
Koristimo formulu za uzorak bez vraćanja:

$$|A| = C_{n_T}^{(r_T)} \cdot C_{n_D}^{(r_D)} = \binom{n_T}{r_T} \cdot \binom{n_D}{r_D} = \binom{n_T}{r_T} \cdot \binom{n - n_T}{r - r_T}.$$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\binom{n_T}{r_T} \cdot \binom{n-n_T}{r-r_T}}{\binom{n}{r}}.$$

Promatramo slučajnu varijablu X = broj ispravnih predmeta u r-članom uzorku iz skupa koji ima n elemenata od toga n_T ispravnih.

Definicija 6.5 *(HIPERGEOMETRIJSKA DISTRIBUCIJA)*
Slučajne varijable X koje imaju sliku $\mathcal{R}(X) = \{0, 1, 2, .., n_T\}$ i funkciju vjerojatnosti

$$f(x) = \frac{\binom{n_T}{x} \cdot \binom{n-n_T}{r-x}}{\binom{n}{r}}, \quad x \in \mathcal{R}(X)$$

gdje su $n_T \leq n$, $r \leq n$, n, n_T, $r \in \mathbb{N}$ kažemo da imaju Hipergeometrijsku distribuciju s parametrima n, r, n_T i označavamo $X \sim Hip\,(n, r, n_T)$.

PRIMJER 6.15 *Funkcija $f(x) = \dfrac{\binom{n_T}{x} \cdot \binom{n-n_T}{r-x}}{\binom{n}{r}}$, $x \in \mathcal{R}(X)$ je funkcija vjerojatnosti.*

Rješenje: Koristimo kombinatorni identitet

$$\sum_{0 \leq x \leq n_T} \binom{n_T}{x} \cdot \binom{n - n_T}{r - x} = \binom{n}{r}$$

PRIMJER 6.16 *Slučajnu varijablu $X \sim Hip(n, r, n_T)$ ima očekivanje*

$$E(X) = r \cdot \frac{n_T}{n};$$

i varijancu

$$Var(X) = \frac{r \cdot n_T \cdot (n - n_T) \cdot (n - r)}{n^2 \cdot (n-1)}.$$

PRIMJER 6.17 *Slučajni pokus: izbor od $r = 3$ elemenata (bez vraćanja) iz skupa koji ima $n = 10$ elemenata od kojih je $n_T = 7$.*
Promatramo slučajnu varijablu X = broj ispravnih predmeta u r članom uzorku iz skupa koji ima n elemenata, od toga n_T ispravnih (bez vraćanja).
(a) Izračunajte vjerojatnost da su u uzorku dva ispravna elementa (točno)?
(b) Izračunajte vjerojatnost da su dva ispravna, ako je uzorak s vraćanjem?

Rješenje:
(a) Uzorak bez vraćanja pa X = broj ispravnih predmeta u r članom uzorku iz skupa koji ima n elementata od toga n_T ispravnih (bez vraćanja) ima hipergeometrijsku distribuciju $X \sim Hip(n, r, n_T)$. Za $n = 10$, $r = 3$, $n_T = 7, X \sim Hip\,(10, 3, 7)$

$$P(X = 2) = f(2) = \frac{\binom{n_T}{2} \cdot \binom{n-n_T}{r-2}}{\binom{n}{r}} = \frac{\binom{7}{2} \cdot \binom{10-7}{3-2}}{\binom{10}{3}} = \frac{\binom{7}{2} \cdot 3}{\binom{10}{3}} = 0.525.$$

(b) Uzorak s vraćanjem pa X = broj ispravnih predmeta u r članom uzorku iz skupa koji ima n elementata od toga n_T ispravnih (s vraćanjem) ima binomnu distribuciju $X \sim B(r, p = \frac{n_T}{n})$. Za $n = 10$, $r = 3$, $n_T = 7, X \sim B(3, \frac{7}{10})$

$$P(X = 2) = f(2) = \binom{r}{2} p^2 (1-p)^{r-2} = \binom{3}{2} (\frac{7}{10})^2 \cdot (1 - \frac{7}{10})^{3-2} = 0.441.$$

Prisjetimo se motivacijskog primjera:

PRIMJER 6.18 *motiv*

U velikom pakovanju od 100 cigli distributer garantira da je najviše $p = 10\%$ oštećenih. Kontrolor prihvaća pošiljku samo ako u uzorku od 10 proizvoda iz pakovanja (bez vraćanja) nema oštećenih. Kolika je vjerojatnost da će kontrolor odbiti pošiljku?

Rješenje:
Uzorak bez vraćanja pa X = broj neispravnih predmeta u r članom uzorku iz skupa

koji ima n elementata od toga n_D neispravnih (bez vraćanja) ima hipergeometrijsku distribuciju $X \sim Hip(n, r, n_D)$. Za $n = 100,\ r = 10,\ n_D = 10$, $X \sim Hip\,(100, 10, 10)$

$$\begin{aligned}
P(X \geq 1) &= 1 - P(X < 1) = 1 - P(X = 0) = 1 - f(0) = 1 - \frac{\binom{n_D}{0} \cdot \binom{n-n_D}{r-0}}{\binom{n}{r}} \\
&= 1 - \frac{\binom{10}{0} \cdot \binom{100-10}{10-0}}{\binom{100}{10}} \\
&= 1 - \frac{\binom{90}{10}}{\binom{100}{10}} \\
&= 0.67.
\end{aligned}$$

Vjerojatnost da će kontolor odbiti pošiljku je 0.67.

NAPOMENA 6.7 *Ako je $n,\ n_T,\ n - n_T$ veliki u odnosu na r onda nije važno je li uzorak s vraćanjem ili bez vraćanja i $Hip\,(n, r, n_T)$ distribucija teži $B(r, p = \frac{n_T}{n})$.*
Ako je uzorak iz nepoznato velike populacije možemo opet koristiti binomnu distribuciju bez obzira na vraćanje.

6.4 GEOMETRIJSKA DISTRIBUCIJA

MOTIV 6.4 *Student izlazi na ispit dok ga ne položi. Ako je vjerojatnost polaganja ispita svaki put jednaka $p = \frac{1}{5}$, (jer je student naučio petinu gradiva) kolika je vjerojatnost da će student 4 puta izlaziti na ispit dok ga ne položi?*

Definicija 6.6 *(Geometrijska slučajna varijabla)*

Slučajnu varijablu X = broj ponavljanja pokusa (nezavisnih) dok se ne dogodi događaj A čija je vjerojatnost $p \in (0,1)$, zovemo geometrijska slučajna varijabla.
Slika geometrijske slučajne varijable je $\mathcal{R}(X) = \{1, 2, ...\}$.
Funkcija vjerojatnosti Binomne slučajne varijable je

$$f(x) := \begin{cases} P(X = k) = (1-p)^{k-1} \cdot p, & x = k \in \mathcal{R}(X) \\ 0, & \text{inače.} \end{cases}$$

Definicija 6.7 *(GEOMETRIJSKA DISTRIBUCIJA)*

Za sve slučajne varijable koje imaju sliku $\mathcal{R}(X) = \{1, 2, ...\}$ i funkciju vjerojatnosti

$$f(x) = (1-p)^{x-1} \cdot p, \ x \in \mathcal{R}(X)$$

kažemo da imaju GEOMETRIJSKU DISTRIBUCIJU (RAZDIOBU)
s parametrom p i označavamo

$$X \sim G(p).$$

PRIMJER 6.19 *Funkcija $f(x) = (1-p)^{x-1} \cdot p$ je funkcija vjerojatnosti.*
Prisjetimo se sume geometrijskog reda i iskoristimo za računanje

$$\sum_{1 \le k \le \infty} f(x) = \sum_{1 \le k \le \infty} (1-p)^{k-1} \cdot p = p \sum_{1 \le k \le \infty} (q)^{k-1} = p \cdot \frac{1}{1-q} = p \cdot \frac{1}{p} = 1$$

PRIMJER 6.20 *Očekivanje diskretne slučajne varijable koja ima geometrijsku distribuciju $X \sim G(p)$ je*

$$E(X) = \frac{1}{p}$$

Rješenje:

Prisjetimo se da konvergentan red (npr. geometrijski) možemo derivirati član po član, pa vrijedi

$$\sum_{k=1}^{\infty} kq^{k-1} = (\frac{1}{1-q})' = \frac{1}{(1-q)^2}.$$

$$\begin{aligned} E(X) &= \sum_{1\leq k\leq\infty} k\cdot f(k) = \sum_{k=1}^{\infty} k(1-p)^{k-1}\cdot p = p\cdot\sum_{k=1}^{\infty} kq^{k-1} = p\cdot\frac{1}{(1-q)^2} \\ &= p\cdot\frac{1}{p^2} = \frac{1}{p}. \end{aligned}$$

PRIMJER 6.21 *Varijanca diskretne slučajne varijable koja ima geometrijsku distribuciju $X \sim G(p)$ je*

$$Var(X) = \frac{1-p}{p^2}$$

Rješenje: (pokušajte sami)

Prisjetimo se motivacijskog primjera:

PRIMJER 6.22 *motiv*

Student izlazi na ispit dok ga ne položi. Ako je vjerojatnost polaganja ispita svaki put jednaka $p = \frac{1}{5}$, kolika je vjerojatnost da će student 4 puta izlaziti na ispit dok ga ne položi?

Rješenje: $X \sim G(p)$, $p = 0.2$,

$$P(X=4) = f(4) = (1-p)^{4-1}\cdot p = (1-p)^3\cdot p = \frac{64}{625} = 0.102.$$

6.5 Ponovimo

BINOMNA DISTRIBUCIJA (RAZDIOBA)

binomna distribucija	$X \sim B(m,p)$
slika binomne sl. var.	$\mathcal{R}(X) = \{0,1,2,...,m\}$
funkcija vjerojatnosti	$f(x) = \binom{m}{x} p^x (1-p)^{m-x}$
očekivanje	$E(X) = mp$
varijanca	$Var(X) = mp(1-p)$

POISOONOVA DISTRIBUCIJA

Poissonova distribucija	$X \sim Po(\lambda)$
slika Poissonove sl. var.	$\mathcal{R}(X) = \{0,1,2,...\}$
funkcija vjerojatnosti	$f(x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$
očekivanje	$E(X) = \lambda$
varijanca	$Var(X) = \lambda$

HIPERGEOMETRIJSKA DISTRIBUCIJA

hipergeometrijska distribucija	$X \sim Hip\,(n,r,n_T)$
slika Poissonove sl. var.	$\mathcal{R}(X) = \{0,1,2,..,n_T\}$
funkcija vjerojatnosti	$f(x) = \frac{\binom{n_T}{x} \cdot \binom{n-n_T}{r-x}}{\binom{n}{r}}$
očekivanje	$E(X) = r \cdot \frac{n_T}{n}$
varijanca	$Var(X) = \frac{r \cdot n_T \cdot (n-n_T) \cdot (n-r)}{n^2 \cdot (n-1)}$

GEOMETRIJSKA DISTRIBUCIJA

geometrijska distribucija	$X \sim G(p)$
slika Poissonove sl. var.	$\mathcal{R}(X) = \{1,2,...\}$
funkcija vjerojatnosti	$f(x) = (1-p)^{x-1} \cdot p$
očekivanje	$E(X) = \frac{1}{p}$
varijanca	$Var(X) = \frac{1-p}{p^2}$

Poglavlje 7

KONTINUIRANA SLUČAJNA VARIJABLA

MOTIV 7.1 *Neka je omjer prodaje i profita slučajna varijabla* $X = \frac{100}{profit}$. *Neka je poslovanje takvo da je nemoguće da će profit biti veći ili jednak* 50% , *a sigurno će profit biti veći ili jednak* 33.3%. *Neka je vjerojatnost da će omjer X poprimiti vrijednost manju ili jednaku x zadan sa funkcijom* $F(x) = \frac{(x^2-4)}{5}$. *Kolika je vjerojatnost da će profit biti između* 40% *i* 20%? *Koliki je očekivani profit?*

MOTIV 7.2 *Vrijeme trajanja sijalica je slučajna varijabla X. Uzimamo uzorak i 5% sijalica traje do 100 sati. Kolika je vjerojatnost da će nova sijalica trajati duže od 200 sati tj.* $P(X > 200)$?

Definicija 7.1 *KONTINUIRANA SLUČAJNA VARIJABLA*

Neka je $(\Omega, \mathcal{F}, P)$ *vjerojatnosni prostor. Za funkciju* $X : \Omega \to \mathbb{R}$ *kažemo da je kontinuirana slučajna varijabla ako je slika* $\mathcal{R}(X)$ *interval* $I \subseteq \mathbb{R}$, *ako slika ne sadrži izolirane točke i ako za svaki interval* $I \subset \mathbb{R}$ *praslika* $\{\omega \in \Omega : X(\omega) \in I\}$ *je familija skupova u* $\mathcal{F}$.

Definicija 7.2 *Za kontinuiranu slučajnu varijablu* $X : \Omega \to \mathbb{R}$ *definiramo vjerojatnost da poprimi vrijednosti iz intervala* $I \subset \mathbb{R}$:

$$P(X \in I) = P(\{\omega \in \Omega : X(\omega) \in I\});$$

$P(X = a) = 0$ *za bilo koji* $a \in \mathbb{R}$.

Definicija 7.3 *(FUNKCIJA GUSTOĆE VJEROJATNOSTI SL. VAR. engl. probability density function)*

Neka je $X : \Omega \to \mathbb{R}$ *kontinuirana slučajna varijabla sa slikom* $\mathcal{R}(X)$. *Funkciju* $f : \mathbb{R} \to \mathbb{R}$ *definiranu na sljedeći način:*

$$f(x) := \lim_{\triangle x \to 0} \frac{P(x \leq X \leq x + \triangle x)}{\triangle x}, \; x \in \mathcal{R}(X)$$

zovemo funkcija gustoće vjerojatnosti kontinuirane slučajne varijable X.
Za funkciju gustoće vjerojatnosti slučajne varijable vrijedi

$$\int_a^b f(x)\,dx = P(a \leq X \leq b).$$

NAPOMENA 7.1 *Geometrijsku interpretaciju funkcije gustoće* f *dobijemo pomoću geometrijske interpretacije određenog integrala funkcije* f :

$P(a \leq X \leq b)$ *je površina ispod krivulje gustoće vjerojatnosti* f *na segmentu* $[a, b]$.

T: SVOJSTVA funkcije gustoće vjerojatnosti slučajne varijable X.
(i) $f(x) \geq 0, \; x \in \mathcal{R}(X)$
(ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1.$

D:
(i) Za $x \in \mathcal{R}(X), \; x \in I \subseteq \mathbb{R}, \; f(x) = \lim_{\triangle x \to 0} \frac{P(x \leq X \leq x + \triangle x)}{\triangle x} \geq 0,$
(ii) Za $\mathcal{R}(X) = \mathbb{R}$ svojstvo (ii) slijedi iz ekvivalentne definicije od f i svojstva (P2), normiranosti funkcije P: $\int_{-\infty}^{\infty} f(x)\,dx = P(X \in \mathbb{R}) = P(\Omega) = 1.$

NAPOMENA 7.2 *Budući da slika kontinuirane slučajne varijable ne sadrži izolirane točke*
$P(X = a) = 0$ *onda vrijedi (za razliku od diskretnih slučajnih varijabli):*

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b).$$

NAPOMENA 7.3 *tko želi znati više*

Svojstvo $\int_a^b f(x)\,dx = P(a \leq X \leq b)$ *ekvivalentno je definiciji od* f.

Promatramo segment $[a, b]$ *i gledamo subdivizije*

$$a = x_0 < x_1 < ... < x_n = b, \;\; x_{i+1} = x_i + \triangle x_i, \;\; i = 1, .., n.$$

Postoji $t_i \in [x_i, x_i + \triangle x_i]$ *tako da* $P(x_i \leq X \leq x_i + \triangle x) \approx f(t_i) \cdot \triangle x_i,$

$\sum_i P(x_i \leq X \leq x_i + \triangle x_i) \approx \sum_i f(t_i) \cdot \triangle x_i$.

Ako promatramo $\lim_{\triangle x_i \to 0}$ *po svim subdivizijama, na desnoj strani prepoznajemo definiciju određenog integrala funkcije* f *na segmentu* $[a, b]$, *pa slijedi*

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx.$$

PRIMJER 7.1 *Kontinuirana slučajna varijabla je zadana sa svojom slikom* $\mathcal{R}(X) \subseteq \mathbb{R}$. *Zadana je funkcija* f:

$$f(x) := \begin{cases} a \cdot x, & 0 \leq x \leq 5 \\ 0, & x < 0,\ x > 5. \end{cases}$$

(a) Odredite konstantu a tako da funkcija f *bude gustoća vjerojatnosti slučajne varijable* X.
(b) Izračunajte vjerojatnost $P(0 < X < 2)$.
(c) Skicirajte graf funkcije f.

Rješenje:
(a) Da bi f bila funkcija gustoće vjerojatnosti slučajne varijable ona mora imati svojstva nenegativnosti i svojstvo $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$.
Zato, iz $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = \int_0^5 a \cdot x\,dx = 1$ dobivamo da je $a = \dfrac{1}{\int_0^5 x\,dx} = \dfrac{1}{25/2} = \dfrac{2}{25}$.
Funkcija gustoće vjerojatnosti slučajne varijable je

$$f(x) = \begin{cases} \frac{2}{25}x, & 0 \leq x \leq 5 \\ 0, & x < 0,\ x > 5. \end{cases}$$

(b) $P(0 < X < 2) = \displaystyle\int_0^2 f(x)\,dx = \int_0^2 \frac{2}{25}x\,dx = \frac{2}{25} \cdot \frac{4}{2} = \frac{4}{25}$.
(c) Funkcija f je definirana na cijelom $\mathbb{R}$, neprekinuta je na $\mathbb{R} \setminus \{5\}$.

Definicija 7.4 *(FUNKCIJA DISTRIBUCIJE KONTINUIRANE SLUČAJNE VARIJABLE, engl. distribution function)*

Neka je $X : \Omega \to \mathbb{R}$ *kontinuirana slučajna varijabla sa slikom* $\mathcal{R}(X) \subseteq \mathbb{R}$. *Funkciju* $F : \mathbb{R} \to \mathbb{R}$ *definiranu na sljedeći način:*

$$F(x) := P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$

zovemo funkcija distribucije kontinuirane slučajne varijable X.
Veza funkcije vjerojatnosti i funkcije distribucije slučajne varijable je:

$$f(x) = F'(x).$$

T: SVOJSTVA FUNKCIJE DISTRIBUCIJE slučajne varijable:
(F1) $\lim\limits_{x\to-\infty} F(x) = F(-\infty) = 0$
(F2) $\lim\limits_{x\to\infty} F(x) = F(\infty) = 1$
(F3) $0 \le F(x) \le 1$
(F4) F je neprekinuta funkcija, $F(x) := P(X \le x) = P(X < x)$,
(F5) $P(a < X \le b) = F(b) - F(a) = \displaystyle\int_a^b f(x)\,dx,\ \ a, b \in \mathbb{R},\ a < b$
(F6) F je rastuća funkcija.
D:
(F1) Kako je događaj $X \le -\infty$ nemoguć događaj

$$F(-\infty) := P(X \le -\infty) = P(\emptyset) = 0.$$

(F2) Kako je događaj $X \le \infty$ siguran događaj

$$F(\infty) := P(X \le \infty) = P(\Omega) = 1.$$

(F3) tvrdnja slijedi iz svojstava (F1) i (F2).
(F4) Iz definicije kontinuirane sl. varijable, slika nema izoliranih vrijednosti, pa je $P(X = a) = 0$ te na integral ne utječu točke prekida funkcije f.
(F5) Neka su $a, b \in \mathbb{R},\ a < b$. Računamo

$$\begin{aligned} F(b) &= P(X \le b) = P((X \le a) \cup (a < X \le b)) \\ &= P(X \le a) + P(a < X \le b) = F(a) + P(a < X \le b) \end{aligned}$$

pa slijedi $P(a < X \le b) = F(b) - F(a)$.

$$F(b) - F(a) = \int_{-\infty}^{b} f(x)\,dx - \int_{-\infty}^{a} f(x)\,dx = \int_a^b f(x)\,dx.$$

(F6) Neka su $a, b \in \mathbb{R},\ a < b$. Iz svojstva (F5) slijedi da je $F(a) \le F(b)$, funkcija je rastuća.

PRIMJER 7.2 *Za funkciju distribucije kontinuirane slučajne varijable* $F(x) = P(X \le x)$ *vrijedi:*

$$P(a < X < b) = P(a \le X \le b) = F(b) - F(a),\ \ a, b \in \mathbb{R}.$$

Rješenje:

$$\begin{aligned} F(b) &= P(X < b) = P((X \le a) \cup (a < X < b)) \\ &= P(X \le a) + P(a < X < b) = F(a) + P(a < X < b) \end{aligned}$$

PRIMJER 7.3 *Slučajna varijabla X iz prethodnog primjera je zadana s funkcijom gustoće vjerojatnosti*

$$f(x) = \begin{cases} \frac{2}{25}\,x, & 0 \le x \le 5 \\ 0, & x < 0,\, x > 5. \end{cases}$$

(a) Napišite funkciju distribucije.
(b) Izračunajte vjerojatnost da slučajna varijabla poprimi vrijednosti veće od 0 i manje od 2 , $P(0 < X < 2) = ?$
(c) Skicirajte graf funkcije distribucije $F(x)$.

Rješenje:
(a)

$$F(x) := P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$$

$$F(x) = \begin{cases} 0, & x < 0 \\ \displaystyle\int_0^x \frac{2}{25}\,t\,dt, & 0 \le x < 5 \\ 1, & 5 \le x. \end{cases} = \begin{cases} 0, & x < 0 \\ \frac{x^2}{25}, & 0 \le x < 5 \\ 1, & 5 \le x. \end{cases}$$

(b) $P(0 < X < 2) = F(2) - F(0) = \frac{2^2}{25} - 0 = \frac{4}{25}$.
(c) Funkcija $F(x)$ je neprekinuta na $\mathbb{R}$.

Definicija 7.5 *(OČEKIVANJE KONTINUIRANE SL. VAR., engl. mean ili mathematical expectation)*

Neka je $X : \Omega \to \mathbb{R}$ kontinuirana slučajna varijabla sa slikom $\mathcal{R}(X)$ i funkcijom vjerojatnosti $f(x)$. Kažemo da kontinuirana slučajna varijabla X ima očekivanje ako integral $\displaystyle\int_{-\infty}^{\infty} x \cdot f(x)\,dx$ konvergira i označavamo

$$E(X) := \int_{-\infty}^{\infty} x \cdot f(x)\,dx.$$

Definicija 7.6 *(VARIJANCA ILI DISPERZIJA KONTINUIRANE SL. VAR., engl. variance)*

Neka je $X : \Omega \to \mathbb{R}$ kontinuirana slučajna varijabla sa slikom $\mathcal{R}(X)$ i funkcijom gustoće vjerojatnosti $f(x)$. Kažemo da kontinuirana slučajna varijabla X ima varijancu ako integral $\displaystyle\int_{-\infty}^{\infty} (x - E(X))^2 f(x)\,dx$ konvergira i označavamo

$$Var(X) := \int_{-\infty}^{\infty} (x - E(X))^2 f(x)\,dx.$$

Možemo računati varijancu i pomoću relacije

$$Var(X) := \int_{-\infty}^{\infty} x^2 \cdot f(x)\, dx - (E(X))^2.$$

(Dokaz vidi kasnije - očekivanje funkcije slučajne varijable).

Definicija 7.7 *(STANDARDNA DEVIJACIJA, engl. standard deviation)*
Standardna devijacija slučajne varijable X definira se kao

$$\sigma(X) = \sqrt{Var(X)}.$$

PRIMJER 7.4 *Neka je zadana slučajna varijabla iz prethodnog primjera s funkcijom gustoće vjerojatnosti* $f(x) = \begin{cases} \frac{2}{25}x, & 0 \le x \le 5 \\ 0, & x < 0, x > 5. \end{cases}$
(a) Izračunajte očekivanje slučajne varijable $E(X)$.
(b) Izračunajte varijancu i standardnu devijaciju $Var(X)$, $\sigma(X)$.

Rješenje:
(a) $E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^5 x \frac{2}{25} x dx = \frac{2}{25} \int_0^5 x^2 dx = \frac{10}{3} = 3.333$
(b)

$$\begin{aligned} Var(X) &= \int_{-\infty}^{\infty} x^2 \cdot f(x) dx - (E(X))^2 = \int_0^5 x^2 \cdot \frac{2}{25} x dx - (\frac{10}{3})^2 \\ &= \frac{2}{25} \int_0^5 x^3 dx - (\frac{10}{3})^2 = \frac{25}{18}. \end{aligned}$$

$\sigma(X) = \sqrt{Var(X)} = \sqrt{\frac{25}{18}} = \frac{5}{6}\sqrt{2} = 1.178.$

Prisjetimo se motivacijskog primjera:

PRIMJER 7.5 *motiv*

Neka je omjer prodaje i profita slučajna varijabla $X = \frac{100}{profit}$. *Neka je poslovanje takvo da je nemoguće da će profit biti veći ili jednak* 50% *, a sigurno će profit biti veći ili jednak* 33.3%. *Neka je vjerojatnost da će omjer X poprimiti vrijednost manju ili jednaku x zadan sa funkcijom* $F(x) = \frac{(x^2-4)}{5}$. *Kolika je vjerojatnost da će profit biti između* 40% *i* 20%*? Koliki je očekivani profit?*

Rješenje:
Slučajna varijabla $X = \frac{100}{profit}$ poprima vrijednosti npr. za profit 50% vrijednost

$X = 2$, a za profit 33.3% vrijdnost $X = 3$. Zaključujemo da je zadana funkcija distribucije $P(X \leq x) = F(x)$

$$F(x) = \begin{cases} 0, & x < 2 \\ \frac{(x^2-4)}{5}, & 2 \leq x < 3 \\ 1, & 3 \leq x. \end{cases}$$

Vjerojatnost da će profit biti između 40% i 20% odgovara pitanju $P(2.5 \leq X \leq 5)$.
$P(2.5 \leq X \leq 5) = F(5) - F(2.5) = 1 - \frac{9}{20} = \frac{11}{20} = 0.55$ Očekivani profit odgovara očekivanoj vrijednosti od X.
Da bi izračunali očekivanje moramo izraziti funkciju gustoće vjerojatnosti f. Veza funkcije gustoće i funkcije distribucije je $f(x) = F'(x)$.

$$f(x) = \begin{cases} 0, & x < 2 \\ \frac{2x}{5}, & 2 \leq x < 3 \\ 0, & 3 \leq x. \end{cases}$$

Računamo očekivanje

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_{2}^{3} x \frac{2x}{5} dx = \frac{38}{15} = 2.533.$$

Očekivani profit je dakle 39.47%.

7.1 FUNKCIJA SLUČAJNE VARIJABLE

MOTIV 7.3 *Neka je* $X \sim \begin{pmatrix} 1 & 2 & 3 \\ 0.1 & 0.3 & 0.6 \end{pmatrix}$. *Opišite slučajnu varijablu* $Y = 3X + 1$.

Definicija 7.8 *(FUNKCIJA SLUČAJNE VARIJABLE) Neka je zadana po dijelovima neprekinuta funkcija* $h : \mathbb{R} \to \mathbb{R}$ *i slučajna varijabla* $X : \Omega \to \mathbb{R}$. *Funkcija slučajne varijable* X *je slučajna varijabla* $Y : \Omega \to \mathbb{R}$ *definirana kao kompozicija* $Y = h \circ X$.

TEOREM 7.1 *Ako je* $h(x)$ *strogo monotona i derivabilna onda funkciju slučajne varijable* $Y = h(X)$ *možemo definirati pomoću slučajne varijable* X :
(a) Ako je X *diskretna slučajna varijabla*

$$f_Y(y) = P(Y = y) = P(h(X) = y) = P(X = h^{-1}(y)) = f(h^{-1}(y)).$$

$$F_Y(y) = P(Y \leq y) = \sum_{x_i;\, h(x_i) \leq y} f(x_i).$$

$$E(Y) = \sum_i h(x_i) \cdot f(x_i)\,.$$

(b)Ako je X *kontinuirana slučajna varijabla*

$$f_Y(y) = \begin{cases} f(h^{-1}(y)) \cdot |(h^{-1}(y))'|, & m < y < M \\ 0, & y \leq m,\ y \geq M. \end{cases}$$

$$F_Y(y) = \begin{cases} 0, & & y \leq m \\ F(h^{-1}(y)), & h'(x) > 0, & m < y < M \\ 1 - F(h^{-1}(y)), & h'(x) < 0, & m < y < M \\ 1, & & y \geq M. \end{cases}.$$

gdje smo označili $m = \min h(x),\ M = \max h(x)$.

$$E(Y) = \int_{-\infty}^{\infty} h(x) \cdot f(x)\, dx.$$

Dokaz: *tko želi znati više*
(a) $f_Y(y) = P(Y = y) = P(h(X) = y) = P(X = h^{-1}(y)) = f(h^{-1}(y))$.

$$F_Y(y) = P(Y \leq y) = P(h(X) \leq y) = P(X \leq h^{-1}(y)) = \sum_{x_i:\, x_i \leq h^{-1}(y)} f(x_i).$$

$$E(Y) = \sum_i y_i f_Y(y_i) = \sum_i h(x_i) f(h^{-1}(y_i)) = \sum_i h(x_i) f(x_i).$$

(b) Ako je X kontinuirana slučajna varijabla

$$F_Y(y) = P(Y \leq y) = P(h(X) \leq y) = \begin{cases} P(X \leq h^{-1}(y), & h'(x) > 0, \\ 1 - P(X \leq h^{-1}(y), & h'(x) < 0. \end{cases}$$

$$f_Y(y) = (F_Y(y))\prime = (F_X(h^{-1}(y)))' \cdot (h^{-1}(y))' = f(h^{-1}(y)) \cdot |(h^{-1}(y))'|.$$

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_{-\infty}^{\infty} h(x) f(h^{-1}(y))(h^{-1}(y))' dy = \int_{-\infty}^{\infty} h(x) f(x) dx.$$

PRIMJER 7.6 *Neka je zadana funkcija $h(x) = a \cdot x + b$, i slučajna varijabla X. Neka je Y funkcija slučajne varijable X, $Y = h(X)$, tj. $Y = a \cdot X + b$.*
(a) Neka je X diskretna slučajna varijabla. Tada je Y definirana s:

$$f_Y(y) = f(\frac{y-b}{a}),$$

$$F_Y(y) = P(Y \leq y) = \sum_{x_i;\, x_i \leq \frac{y-b}{a}} f(x_i),$$

$$E(Y) = a \cdot E(X) + b,$$

$$Var(Y) = a^2 \cdot Var(X).$$

(b) Neka je X kontinuirana slučajna varijabla. Tada je Y definirana s:

$$f_Y(y) = f(\frac{y-b}{a}) \cdot \frac{1}{a},$$

$$F_Y(y) = \begin{cases} F(\frac{y-b}{a}), & a > 0, \\ 1 - F(\frac{y-b}{a}), & a < 0, \end{cases}$$

$$E(Y) = a \cdot E(X) + b,$$

$$Var(X) = a^2 \cdot Var(X).$$

Rješenje:

(a) $f_Y(y) = P(Y = y) = P(aX + b = y) = P(X = \dfrac{y-b}{a}) = f(\dfrac{y-b}{a})$.

(b) $h(x)$ je strogo monotona za $a \neq 0$, $h^{-1}(y) = \dfrac{y-b}{a}$, $(h^{-1}(y))' = \dfrac{1}{a}$.

$$F_Y(y) = P(Y \leq y) = P(a \cdot X + b \leq y) = \begin{cases} P(X \leq \frac{y-b}{a}), & a > 0, \\ 1 - P(X \leq \frac{y-b}{a}), & a < 0. \end{cases}$$

Prisjetimo se motivacijskog primjera:

PRIMJER 7.7 *motiv*

Neka je $X \sim \begin{pmatrix} 1 & 2 & 3 \\ 0.1 & 0.3 & 0.6 \end{pmatrix}$. *Opišite slučajnu varijablu* $Y = 3X + 1$.

Rješenje:

$$Y \sim \begin{pmatrix} 4 & 7 & 10 \\ 0.1 & 0.3 & 0.6 \end{pmatrix}$$

PRIMJER 7.8 *Neka je zadana slučajna varijabla* X *s funkcijom gustoće vjerojatnosti* $f(x) = \begin{cases} \frac{2}{25}x, & 0 \leq x \leq 5, \\ 0, & x < 0, x > 5, \end{cases}$ $F(x) = \begin{cases} 0, & x < 0, \\ \frac{x^2}{25}, & 0 \leq x < 5, \\ 1, & 5 \leq x, \end{cases}$

$E(X) = \frac{10}{3}$, $Var(X) = \frac{25}{18}$. *Opišite slučajnu varijablu* $Y = 3X + 1$.

Rješenje:

$h(x) = 3x + 1$, $m = min\, h(x) = 1$, $M = max\, h(x) = 16$, $x \in [0, 5]$.

$$f_Y(y) = \begin{cases} f(\frac{y-1}{3}) \cdot \frac{1}{3}, & m < y < M \\ 0, & y \leq m, y \geq M \end{cases} = \begin{cases} \frac{2}{25 \cdot 3}(\frac{y-1}{3}), & 1 < y < 16 \\ 0, & y \leq 1, y \geq 16. \end{cases}$$

$$F_Y(y) = \begin{cases} 0, & y \leq m \\ F(\frac{y-1}{3}), & m < y < M \\ 1, & y \geq M \end{cases} = \begin{cases} 0, & y \leq 1 \\ \frac{1}{25}(\frac{y-1}{3})^2, & 1 < y < 16 \\ 1, & y \geq 16 \end{cases}$$

$$E(Y) = 3 \cdot E(X) + 1 = 3 \cdot \frac{10}{3} + 1 = 11$$

$$Var(Y) = 9 \cdot Var(X) = 9 \cdot \frac{25}{18} = \frac{25}{2}.$$

PRIMJER 7.9 *Varijanca slučajne varijable X je očekivanje slučajne varijable* $Y = (X - E(X))^2$

(a) $Var(X) = E(Y) = E((X - E(X))^2)$

(b) $Var(X) = E(X^2) - (E(X))^2$

Rješenje:

(a) slijedi iz defininicije varijance,

(b)

$$\begin{aligned} Var(X) &= E(X^2 - 2X \cdot E(X) + (E(X))^2) \\ &= E(X^2) - 2 \cdot E(X) \cdot E(X) + (E(X))^2 = E(X^2) - (E(X))^2 \end{aligned}$$

Definicija 7.9 *(k-TI MOMENT)*

k-ti moment slučajne varijable X je očekivanje funkcije slučajne varijable X^k*:*

$$\mu_k = E(X^k).$$

PRIMJER 7.10 *Očekivanje je 1. moment slučajne varijable:*

$$\mu_1 = E(X) = \mu.$$

Definicija 7.10 *(k-TI CENTRALNI MOMENT)*

k-ti centralni moment slučajne varijable X je očekivanje funkcije slučajne varijable $(X - E(X))^k$*:*

$$\beta_k = E((X - E(X))^k).$$

PRIMJER 7.11 *Varijanca je 2. centralni moment slučajne varijable:*

$$\beta_2 = E((X - E(X))^2) = Var(X) = \sigma^2.$$

7.2 Ponovimo

KONTINUIRANA SLUČAJNA VARIJABLA

slika kontinuirane slučajne varijable	$\mathcal{R}(X) = I \subset \mathbb{R}$
funkcija gustoće vjerojatnosti kon. sl. var.	$f(x) := \lim_{\triangle x \to 0} \frac{P(x \leq X \leq x + \triangle x)}{\triangle x}$
funkcija distribucije kon. sl. var. X	$F(x) := P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt$
očekivanje kon. sl. var. X	$E(X) := \int_{-\infty}^{\infty} x \cdot f(x)\, dx$
varijanca kon. sl. var. X	$Var(X) := \int_{-\infty}^{\infty} (x - E(X))^2 f(x)\, dx$

FUNKCIJA OD SLUČAJNE VARIJABLE

Y je funkcija od slučajne varijable X	$Y = h(X)$,
$Y = a \cdot X + b$	$Y : \Omega \to \mathbb{R}$
očekivanje od Y	$E(Y) = a \cdot E(X) + b$
varijanca od Y	$Var(Y) = a^2 Var(X)$

FUNKCIJA od DISKRETNE SLUČAJNE VARIJABLE $Y = a \cdot X + b$

funkcija vjerojatnosti od Y	$f_Y(y) = f(\frac{y-b}{a})$
funkcija distribucije od Y	$F_Y(y) = P(Y \leq y) = \sum_{x_i;\, x_i \leq \frac{y-b}{a}} f(x_i)$

FUNKCIJA od KONTINUIRANE SLUČAJNE VARIJABLE $Y = a \cdot X + b$

funkcija gustoće vjerojatnosti od Y	$f_Y(y) = f(\frac{y-b}{a}) \cdot \frac{1}{a}$
funkcija distribucije od Y	$F_Y(y) = \begin{cases} F(\frac{y-b}{a}), & a > 0, \\ 1 - F(\frac{y-b}{a}), & a < 0, \end{cases}$

Poglavlje 8

PRIMJERI KONTINUIRANIH SLUČAJNIH VARIJABLI

U ovom poglavlju istaknut ćemo kontinuirane slučajne varijable koje se pojavljuju u određenim "scenarijima" i njihove funkcije gustoće vjerojatnosti definirat će se kao specifične distribucije (razdiobe): normalna, uniformna, eksponencijalna, gama, hi-kvadrat, Studentova. Normalna distribucija, hi-kvadrat i studentova distribucija imaju veliku ulogu u matematičkoj statistici.

MOTIV 8.1 *Potrošnja materijala u nekom proizvodnom procesu je slučajan pokus. U prosjeku svaki dan se potroši 20 komada. Svaki mjesec se nabavlja 640 komada potrošnog materijala. Neka je X slučajna varijabla = vrijeme potrebno da se potroši zaliha (dogodi događaj α puta).*
(a) Kolika je vjerojatnost da ponestane potrošnog materijal?
(b) Kolika mora biti mjesečna nabavka da vjerojatnost nestašice bude 0.01?

8.1 NORMALNA DISTRIBUCIJA

Najvažnija kontinuirana distibucija je normalna distribucija. Ona se pojavljuje kao aproksimacija mnogih drugih distribucija i pojavljuje se u mnogim statističkim testovima. Njemački matematičar Carl F. Gauss (1777-1855) je 1809. godine objavio monografiju u kojoj je dao normalnu distribuciju kao model za pogreške u eksperimentu.

MOTIV 8.2 *Neka je kritična čvrstoća jednog tipa plastičnih ploča normalna slučajna varijabla X s očekivanom čvrstoćom 1250 kg i standardnom devijacijom 55 kg. Koje je maksimalno opterćenje takvo da je očekivani broj slomljenih ploča najviše* 5%*?*

Definicija 8.1 *(NORMALNA DISTRIBUCIJA)*

Za kontinuiranu slučajnu varijablu $X^* : \Omega \to \mathbb{R}$ kažemo da ima standardnu normalnu distribuciju ili standardnu Gaussovu distribuciju s parametrima 0 i 1 i označavamo $X^* \sim N(0,1)$ ako ima funkciju gustoće vjerojatnosti

$$f^*(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2}.$$

Za kontinuiranu slučajnu varijablu $X : \Omega \to \mathbb{R}$ kažemo da ima normalnu distribuciju ili Gaussovu distribuciju s parametrima μ i σ i označavamo
$X \sim N(\mu, \sigma^2)$ ako ima funkciju gustoće vjerojatnosti

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}.$$

NAPOMENA 8.1 *Ako $X^* \sim N(0,1)$ ima standardnu normalnu distribuciju, onda funkcija slučajne varijable $X = \sigma \cdot X^* + \mu$ ima normalnu distribuciju $X \sim N(\mu, \sigma^2)$.*

PRIMJER 8.1 *Funkcija $f^*(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2}$ je funkcija gustoće vjerojatnosti.*

Rješenje: Koristimo izvod $\displaystyle\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx = \sqrt{2\pi}$.

$$\int_{-\infty}^{\infty} f^*(x)\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx = \frac{1}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = 1.$$

PRIMJER 8.2 *Funkcija distribucije standardne normalne slučajne varijable $X^* \sim N(0,1)$ je*

$$F^*(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\,dt.$$

$$F^*(x) = 1 - F^*(-x)$$

Funkcija distribucije normalne slučajne varijable $X \sim N(\mu, \sigma^2)$ je

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2}\,dt.$$

Rješenje: Koristimo definiciju funkcije distribucije kontinuirane slučajne varijable $F^*(x) = \int_{-\infty}^{x} f^*(t)\,dt$ i formulu za funkciju distribucije za funkciju slučajne varijable $X = \sigma \cdot X^* + \mu$, $F(y) = F^*(\frac{y-\mu}{\sigma})$.

PRIMJER 8.3 *(a) Za* $X^* \sim N(0,1)$, $E(X^*) = 0$, $Var(X) = 1$.
(b) Za $X \sim N(\mu, \sigma^2)$, $E(X^*) = \mu$, $Var(X) = \sigma^2$.

Rješenje: Koristimo izvod $\int_{-\infty}^{\infty} x^2 \cdot e^{-\frac{1}{2}x^2}\,dx = \sqrt{2\pi}$.
(a) $E(X^*) = \int_{-\infty}^{\infty} x f^*(x) dx = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} dx = 0$ (neparna funkcija).

$$\begin{aligned} Var(X^*) &= E((X^*)^2) - (E(X^*))^2 = \int_{-\infty}^{\infty} x^2 f^*(x) dx = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = 1. \end{aligned}$$

(b)

$$E(X) = E(\sigma \cdot X^* + \mu) = \sigma \cdot E(X^*) + \mu = \mu,$$

$$Var(X) = Var(\sigma \cdot X^* + \mu) = \sigma^2 Var(X^*) = \sigma^2.$$

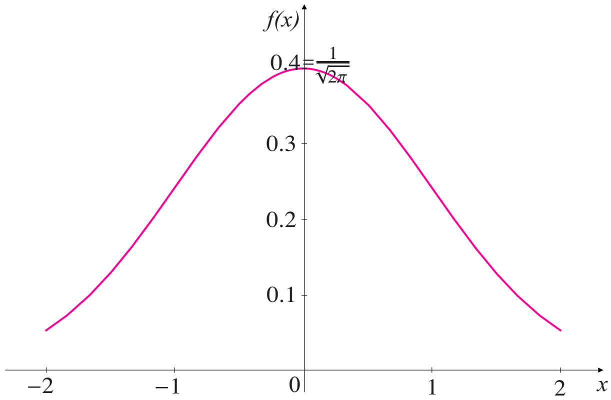
PRIMJER 8.4 *Skicirajte graf funkcije gustoće vjerojatnosti* $f(x)$. *Krivulja se zove Gaussova zvonolika krivulja. Krivulja je simetrična u odnosu na pravac* $x = \mu$. *dostiže maksimum u točki* $(\mu, \frac{1}{\sigma\sqrt{2\pi}})$, *a točke infleksije su u* $\mu - \sigma$ *i* $\mu + \sigma$. *Os x je horizontalna asimptota. Ako* $\sigma < 1$ *graf se sužava i maksimala vrijednost raste, a ako je* $\sigma > 1$ *graf se širi i maksimalna vrijednost pada.*

PRIMJER 8.5 *Neka je* $X^* \sim N(0,1)$. *Tada možemo izračunati vjerojatnost da slučajna varijabla* X^* *poprimi vrijednost iz nekog intervala* $[a,b]$:

$$P(\alpha \leq X^* \leq \beta) = F^*(\beta) - F^*(\alpha).$$

Neka je $X \sim N(\mu, \sigma^2)$. *Tada možemo izračunati vjerojatnost da slučajna varijabla* $X = \sigma \cdot X^* + \mu$ *poprimi vrijednost iz nekog intervala* $[a,b]$:

$$P(a \leq X \leq b) = F(b) - F(a) = F^*(\frac{b-\mu}{\sigma}) - F^*(\frac{a-\mu}{\sigma}).$$



Slika 8.1: Graf funkcije gustoće vjerojatnosti $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ jediniďż˝ne normalne sluďż˝ajne varijable $X \sim N(0, 1)$.

NAPOMENA 8.2 *U literaturi je poznata Laplaceova funkcija*

$$L(x) = \frac{1}{\sqrt{2\pi}} \int_0^x exp^{-\frac{1}{2}t^2}\, dt.$$

Ona je obično tabelirana. Veza funkcija F^*, F i $L(x)$ je sljedeća:

$$F^*(x) = \frac{1}{2} + L(x)$$

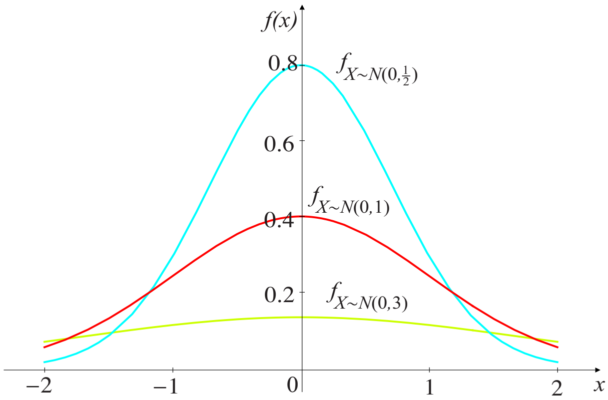
$$F(x) = \frac{1}{2} + L(\frac{x-\mu}{\sigma})$$

PRIMJER 8.6 *Slučajna varijabla $X \sim N(20, 4)$. Provjerite*
(a) $P(18 \leq X \leq 22) = 0.68$
(b) $P(16 \leq X \leq 24) = 0.95$
(c) $P(14 \leq X \leq 26) = 0.99$

Rješenje:

$$\begin{aligned}(a)\ \ P(18 \leq X \leq 22) &= F^*(\frac{22-20}{2}) - F^*(\frac{18-20}{2}) = F^*(1) - F^*(-1) \\ &= F^*(1) - (1 - F^*(1)) = 2F^*(1) - 1 \\ &= 2 \cdot 0.8413 - 1 = 1.6826 - 1 = 0.68\end{aligned}$$

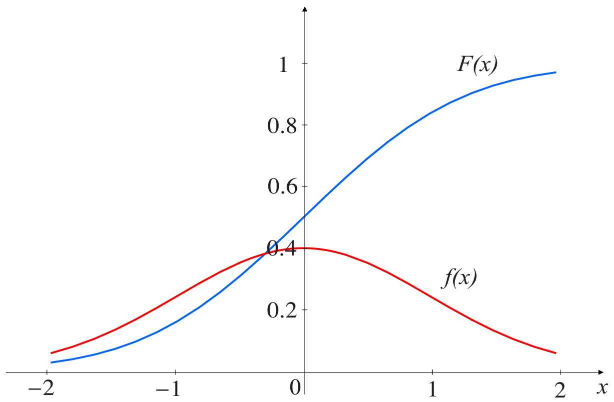
PRIMJER 8.7 *Slučajna varijabla $X \sim N(\mu, \sigma^2)$. Provjerite pravilo 3σ :*
(a) $P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.68$
(b) $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.95$
(c) $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.99$



Slika 8.2: Grafovi funkcija gustoće vjerojatnosti normalnih sluďż˝ajnih varijabli $X_1 \sim N\left(0, \frac{1}{2}\right)$, $X_2 \sim N(0, 1)$ i $X_3 \sim N(0, 3)$.

PRIMJER 8.8 *Debljina željeznih ploča ja slučajna varijabla. Možemo pretpostaviti da je to kontinuirana slučajna varijabla koja ima normalnu distribuciju s očekivanjem* 10*mm i standardnom devijacijom* 0.02*mm. Kolika je vjerojatnost defektne ploče ako je kontrola dala kriterij:*
(a) ploča tanja od 9.97 mm,
(b) ploča deblja od 10.05 mm,
(c) ploča odstupa 0.03 mm od 10 mm. U kojim granicama treba biti debljina da bi u očekivani postotak defektnih ploča bio 5%*?*

Rješenje:
Za slučajnu varijablu $X \sim N(10, 0.02^2)$ računamo:
(a) $P(X < 9.97) = F^*(\frac{9.97-10}{0.02}) = F^*(-1.5) = 0.0668$. Uz ovaj kriterij očekuje se 6.7% oštećenih ploča.
(b) $P(X > 10.05) = 1-F^*(10.05) = 1-F^*(\frac{10.05-10}{0.02}) = 1-F^*(2.5) = 1-0.9938 = 0.0062$. Uz ovaj kriterij očekuje se 0.6% oštećenih ploča.
(c) $P(9.97 < X < 10.03) = F^*(\frac{10.03-10}{0.02}) - F^*(\frac{9.97-10}{0.02}) = F^*(1.5) - F^*(-1.5) = 0.8664$.
$1 - P(9.97 < X < 10.03) = 0.1336$
Uz ovaj kriterij očekuje se 13.3% oštećenih ploča.
Prema pravilu 3σ :
$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.95$,
$P(10 - 2 \cdot 0.02 \leq X \leq 10 + 2 \cdot 0.02) = 0.95$.
Uz ovaj kriterij očekuje se 5% oštećenih ploča.
Za $2\sigma = 0.04$ imamo interval dobrih ploča $[9.96, 10.04]$.



Slika 8.3: Grafovi funkcije gustoće vjerojatnosti i funkcije distribucije jedinidž˝ne normalne sludž˝ajne varijable $X \sim N(0, 1)$.

Prisjetimo se motivacijskog primjera:

PRIMJER 8.9 *motiv*

Neka je kritična čvrstoća jednog tipa plastičnih ploča normalna slučajna varijabla X s očekivanom čvrstoćom 1250 kg i standardnom devijacijom 55 kg. Koje je maksimalno opterćenje takvo da je očekivani broj slomljenih ploča najviše 5%*?*

Rješenje:
Za slučajnu varijablu $X \sim N(1250, 55^2)$ računamo prema pravilu 3σ :
$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.95,$
$P(1250 - 2 \cdot 55 \leq X \leq 1250 + 2 \cdot 55) = 0.95$. Za maksimalno opterećenje od 1360 kg očekuje se 5% slomljenih ploča.

PRIMJER 8.10 *Na prvoj godini GF studira 200 studenata. Očekivana težina je 75 kg a standardna devijacija 7 kg. Pretpostavimo da je težina normalno distribuirana. Koliko studenata ima težinu između 68 kg i 82 kg (tj. zaokruženo od 67.5.5 kg do 82.5 kg)?*

Rješenje:
Za slučajnu varijablu težina studenata $X \sim N(75, 7^2)$ računamo:
$P(67.5 \leq X \leq 82.5) = F^*(\frac{82.5-75}{7}) - F^*(\frac{67.5-75}{7}) = 2F^*(1.07) - 1 = 2 \cdot 0.8577 - 1 = 0.715.$
Ukupno studenata koji imaju težinu između 68 kg i 82 kg (tj. zaokruženo od 67.5.5 kg do 82.5 kg) je $200 \cdot 0.715 = 143$.

NAPOMENA 8.3 *U graničnom slučaju za velike m binomna distribucija $X \sim B(m, p)$ se može aproksimirati standardnom normalnom distribucijom. Prema integralnom Moivre-Laplaceov teoremu (centralni granični teorem za binomnu sl. var.) u poglavlju 11 vrijedi*

$$P(a < X < b) \approx F^*(\frac{b - mp + 0.5}{\sqrt{mp(1-p)}}) - F^*(\frac{a - mp - 0.5}{\sqrt{mp(1-p)}}).$$

Primijetimo da su dodani pribrojnici 0.5 i -0.5 zbog korekcije.

8.2 UNIFORMNA DISTRIBUCIJA

Definicija 8.2 *(UNIFORMNA DISTRIBUCIJA)*

Za kontinuiranu slučajnu varijablu $X : \Omega \to \mathbb{R}$ kažemo da ima uniformnu distribuciju na se segmentu $[a, b]$, ako je slika $R(X) = R$, a funkcija gustoće vjerojatnosti je

$$f(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & a \leq x \leq b \\ 0, & b < x. \end{cases}$$

i označavamo $X \sim U(a, b)$.

PRIMJER 8.11 *Funkcija distribucije uniformne slučajne varijable $X \sim U(a, b)$ je*

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b < x. \end{cases}$$

PRIMJER 8.12 *Za $X \sim U(a, b)$, $E(X) = \frac{a+b}{2}$, $Var(X) = \frac{(b-a)^2}{12}$.*

Rješenje:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b x \, dx = \frac{1}{b-a} \left(\frac{1}{2} b^2 - \frac{1}{2} a^2 \right) \\ &= \frac{a+b}{2}. \end{aligned}$$

$$\int_{-\infty}^{\infty} x^2 f(x) dx = \int_a^b x^2 \frac{1}{b-a} dx = \frac{b^2 + ab + a^2}{3}$$

$$Var(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - (E(X))^2 = \frac{b^2 + ab + a^2}{3} - (\frac{a+b}{2})^2 = \frac{(b-a)^2}{12}.$$

PRIMJER 8.13 *Skiciraj graf funkcije gustoće vjerojatnosti i graf funkcije distribucije vjerojatnosti.*

Funkcija $f(x)$ ima prekid u točkama a i b, a funkcija $F(x)$ je neprekinuta na $\mathbb{R}$.

8.3 EKSPONENCIJALNA DISTRIBUCIJA

Eksponencijalna distribucija se pojavljuje u problemima teorije opsluživanja.

MOTIV 8.3 *Vrijeme trajanja sijalica je slučajna varijabla X. Uzimamo uzorak i 5% sijalica traje do 100 sati. Kolika je vjerojatnost da će nova sijalica trajati duže od 200 sati tj.* $P(X > 200)$?

Definicija 8.3 *(EKSPONENCIJALNA DISTRIBUCIJA)*

Za kontinuiranu slučajnu varijablu $X : \Omega \to \mathbb{R}$ *kažemo da ima eksponencijalnu distribuciju s parametrom* λ, *ako je slika* $R(X) = R$, *a funkcija gustoće vjerojatnosti je*

$$f(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & 0 \le x \end{cases}$$

i označavamo $X \sim Exp(\lambda)$.

PRIMJER 8.14 *Funkcija distribucije eksponencijalne slučajne varijable* $X \sim Exp(\lambda)$ *je*

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & 0 \le x \end{cases}$$

PRIMJER 8.15 *Za* $X \sim Exp(\lambda)$, $E(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$.

Rješenje:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_{0}^{\infty} x\lambda e^{-\lambda x} dx = parc.int. = \frac{1}{\lambda}.$$

$$\int_{-\infty}^{\infty} x^2 f(x) dx = \int_{0}^{\infty} x^2 \lambda e^{-\lambda x} dx = parc.int. = \frac{2}{\lambda^2}$$

$$Var(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - (E(X))^2 = \frac{2}{\lambda^2} - (\frac{1}{\lambda})^2 = \frac{1}{\lambda^2}.$$

PRIMJER 8.16 *Skiciraj graf funkcije gustoće i funkcija distribucije eksponencijalne slučajne varijable* $X \sim Exp(\lambda)$.
Funkcija $f(x)$ *je padajuća funkcija na* $[0, \infty)$, *prekinuta u nuli. Funkcija distribucije je neprekinuta funkcija R, rastuća, konkavna, ima horizontalnu asimptotu* $y = 1$.

Prisjetimo se motivacijskog primjera:

PRIMJER 8.17 *motiv*

Vrijeme trajanja sijalica je slučajna varijabla koja ima eksponencijalnu distribuciju s parametrom λ. Uzimamo uzorak i 5% sijalica traje do 100 sati.
(a) Odredite parametar λ.
(b) Kolika je vjerojatnost da će nova sijalica trajati duže od 200 sati?

Rješenje:

(a) $X \sim Exp(\lambda),\ P(X < 100) = 0.05 \ \Rightarrow\ F(100) = 0.05$

$F(x) = 1 - e^{-\lambda x}, \quad 0 \le x, \quad \Rightarrow \quad 1 - e^{-\lambda \cdot 100} = 0.05$

$e^{-\lambda \cdot 100} = 0.95 \quad \Rightarrow \quad \lambda \cdot 100 \approx 0.051 \quad \Rightarrow \quad \lambda \approx 0.00051.$

$$\begin{aligned}(b)P(X > 200) &= 1 - P(X < 200) = e^{-\lambda \cdot 200} \approx (e^{-\lambda \cdot 100})^2 \\ &= (0.95)^2 = 0.9025.\end{aligned}$$

8.4 GAMA DISTRIBUCIJA

Definicija 8.4 *(GAMA DISTRIBUCIJA)*

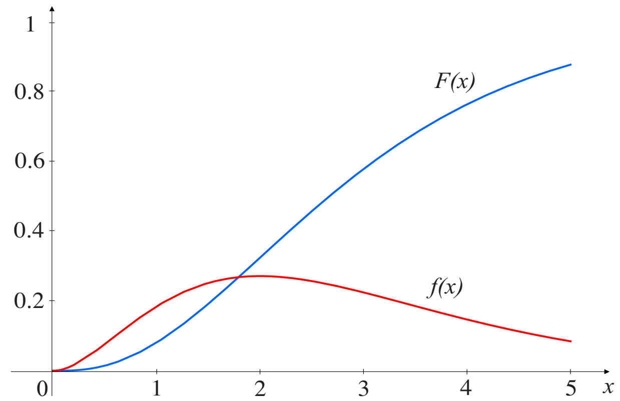
Gama distribucija je generalizacija eksponencijalne distribucije.
Neka je slučajni pokus ponavljanje događaja u vremenu s zadanim konstantnim intezitetom (λ). *Slučajna varijabla koja daje vrijeme potrebno da se događaj dogodi određeni broj puta* (α) *ima gama distibuciju s parametrima* α *i* λ.

Definicija 8.5 *(GAMA DISTRIBUCIJA)*

Za kontinuiranu slučajnu varijablu $X : \Omega \to \mathbb{R}$ *kažemo da ima gama distribuciju s parametrima* α *i* λ, $(\alpha \text{ i } \lambda > 0)$ *, ako je slika* $\mathcal{R}(X) = \mathbb{R}$*, a funkcija gustoće vjerojatnosti je*

$$f(x) = \begin{cases} 0, & x \leq 0 \\ C \cdot x^{\alpha-1} \cdot e^{-\lambda x}, & x > 0 \end{cases}$$

gdje je $C = \dfrac{\lambda^\alpha}{\Gamma(\alpha)}$*, a* $\Gamma(x) = \displaystyle\int_0^\infty t^{x-1} \cdot e^{-t} dt$*, Gama funkcija,* $x > 0$ *i označavamo* $X \sim \Gamma(\alpha, \lambda)$.



Slika 8.4: Grafovi funkcije gustoće vjerojatnosti i funkcije distribucije sluďż˝ajne varijable $X \sim \Gamma(3, 1)$.

PRIMJER 8.18 *Funkcija distribucije gama distribucije* $X \sim \Gamma(\alpha, \lambda)$ *je*

$$F(x) = \begin{cases} 0, & x < 0 \\ C \cdot \displaystyle\int_0^x t^{\alpha-1} \cdot e^{-\lambda t} dt, & x \geq 0. \end{cases}.$$

PRIMJER 8.19 *SVOJSTVA gama funkcije:*
(a) $\Gamma(\alpha+1)=\alpha\cdot\Gamma(\alpha)$,
(b) $\Gamma(1)=\displaystyle\int_0^{\infty} e^{-t}dt=1$,
(c) $\Gamma(n+1)=n\cdot\Gamma(n)=n\cdot(n-1)\cdot\Gamma(n-2)=...=n!,\ n\in N$,

PRIMJER 8.20 *Za* $X\sim\Gamma(\alpha,\lambda),\ E(X)=\dfrac{\alpha}{\lambda},\ Var(X)=\dfrac{\alpha}{\lambda^2}$.

PRIMJER 8.21 *Za* $\alpha>1$ *funkcije gustoće gama distribucije ima maksimum u* $x=\frac{\alpha-1}{\lambda}$, *ima zvonoliki oblik i os x je horizontalna asimptota, a za* $\alpha<1$, $f(x)$ *je strogo padajuća, konkavna funkcija koja ima vertikalnu asimptotu os y, a horizontalnu os x.*

Definicija 8.6 *(NEPOTPUNA GAMA FUNKCIJA)*
Tabelirana je nepotpuna gama funkcija $\gamma(z,\alpha)=\dfrac{1}{\Gamma(\alpha)}\displaystyle\int_0^z t^{\alpha-1}\cdot e^{-t}dt$.
Veza je $F(x;\ \alpha,\lambda)=\gamma(\lambda x,\alpha)$.

Prisjetimo se motivacijskog primjera:

PRIMJER 8.22 *motiv*
Potrošnja materijala u nekom proizvodnom procesu je slučajan pokus. U prosjeku svaki dan se potroši 20 komada. Svaki mjesec se nabavlja 640 komada potrošnog materijala. Neka je X slučajna varijabla = vrijeme potrebno da se potroši zaliha (dogodi događaj α *puta).*
(a) Kolika je vjerojatnost da ponestane potrošnog materijal?
(b) Kolika mora biti mjesečna nabavka da vjerojatnost nestašice bude 0.01?
$\lambda=20,\ \ \alpha=640,\ \ X\sim\Gamma(\alpha,\lambda)=\Gamma(640,20)$.

Rješenje:
Potrebni podaci: $\lambda=20,\ \ \alpha=640,\ \ X\sim\Gamma(\alpha,\lambda)=\Gamma(640,20)$.

(a) $P(X<30)=F(30)=F(30;640,20),\ \ F(x;\alpha,\lambda)=\gamma(\lambda x,\alpha)$
$F(30;640,20)=\gamma(20\cdot 30,640)=\gamma(600,640)=0.057$.
Vjerojatnost da bude nestašica je 0.057.

(b) $P(X<30)=0.01,\ \ F(30;\alpha,20)=0.01$
$F(30;\alpha,20)=\gamma(20\cdot 30,\alpha)=\gamma(600,\alpha)=0.01 \quad\Rightarrow\quad \alpha=660$.
Potrebne mjesečne zalihe potrošnog materijala su 660 komada da bi vjerojatnost nestašice bila 0.01 (mala).

PRIMJER 8.23 *Za* $\alpha=1$, *gama distribucija je eksponencijalna distribucija* $X\sim\Gamma(1,\lambda)=Exp(\lambda)$.

8.5 HI KVADRAT DISTRIBUCIJA

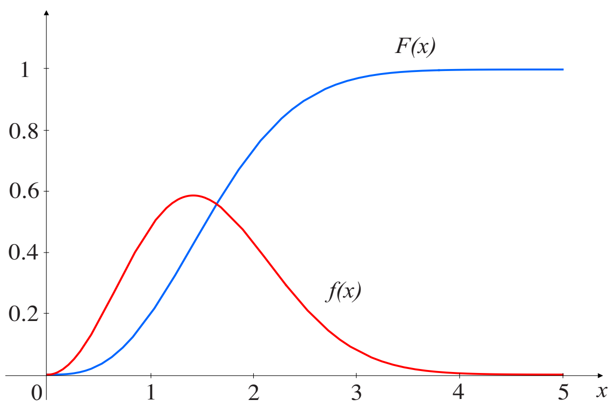
Definicija 8.7 *(HI KVADRAT DISTRIBUCIJA)*

Za $\alpha = \frac{n}{2}$, $\lambda = \frac{1}{2}$ *gama distribucija je* $\chi^2(n)$, *hi kvadrat distribucija s parametrom* n, $X \sim \Gamma(\frac{n}{2}, \frac{1}{2}) = \chi^2(n)$.

Funkcija gustoće vjerojatnosti je

$$f(x) = \begin{cases} 0, & x \leq 0 \\ C \cdot x^{\frac{n}{2}-1} \cdot e^{-\frac{1}{2}x}, & x > 0 \end{cases}$$

gdje je $C = \dfrac{(\frac{1}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}$.



Slika 8.5: Grafovi funkcije gustoće vjerojatnosti i funkcije distribucije sluďž˝ajne varijable $X \sim \chi^2(3)$.

PRIMJER 8.24 *Tabelira se* $\chi^2(n)$, *za* $n = 1, 2, ..., 30$, *ali u obliku:*
za $X \sim \chi^2(n)$ *i zadanu vjerojatnost* $p = P(X > x_p)$ *u tabeli možemo očitati vrijednosti* x_p.
Najčešće su tražene vrijednosti za x_p, *ako su zadane vjerojatnosti* $p = 0.99$,
$p = 0.95$, $p = 0.5$, $p = 0.1$, $p = 0.05$

PRIMJER 8.25 *Neka je* $X \sim \chi^2(20)$ *i* $P(X > x_p) = 0.1$. *Od koje ďž˝e vrijednosti slučajna varijabla poprimi veću vrijednost s vjerojatnošću 0.05?*
U tablici očitamo za $n = 20$, *i* $p = 0.1$, $x_p = 28.41$.

NAPOMENA 8.4 *Neka su slučajne varijable* $X_1, X_2, ..., X_n$ *takve da sve imaju standardnu normalnu distribuciju,* $X_i^* \sim N(0, 1)$, $i = 1, ..., n$.
Tada slučajna varijabla $Y = \sum_{i=1}^{n} (X_i)^2$ *ima hi kvadrat distribuciju,* $Y \sim \chi^2(n)$.

PRIMJER 8.26 *Za* $X \sim \chi^2(n),\ E(X) = n,\ Var(X) = 2n.$

Rješenje:
Koristimo formulu za očekivanje i varijancu gama distribucije s parametrima $\alpha = \frac{n}{2},\ \lambda = \frac{1}{2}:\ \ E(X) = \frac{\alpha}{\lambda} = \frac{\frac{n}{2}}{\frac{1}{2}} = n,\ Var(X) = \frac{\alpha}{\lambda^2} = \frac{\frac{n}{2}}{(\frac{1}{2})^2} = 2n.$

NAPOMENA 8.5 *Za* $X \sim \chi^2(n)$ *i* $n \to \infty,\ X \sim N(n, 2n).$
Za $X \sim \chi^2(n)$ *i* $n > 30$, *dobra aproksimacija je* $X \sim N(n, 2n)$.

8.6 STUDENTOVA DISTRIBUCIJA

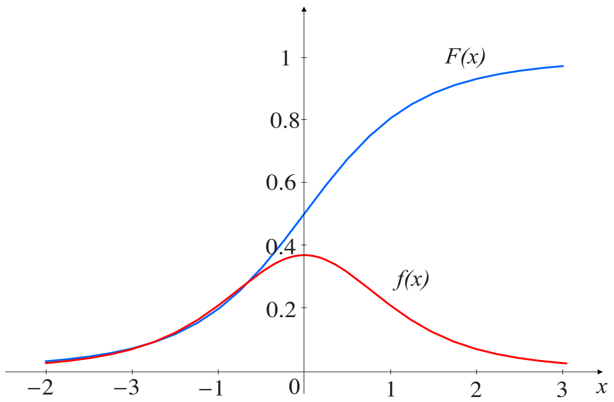
U matematičkoj statistici važna je Studentova distribucija koju je 1908 definirao S. Gosset pod pseudonimom Student.

Definicija 8.8 *(STUDENTOVA DISTRIBUCIJA)*

Za kontinuiranu slučajnu varijablu $X : \Omega \to \mathbb{R}$ kažemo da ima Studentovu distribuciju ili t -distribuciju s parametrom n (stupanj slobode) i označavamo $X \sim t(n)$ ako ima funkciju gustoće vjerojatnosti

$$f(x) = \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \cdot (1 + \frac{x^2}{n})^{-\frac{n+1}{2}}, \quad x \in R$$

PRIMJER 8.27 *Skicirati graf funkcije gustoće vjerojatnosti $f(x)$. Funkcija je pozitivna, simetrična u odnosu na os y, parna, dostiže maksimum u $x = 0$, os x je horizontalna asimptota. Kad $n \to \infty$, graf postaje Gaussova krivulja.*



Slika 8.6: Grafovi funkcije gustoće vjerojatnosti i funkcije distribucije sluďż˝ajne varijable $X \sim t(3)$.

PRIMJER 8.28 *Tabelira se $t(n)$, za $n = 1, 2, ..., 30$, ali u obliku:*
za $X \sim t(n)$ i zadanu vjerojatnost $p = P(|X| > x_p)$ u tabeli možemo očitati vrijednosti x_p. Najčešće su tražene vrijednosti za x_p, ako su zadane vjerojatnosti $p = 0.9$,
$p = 0.8,\ p = 0.7, ...,\ p = 0.1$.

PRIMJER 8.29 *Neka je $X \sim t(20)$ i $P(|X| > x_p) = 0.1$. Izvan kojih granica slučajna varijabla poprimi vrijednost s vjerojatnošću 0.1?*
U tablici očitamo za $n = 20$, i $p = 0.1$ $x_p = 1.725$.

PRIMJER 8.30 *Neka su slučajne varijable* $X^* \sim N(0,1)$ *i* $Y \sim \chi^2(n)$. *Tada slučajna varijabla* $T = \frac{X^*}{\sqrt{\frac{Y}{n}}}$ *ima Studentovu distribuciju,* $T \sim t(n)$.

NAPOMENA 8.6 *Za* $X \sim t(n)$, $E(X) = 0$, $Var(X) = \dfrac{n}{n-2}$, $n > 2$.

NAPOMENA 8.7 *Za* $X \sim t(n)$ *i* $n \to \infty$, $X \sim N(0,1)$.
Za $X \sim t(n)$ *i* $n > 30$, *dobra aproksimacija je* $X \sim N(0,1)$.

NAPOMENA 8.8 *Za* $n = 1$, *Studentova distribucija je Cauchyjeva distribucija.*

8.7 Ponovimo

STANDARDNA NORMALNA DISTRIBUCIJA (RAZDIOBA)

standardna normalna distribucija	$X^* \sim N(0,1)$
funkcija gustoće vjerojatnosti	$f^*(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2}$
funkcija distribucije vjerojatnosti	$F^*(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt.$
očekivanje	$E(X) = 0$
varijanca	$Var(X) = 1$

NORMALNA DISTRIBUCIJA

normalna distribucija	$X \sim N(\mu, \sigma^2)$
funkcija gustoće vjerojatnosti	$f(x) = \frac{1}{\sigma} \cdot f^*(\frac{x-\mu}{\sigma})$
funkcija distribucije vjerojatnosti	$F(x) = F^*(\frac{x-\mu}{\sigma})$
očekivanje	$E(X) = \mu$
varijanca	$Var(X) = \sigma^2$
PRAVILO 3σ	za $X \sim N(\mu, \sigma)$
	$P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.68$
	$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.95$
	$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.99$

UNIFORMNA DISTRIBUCIJA

uniformna distribucija	$X \sim U(a,b)$
funkcija gustoće vjerojatnosti	$f(x) = \frac{1}{b-a}$, za $a \leq x \leq b$,
funkcija distribucije vjerojatnosti	$F(x) = \frac{x-a}{b-a}$, za $a \leq x \leq b$,
očekivanje	$E(X) = \frac{a+b}{2}$
varijanca	$Var(X) = \frac{(b-a)^2}{12}$

EKSPONENCIJALNA DISTRIBUCIJA

eksponencijaln distribucija	$X \sim Exp(\lambda)$
funkcija gustoće vjerojatnosti	$f(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & 0 \leq x \end{cases}$
funkcija distribucije vjerojatnosti	$F(x) == \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & 0 \leq x \end{cases}$
očekivanje	$E(X) = \frac{1}{\lambda}$
varijanca	$Var(X) = \frac{1}{\lambda^2}$

GAMA DISTRIBUCIJA, HI-kvadrat, eksponencijalna

gama distribucija	$X \sim \Gamma(\alpha, \lambda)$
hi-kvadrat distribucija	$X \sim \chi^2(n) = \Gamma(\frac{n}{2}, \frac{1}{2})$
eksponencijalna distribucija	$X \sim Exp(\lambda) = \Gamma(1, \lambda)$
očekivanje	$E(X) = \frac{\alpha}{\lambda}$
varijanca	$Var(X) = \frac{\alpha}{\lambda^2}$
hi-kvadrat $Y \sim \chi^2(n)$	$Y = \sum_{i=1}^{n} (X_i)^2$
za	$X_i \sim N(0, 1)$

STUDENTOVA ili t- DISTRIBUCIJA

Studentova distribucija	$X \sim t(n)$
očekivanje	$E(X) = 0$
varijanca	$Var(X) = \frac{n}{n-2}$, za $n > 2$
Studentova razdioba $T \sim t(n)$	$T = \frac{X^*}{\sqrt{\frac{Y}{n}}}$,
ako je	$X^* \sim N(0, 1)$ i $Y \sim \chi^2(n)$.

Poglavlje 9

DVODIMENZIONALNI SLUČAJNI VEKTOR

Pojam slučajne varijable generalizira se za dvije i n varijabli. U ovom poglavlju posebno će se obraditi diskretni dvodimenzionali slučajni vektor. Pojam n-dim slučajnog vektora važan je za definiciju slučajnog uzorka.

Definicija 9.1 *(n-dim SLUČAJNI VEKTOR, engl. random vector)*

Neka su su $X_1 : \Omega \to \mathbb{R}$, $X_2 : \Omega \to \mathbb{R}$,..., $X_n : \Omega \to \mathbb{R}$ *slučajne varijable definirane na vjerojatnosnom prostoru* $(\Omega, \mathcal{F}, P)$. *Funkciju* $(X_1, X_2, ..., X_n) : \Omega \to \mathbb{R}^n$ *zovemo n-dim slučajni vektor ako vrijedi* $\forall x_1, x_2, ..., x_n \in \mathbb{R}$,

$$\{\omega \in \Omega : \ X_1(\omega) \leq x_1, \ X_2(\omega) \leq x_2, ..., \ X_n(\omega) \leq x_n\} \subset \mathcal{F}$$

i označavamo $(X_1, X_2, ..., X_n)(\omega) = (X_1(\omega), X_2(\omega), ..., X_n(\omega)), \ \omega \in \Omega.$

Definicija 9.2 *(FUNKCIJA DISTRIBUCIJE n-dim SLUČAJNOG VEKTORA, engl. distribution function of random vector)*

Neka je $(X_1, X_2, ..., X_n) : \Omega \to \mathbb{R}^n$ *n-dim slučajni vektor. Funkciju* $F : \mathbb{R}^n \to \mathbb{R}$ *definiranu na sljedeći način:*

$$F(x_1, x_2, ..., x_n) := P(X_1 \leq x_1, ..., X_n \leq x_n)$$

zovemo funkcija distribucije diskretnog n-dim slučajnog vektora $(X_1, X_2, .., X_n)$.

9.1 DISKRETNI DVODIMENZIONALNI SLUČAJNI VEKTOR

MOTIV 9.1 *Promatramo slučajni pokus bacanje 2 igraće kocke i slučajnu varijablu $X =$ suma brojeva koji su pali i slučajnu varijablu $Y =$ broj 1 ako su pali jednaki brojevi, inače 0. (a) Nađite funkciju vjerojatnosti slučajnog vektora (X, Y). (b)Izračunajte vjerojatnost da je zbroj brojeva koji su pali veći od 3 a manji ili jednak 6 i da su pali isti brojevi? (c) Jesu li X i Y nezavisne varijable? (d) Izračunajte kovarijancu. (e) Jesu li X i Y korelirane varijable?*

Definicija 9.3 *DISKRETNI n-dim SLUČAJNI VEKTOR*

Ako su slučajne varijable X_1, $X_2, ..., X_n$ diskretne slučajne varijable, onda za slučajni vektor $(X_1, X_2, ..., X_n)$ kažemo da je diskretni n-dim slučajni vektor.

Definicija 9.4 *(FUNKCIJA VJEROJATNOSTI n-dim SLUČAJNOG VEKTORA, engl. probability function of random vector)*

Neka je $(X_1, X_2, ..., X_n) : \Omega \to \mathbb{R}^n$ n-dim diskretni slučajni vektor. Funkciju $f : \mathbb{R}^n \to \mathbb{R}$ definiranu na sljedeći način:

$$f(x_1, x_2, ..., x_n) := \begin{cases} P(X_1 = x_1, X_2 = x_2, ...X_n = x_n), & (x_1, ..., x_n) \in \mathcal{R}(X_1, X_2, ..., X_n), \\ 0, & \textit{inače} \end{cases}$$

zovemo funkcija vjerojatnosti diskretnog n-dim slučajnog vektora $(X_1, X_2, ..., X_n)$.

Definicija 9.5 *DISKRETNI 2-dim SLUČAJNI VEKTOR*

Ako su slučajne varijable X i Y diskretne slučajne varijable, onda za slučajni vektor (X, Y) kažemo da je diskretni dvodimenzionalni slučajni vektor.

Definicija 9.6 *(FUNKCIJA VJEROJATNOSTI diskretnog 2-dim SLUČAJNOG VEKTORA)*

Neka su $X : \Omega \to \mathbb{R}$, $Y : \Omega \to \mathbb{R}$ diskretne slučajne varijable sa slikama $\mathcal{R}(X) = \{x_1, x_2, ...\}$, $\mathcal{R}(Y) = \{y_1, y_2, ...\}$ Funkciju $f : \mathbb{R}^2 \to \mathbb{R}$ definiranu na sljedeći način:

$$f(x, y) := \begin{cases} P(X = x_i, Y = y_j), & x = x_i \in \mathcal{R}(X),\ y = y_j \in \mathcal{R}(X), \\ 0, & \textit{inače} \end{cases}$$

zovemo funkcija vjerojatnosti diskretnog 2-dim slučajnog vektora (X, Y).

T: SVOJSTVA funkcije vjerojatnosti slučajne varijable X.
(i) $0 \leq f(x_i, y_j) \leq 1$,
(ii) $\sum\limits_{i=1}^{\infty} \sum\limits_{j=1}^{\infty} f(x_i, y_j) = 1$.
Dokaz: *tko želi znati više*

(i) Za $(x_i, y_j) \in \mathbb{R}^2$, $P(X = x_i, Y = y_j) = P(\{\omega : X(\omega) = x_i, Y(\omega) = y_j\})$
$\Rightarrow \quad 0 \leq f(x_i, y_i) \leq 1$.
(ii) Za $(x_i, y_j) \neq (x_k, y_l) \ \Rightarrow$
$\{\omega : X(\omega) = x_i, Y(\omega) = y_j\} \bigcap \{\omega : X(\omega) = x_k, Y(\omega) = y_l\} = \emptyset$,
svojstvo (ii) slijedi prema svojstvu (P3), prebrojive aditivnosti funkcije P:

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f(x_i, y_j) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P(X = x_i, Y = y_j) = P(\Omega) = 1.$$

NAPOMENA 9.1 *Diskretni 2-dim slučajni vektor je zadan sa svojom slikom* $\mathcal{R}((X, Y)) = \{(x_i, y_j), i = 1, ...; j = 1, ...\}$ *i funkcijom vjerojatnosti* f *slučajne varijable, vrijednostima* $\{f(x_i, y_j),\ i = 1, ...;\ j = 1, ...\}$.
U literaturi se zato može naći zapis slučajnog vektora (X, Y) *kao uređene sheme:*

$$(X, Y) \sim \begin{pmatrix} X/Y & y_1 & y_2 & \cdots & y_j & \cdots \\ x_1 & p_{11} & p_{12} & & p_{1j} & \cdots \\ x_2 & p_{21} & & & p_{2j} & \\ \cdots & & & & & \\ x_i & p_{i1} & & & p_{ij} & \\ \cdots & & & & \cdots & \cdots \end{pmatrix}$$

gdje su $p_{ij} = f(x_i, y_j),\ i, j \in \{1, 2, ...\}$.

PRIMJER 9.1 *Neka je* $(\Omega, \mathcal{P}(\Omega), P)$ *diskretni vjerojatnosni prostor slučajnog pokusa bacanje igraće kocke.* $\Omega = \{\omega_1, \omega_2, \omega_3, ..., \omega_6\}$, $P(\{\omega_i\}) = \frac{1}{6}$, $i = 1, .., 6$.
Neka je $X : \Omega \to \mathbb{R}$ *slučajna varijabla definirana na sljedeći način:* X = *broj koji je pao.*
Neka je $Y : \Omega \to \mathbb{R}$ *slučajna varijabla definirana na sljedeći način:* $Y = 1$ *ako je pao paran broj veći od 3, inače 0.*
X je slučajna varijabla. Slika slučajne varijable je skup $\mathcal{R}(X) = \{1, 2, 3, 4, 5, 6\}$.
Y je slučajna varijabla. Slika slučajne varijable je skup $\mathcal{R}(Y) = \{0, 1\}$.
Naći funkciju vjerojatnosti slučajnog vektora (X, Y).

Rješenje: Funkcija vjerojatnosti slučajne varijable X je $f : \mathbb{R}^2 \to \mathbb{R}$

$$
\begin{aligned}
f(x,y) \; &:= \; \begin{cases} P(X = x_i, Y = y_j), & x = x_i \in \mathcal{R}(X), y = y_j \in \mathcal{R}(Y) \\ \quad 0, & \text{inače.} \end{cases} \\
&= \; \begin{cases} \frac{1}{6}, & x = x_i \in \{4,6\},\; y = y_j \in \{1\} \\ \frac{1}{6}, & x = x_i \in \{1,2,3,5\},\; y = y_j \in \{0\} \\ 0, & \text{inače.} \end{cases}
\end{aligned}
$$

Slučajni vektor (X,Y) možemo za zadati s

$$
(X,Y) \sim \begin{pmatrix} X/Y & 0 & 1 \\ 1 & \frac{1}{6} & 0 \\ 2 & \frac{1}{6} & 0 \\ 3 & \frac{1}{6} & 0 \\ 4 & 0 & \frac{1}{6} \\ 5 & \frac{1}{6} & 0 \\ 6 & 0 & \frac{1}{6} \end{pmatrix}
$$

PRIMJER 9.2 *Neka je (X,Y) slučajni vektor za slučajni pokus izbora dva broja iz {1,2,3}, pri čemu je X= prvi izabrani broj, Y= izabrani broj nije manji od prvog. Naći funkciju vjerojatnosti slučajnog vektora.*

Rješenje:

$$
(X,Y) \sim \begin{pmatrix} X/Y & 1 & 2 & 3 \\ 1 & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ 2 & 0 & \frac{1}{6} & \frac{1}{6} \\ 3 & 0 & 0 & \frac{1}{3} \end{pmatrix}.
$$

Definicija 9.7 *(FUNKCIJA DISTRIBUCIJE DISKRETNOG 2-dim SLUČAJNOG VEKTORA)*

Neka je $(X,Y) : \Omega \to \mathbb{R}^2$ *diskretni 2-dim slučajni vektor sa slikom* $\mathcal{R}((X,Y)) = \{(x_i, y_j), i = 1,...; j = 1,...\}$. *Funkciju* $F : \mathbb{R}^2 \to \mathbb{R}$ *definiranu na sljedeći način:*

$$
F(x,y) := P(X \leq x, Y \leq y) = \sum_{i:\, x_i \leq x} \sum_{j:\, y_j \leq y} f(x_i, y_j)
$$

zovemo funkcija distribucije diskretnog 2-dim slučajnog vektora (X,Y).

T: SVOJSTVA funkcije distribucije diskretnog 2-dim slučajnog vektora:

(F1) $\lim\limits_{x\to-\infty,\, y\to-\infty} F(x,y) = F(-\infty,-\infty) = 0$

$\lim\limits_{x\to-\infty} F(x,y) = F(-\infty,y) = 0$

$\lim\limits_{y\to-\infty} F(x,y) = F(x,-\infty) = 0$

(F2) $\lim\limits_{x\to\infty,\, y\to\infty} F(x,y) = F(\infty,\infty) = 1$

(F3) $0 \le F(x,y) \le 1$

(F4)

$$P(a_1 < X \le b_1, a_2 < Y \le b_2) = F(b_1,b_2) - F(a_1,b_2) - F(b_1,a_2) + F(a_1,a_2),$$

$a_1, b_1, a_2, b_2 \in \mathbb{R},\ \ a_1 < b_1,\ a_2 < b_2.$

(F6) F je rastuća funkcija po svakoj varijabli.

Dokaz: *tko želi znati više*

(F1) Kako je događaj $X \le -\infty,\ Y \le -\infty$ nemoguć događaj

$$F(-\infty,-\infty) := P(X \le -\infty,\ Y \le -\infty) = P(\emptyset) = 0.$$

$$F(-\infty,y) = P(X \le -\infty \bigcap Y \le y) = P(\emptyset \bigcap Y \le y) = P(\emptyset) = 0.$$

(F2) Budući je događaj $X \le \infty,\ Y \le \infty$ siguran događaj, onda je

$$F(\infty,\infty) = P(\Omega) = 1.$$

(F3) tvrdnja slijedi iz svojstava (F1) i (F2).

(F4) Neka su $a, b \in \mathbb{R},\ a_1 < b_1,\ a_2 < b_2$. Računamo

$$\begin{aligned} & P(a_1 < X \le b_1, a_2 < Y \le b_2) \\ &\qquad = P(a_1 < X \le b_1, Y \le b_2) - P(a_1 < X \le b_1, Y \le a_2) \\ &\qquad = (F(b_1,b_2) - F(a_1,b_2)) - (F(b_1,a_2) - F(a_1,a_2)). \end{aligned}$$

(F6) Neka su $a_1, b_1 \in \mathbb{R},\ a_1 < b_1$.

$$\begin{aligned} P(a_1 < X \le b_1, Y \le y) &= F(b_1,y) - F(a_1,y) - F(b_1,-\infty) + F(a_1,-\infty) \\ &= F(b_1,y) - F(a_1,y) \ge 0. \end{aligned}$$

Iz svojstva (F5) slijedi da je $F(a_1,y) \le F(b_1,y)$, funkcija je rastuća po prvoj varijabli.

PRIMJER 9.3 *Za slučajni vektor* (X, Y) *iz primjera zadan s*

$$(X,Y) \sim \begin{pmatrix} X/Y & 0 & 1 \\ 1 & \frac{1}{6} & 0 \\ 2 & \frac{1}{6} & 0 \\ 3 & \frac{1}{6} & 0 \\ 4 & 0 & \frac{1}{6} \\ 5 & \frac{1}{6} & 0 \\ 6 & 0 & \frac{1}{6} \end{pmatrix}$$

odredite vrijednost funkcije distribucije $F(1,0), F(1,1)$, $F(2,0)$, $F(2,1)$, $F(4,0)$. *Izračunajte vjerojatnost* $P(1 < X \leq 2, 0 < Y \leq 1)$ *i* $P(1 < X \leq 4, Y \leq 0)$.

Rješenje:

$$F(x,y) = P(X \leq x, Y \leq y) = \sum_{i:\, x_i \leq x} \sum_{j:\, y_j \leq y} f(x_i, y_j)$$

$F(1,0) = f(1,0) = p_{11} = \frac{1}{6}$,
$F(1,1) = f(1,0) + f(1,1) = p_{11} + p_{12} = \frac{1}{6}$,
$F(2,0) = f(1,0) + f(2,0) = p_{11} + p_{21} = \frac{2}{6}$,
$F(2,1) = f(1,0) + f(1,1) + f(2,0) + f(2,1) = p_{11} + p_{12} + p_{21} + p_{22} = \frac{2}{6}$,
$F(4,0) = f(1,0) + f(2,0) + f(3,0) + f(4,0) = p_{11} + p_{21} + p_{31} + p_{41} = \frac{3}{6}$.
Koristimo formule

$$P(a_1 < X \leq b_1,\, a_2 < Y \leq b_2) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$$

$P(a_1 < X \leq b_1,\, Y \leq b_2) = F(b_1, b_2) - F(a_1, b_2)$

$$\begin{aligned} P(1 < X \leq 2,\, 0 < Y \leq 1) &= F(2,1) - F(1,1) - F(2,0) + F(1,0) \\ &= \frac{2}{6} - \frac{1}{6} - \frac{2}{6} + \frac{1}{6} = 0. \end{aligned}$$

$$P(1 < X \leq 4,\, Y \leq 0) = F(4,0) - F(1,0) = \frac{3}{6} - \frac{1}{6} = \frac{2}{6}.$$

ili
$P(1 < X \leq 2,\, 0 < Y \leq 1) = P(\emptyset) = 0$,
$P(1 < X \leq 4,\, Y \leq 0) = P(\{\omega_2\}) + P(\{\omega_3\}) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6}$.

PRIMJER 9.4 *Bacimo istovremeno dva novčića od 1 kune i od 5 kuna. Neka su X i Y slučajne varijable definirane:*

X=1 ako je pao slavuj, inače 0

Y= 1 ako je pao medo, inače 0.
Promatramo slučajni vektor (X,Y). Napišite funkciju vjerojatnosti slučajnog vektora. Odredite $F(1,1)$, $P(0 < X \leq 1, 0 < Y \leq 1)$ *i* $P(0 < X \leq 1, Y \leq 0)$.

Rješenje: $\mathcal{R}(X) = \{0,1\}$, $\mathcal{R}(Y) = \{0,1\}$.
$\mathcal{R}((X,Y)) = \{(x_i, y_j),\ i = 1,2;\ j = 1,2\} = \{(0,0),(1,0),(0,1),(1,1)\}$.

$$f(x,y) := \begin{cases} \frac{1}{4}, & x = x_i \in \{0,1\},\ y = y_j \in \{0,1\} \\ 0, & \text{inače.} \end{cases}$$

$$(X,Y) \sim \begin{pmatrix} X/Y & 0 & 1 \\ 0 & \frac{1}{4} & \frac{1}{4} \\ 1 & \frac{1}{4} & \frac{1}{4} \end{pmatrix}, \quad f(0,0) = f(1,0) = f(0,1) = f(1,1) = \tfrac{1}{4}.$$

$$F(x,y) := P(X \leq x, Y \leq y) = \sum_{i:\, x_i \leq x} \sum_{j:\, y_j \leq y} f(x_i, y_j)$$

$F(0,0) = f(0,0) = p_{11} = \frac{1}{4}$
$F(0,1) = f(0,0) + f(0,1) = p_{11} + p_{12} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.
$F(1,0) = f(0,0) + f(1,0) = p_{11} + p_{21} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$
$F(1,1) = f(0,0) + f(0,1) + f(1,0) + f(1,1) = p_{11} + p_{12} + p_{21} + p_{22} = 1$
Koristimo formule

$$P(a_1 < X \leq b_1, a_2 < Y \leq b_2) = F(b_1,b_2) - F(a_1,b_2) - F(b_1,a_2) + F(a_1,a_2)$$

$P(a_1 < X \leq b_1, Y \leq b_2) = F(b_1,b_2) - F(a_1,b_2)$

$$\begin{aligned} P(0 < X \leq 1,\ 0 < Y \leq 1) &= F(1,1) - F(0,1) - F(1,0) + F(0,0) \\ &= 1 - \frac{1}{2} - \frac{1}{2} + \frac{1}{4} = \frac{1}{4}. \end{aligned}$$

$$P(0 < X \leq 1, Y \leq 0) = F(1,0) - F(0,0) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

ili
$P(0 < X \leq 1,\ 0 < Y \leq 1) = P(\{(s,m)\}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$,
$P(0 < X \leq 1,\ Y \leq 0) = P(\{s,5\}\}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

Prisjetimo se motivacijskog primjera:

PRIMJER 9.5 *motiv*

Promatramo slučajni pokus bacanje 2 igraće kocke i slučajnu varijablu X = suma brojeva koji su pali i slučajnu varijablu Y = broj 1 ako su pali jednaki brojevi, inače 0. (a) Nađite funkciju vjerojatnosti slučajnog vektora (X,Y). (b) Izračunajte vjerojatnost da je zbroj brojeva koji su pali veći od 3 a manji ili jednak 6 i da su pali isti brojevi?

Rješenje: (a) $(X,Y) \sim \begin{pmatrix} X/Y & 0 & 1 \\ 2 & 0 & \frac{1}{36} \\ 3 & \frac{2}{36} & 0 \\ 4 & \frac{2}{36} & \frac{1}{36} \\ 5 & \frac{4}{36} & 0 \\ 6 & \frac{4}{36} & \frac{1}{36} \\ 7 & \frac{6}{36} & 0 \\ 8 & \frac{4}{36} & \frac{1}{36} \\ 9 & \frac{4}{36} & 0 \\ 10 & \frac{2}{36} & \frac{1}{36} \\ 11 & \frac{2}{36} & 0 \\ 12 & 0 & \frac{1}{36} \end{pmatrix}$

(b) $P(3 < X \leq 6, Y = 1) = f(4,1) + f(5,1) + f(6,1) = \frac{1}{36} + 0 + \frac{1}{36} = \frac{2}{36}$.

Definicija 9.8 *(MARGINALNA FUNKCIJA VJEROJATNOSTI komponenata 2-dim SLUČAJNOG VEKTORA, engl. marginal distribution)*

Neka je $(X,Y) : \Omega \to \mathbb{R}^2$ diskretni 2-dim slučajni vektor sa slikom $\mathcal{R}((X,Y)) = \{(x_i, y_j),\ i = 1, ...;\ j = 1, ...\}$ i funkcijom vjerojatnosti
$f : \mathbb{R}^2 \to [0,1]$

$$f(x,y) := \begin{cases} P(X = x_i,\ Y = y_j), & (x,y) = (x_i, y_j) \in \mathcal{R}((X,Y)) \\ 0, & \textit{inače.} \end{cases}$$

Funkciju $f_1(x) = f(X = x,\ Y = proizvoljno) = \sum_j f(x, y_j)$, *zovemo marginalna funkcija vjerojatnosti komponente X slučajnog vektora (X,Y).*
Funkciju $f_2(y) = f(X = proizvoljno,\ Y = y) = \sum_i f(x_i, y)$ *zovemo marginalna funkcija vjerojatnosti komponente Y slučajnog vektora (X,Y).*

$$\begin{pmatrix} X/Y & y_1 & y_2 & \dots & y_j & \dots & f_1(x) \\ x_1 & p_{11} & p_{12} & \dots & p_{1j} & \dots & \sum_j p_{1j} \\ x_2 & p_{21} & & & p_{2j} & \dots & \sum_j p_{2j} \\ \dots & \dots & & & \dots & & \dots \\ x_i & p_{i1} & & & p_{ij} & & \sum_j p_{ij} \\ \dots & \dots & \dots & & \dots & \dots & \dots \\ f_2(y) & \sum_i p_{i1} & \dots & & \sum_i p_{ij} & & \sum_i \sum_j p_{ij} = 1 \end{pmatrix}$$

PRIMJER 9.6 *U prethodnim primjerima slučajnih vektora odredite marginalne funkcije vjerojatnosti komponenti X i Y slučajnog vektora (X, Y).*

Rješenje:

(a) $$\begin{pmatrix} X/Y & 0 & 1 & f_1(x) \\ 1 & \frac{1}{6} & 0 & \frac{1}{6} \\ 2 & \frac{1}{6} & 0 & \frac{1}{6} \\ 3 & \frac{1}{6} & 0 & \frac{1}{6} \\ 4 & 0 & \frac{1}{6} & \frac{1}{6} \\ 5 & \frac{1}{6} & 0 & \frac{1}{6} \\ 6 & 0 & \frac{1}{6} & \frac{1}{6} \\ f_2(y) & \frac{4}{6} & \frac{2}{6} & 1 \end{pmatrix}, \qquad X \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}, \qquad Y \sim \begin{pmatrix} 0 & 1 \\ \frac{4}{6} & \frac{2}{6} \end{pmatrix}$$

(b) $$\begin{pmatrix} X/Y & 1 & 2 & 3 & f_1(x) \\ 1 & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{6}{18} \\ 2 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{6}{18} \\ 3 & 0 & 0 & \frac{1}{3} & \frac{6}{18} \\ f_2(y) & \frac{2}{18} & \frac{5}{18} & \frac{11}{18} & 1 \end{pmatrix}$$

(c) $$\begin{pmatrix} X/Y & 0 & 1 & f_1(x) \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ 1 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ f_2(y) & \frac{1}{2} & \frac{1}{2} & 1 \end{pmatrix}$$

(d) $$\begin{pmatrix} X/Y & 0 & 1 & f_1(x) \\ 2 & 0 & \frac{1}{36} & \frac{1}{36} \\ 3 & \frac{2}{36} & 0 & \frac{2}{36} \\ 4 & \frac{2}{36} & \frac{1}{36} & \frac{3}{36} \\ 5 & \frac{4}{36} & 0 & \frac{4}{36} \\ 6 & \frac{4}{36} & \frac{1}{36} & \frac{5}{36} \\ 7 & \frac{6}{36} & 0 & \frac{6}{36} \\ 8 & \frac{4}{36} & \frac{1}{36} & \frac{5}{36} \\ 9 & \frac{4}{36} & 0 & \frac{4}{36} \\ 10 & \frac{2}{36} & \frac{1}{36} & \frac{3}{36} \\ 11 & \frac{2}{36} & 0 & \frac{2}{36} \\ 12 & 0 & \frac{1}{36} & \frac{1}{36} \\ f_2(y) & \frac{30}{36} & \frac{6}{36} & 1 \end{pmatrix}$$

Definicija 9.9 *(MARGINALNA FUNKCIJA DISTRIBUCIJE VJEROJATNOSTI komponenata 2-dim SLUČAJNOG VEKTORA, engl. marginal distribution)*

Neka je $(X,Y) : \Omega \to \mathbb{R}^2$ *diskretni 2-dim slučajni vektor sa slikom* $\mathcal{R}((X,Y)) = \{(x_i, y_j), i = 1, ...; \ j = 1, ...\}$ *i funkcijom distribucije vjerojatnosti* $F : \mathbb{R}^2 \to [0,1]$

$$F(x,y) := P(X \leq x, Y \leq y) = \sum_{i:\, x_i \leq x} \sum_{j:\, y_j \leq y} f(x_i, y_j)$$

Funkciju $F_1(x) := F(x, \infty) = P(X \leq x, -\infty < Y \leq \infty) = \sum_{u \leq x} f_1(u),$ *zovemo marginalna funkcija distribucije vjerojatnosti komponente* X *slučajnog vektora* (X,Y).
Funkciju $F_2(y) := F(\infty, y) = P(-\infty < X \leq \infty, Y \leq y) = \sum_{u \leq y} f_2(u)$ *zovemo marginalna funkcija distribucije vjerojatnosti komponente* Y *slučajnog vektora* (X,Y).

Definicija 9.10 *(NEZAVISNE SLUČAJNE VARIJABLE 2-dim SLUČAJNOG VEKTORA, engl. independent variables)*

Neka je $(X,Y) : \Omega \to \mathbb{R}^2$ *2-dim slučajni vektor sa funkcijom distribucije vjerojatnosti* $F(x,y)$. *Za slučajne varijable* X *i* Y *kažemo da su nezavisne ako je za svaki* (x,y)

$$F(x,y) = F_1(x) \cdot F_2(y).$$

Neka je $(X, Y) : \Omega \to \mathbb{R}^2$ *diskretni 2-dim slučajni vektor sa funkcijom vjerojatnosti* $f(x, y)$*. Za slučajne varijable X i Y kažemo da su nezavisne ako je za svaki* (x, y)

$$f(x, y) = f_1(x) \cdot f_2(y).$$

Inače su X i Y zavisne.

PRIMJER 9.7 *U primjerima provjeri jesu li slučajne varijable u slučajnim vektorima (X,Y) nezavisne.*

Rješenje:
(a) nisu jer je npr. $f(4, 1) \neq f_1(4) \cdot f_2(1)$, $f(4, 1) = \frac{1}{6}$, $f_1(4) = \frac{1}{6}$, $f_2(1) = \frac{2}{6}$;
(b) nisu jer je npr. $f(1, 1) \neq f_1(1) \cdot f_2(1)$, $f(1, 1) = \frac{1}{9}$, $f_1(4) = \frac{6}{18}$, $f_2(1) = \frac{2}{18}$;
(c) jesu jer je $f(x_i, y_j) = f_1(x_i) \cdot f_2(y_j)$, $i, j = 1, 2$;
(d) nisu jer je npr. $f(6, 1) \neq f_1(6) \cdot f_2(1)$, $f(6, 1) = \frac{1}{36}$, $f_1(6) = \frac{5}{36}$, $f_2(1) = \frac{6}{36}$.

Definicija 9.11 *(OČEKIVANJE 2-dim SLUČAJNOG VEKTORA)*
Neka je $(X, Y) : \Omega \to \mathbb{R}^2$ *2-dim slučajni vektor i neka komponente X i Y su slučajne varijable koje imaju očekivnje E(X) i E(Y) onda se matematičko očekivanje slučajnog vektora definira s*

$$E(X, Y) = (E(X), E(Y)).$$

Definicija 9.12 *(OČEKIVANJE funkcije DISKRETNOG 2-dim SLUČAJNOG VEKTORA)*
Neka je $(X, Y) : \Omega \to \mathbb{R}^2$ *diskretni 2-dim slučajni vektor sa slikom* $\mathcal{R}((X, Y)) = \{(x_i, y_j), i = 1, .; j = 1, ...\}$ *i funkcijom vjerojatnosti* $f : \mathbb{R}^2 \to [0, 1]$

$$f(x, y) := \begin{cases} P(X = x_i, Y = y_j), & (x, y) = (x_i, y_j) \in \mathcal{R}((X, Y)) \\ 0, & inače. \end{cases}$$

Neka je zadana funkcija $h : R^2 \to R$*. Kažemo da je slučajna varijabla* $h(X, Y)$ *funkcija od slučajnog vektora (X,Y) i za nju definiramo očekivanje ako red* $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} h(x_i, y_j) f(x_i, y_j)$ *konvergira i označavamo*

$$E(h(X, Y)) := \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} h(x_i, y_j) f(x_i, y_j).$$

PRIMJER 9.8 $E(X \cdot Y) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} x_i \cdot y_j f(x_i, y_j)$ *za* $h(x, y) = x \cdot y$.

PRIMJER 9.9 *Ako su* X *i* Y *nezavisne slučajne varijable onda je*

$$E(X \cdot Y) = E(X) \cdot E(Y),$$

jer

$$\begin{aligned} E(X \cdot Y) &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} x_i \cdot y_j f(x_i, y_j) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} x_i \cdot y_j f_1(x_i) \cdot f_2(y_j) \\ &= \sum_{i=1}^{\infty} x_i f_1(x_i) \cdot \sum_{j=1}^{\infty} y_j f_2(y_j) = E(X) \cdot E(Y). \end{aligned}$$

PRIMJER 9.10 *Ako je* $(X_1, X_2, ..., X_n)$ *n-dim slučajni vektor i sve varijable nezavisne onda*

$$E(X_1 \cdot X_2 \cdot ... \cdot X_n) = E(X_1) \cdot E(X_2) \cdot ... \cdot E(X_n).$$

PRIMJER 9.11 $E(X + Y) = E(X) + E(Y)$ *za* $h(x, y) = x + y$.

PRIMJER 9.12 *Ako je* $(X_1, X_2, ..., X_n)$ *n-dim slučajni vektor onda*

$$E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n).$$

PRIMJER 9.13 *Neka je* (X, Y) *diskretni 2-dim slučajni vektor takav da* Y *ne poprima vrijednost 0. Odredi* $E(\sin \frac{X}{Y})$.

Rješenje: $h(x, y) = \sin \frac{x}{y}$

$$E(\sin \frac{X}{Y}) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} h(x_i, y_j) f(x_i, y_j) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \sin \frac{x_i}{y_j} f(x_i, y_j)$$

9.2 KONTINUIRANI 2-dim SLUČAJNI VEKTOR *tko želi znati više*

MOTIV 9.2 *Mladić i djevojka su dogovorili sastanak na trgu u 12 sati. Čekat će se najdulje 20 minuta nakon dolaska. Kolika je vjerojatnost da se susretnu ako su oboje došli na trg od 12 do 13 sati?*

Definicija 9.13 *KONTINUIRANI 2-dim SLUČAJNI VEKTOR*

Ako su slučajne varijable X i Y kontinuirane slučajne varijable, onda za slučajni vektor (X, Y) kažemo da je kontinuirani dvodimenzionalni slučajni vektor.

Definicija 9.14 *(FUNKCIJA GUSTOĆE VJEROJATNOSTI kontinuiranog 2-dim SLUČAJNOG VEKTORA)*

Neka su $X : \Omega \to \mathbb{R}$, $Y : \Omega \to \mathbb{R}$ kontinuirane slučajne varijable sa slikama $\mathcal{R}(X)$ i $\mathcal{R}(Y)$. Funkciju $f : \mathbb{R}^2 \to \mathbb{R}$ definiranu na sljedeći način:
(i) $0 \leq f(x, y)$
(ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dxdy = 1.$
zovemo funkcija gustoće vjerojatnosti kontinuiranog 2-dim slučajnog vektora (X, Y).

Definicija 9.15 *(FUNKCIJA DISTRIBUCIJE kontinuiranog 2-dim SLUČAJNOG VEKTORA)*

Neka je $(X, Y) : \Omega \to \mathbb{R}^2$ kontinuirani 2-dim slučajni vektor sa slikom $\mathcal{R}((X, Y))$. Funkciju $F : \mathbb{R}^2 \to \mathbb{R}$ definiranu na sljedeći način:

$$F(x, y) := P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x, y) dxdy$$

zovemo funkcija distribucije kontinuiranog 2-dim slučajnog vektora (X, Y).

NAPOMENA 9.2 *Neka je $(X, Y) : \Omega \to \mathbb{R}^2$ kontinuirani 2-dim slučajni vektor sa slikom $\mathcal{R}((X, Y)) = [a, b] \times [c, d]$.Tada vrijedi*

$$P(a < X \leq b, c < Y \leq d) = \int_{a}^{b} \int_{c}^{d} f(x, y) dxdy.$$

NAPOMENA 9.3 *Neka je $(X, Y) : \Omega \to \mathbb{R}^2$ kontinuirani 2-dim slučajni vektor sa slikom $\mathcal{R}((X, Y))$. Neka je događaj A vezan uz slučajni pokus i vektor (X, Y). Ako označimo područje u xy ravnini koje odgovara događaju A sa D_A onda vjerojatnost događaja A (geometrijsku vjerojatnost) možemo računati pomoću*

$$P(A) = \int\int_{D_A} f(x, y)dxdy.$$

MOTIV 9.3 *Mladić i djevojka su dogovorili sastanak na trgu u 12 sati. Čekat će se najdulje 20 minuta nakon dolaska. Kolika je vjerojatnost da se susretnu ako su oboje došli na trg od 12 do 13 sati?*

Rješenje:

Neka su X i Y slučajne varijable koje predtavljaju vrijeme dolaska djevojke i mladića. Pretpostavit ćemo da jednaki vremenski intervali imaju jednaku vjerojatnost, pa su funkcije gustoće vjerojatnosti za obe varijable jednake:

$$f_1(x) = \begin{cases} 1, & 0 \leq x \leq 60 \\ 0, & \text{inače.} \end{cases}$$

$$f_2(x) = \begin{cases} 1, & 0 \leq x \leq 60 \\ 0, & \text{inače.} \end{cases}$$

Budući su slučajne varijable X i Y nezavisne možemo odrediti funkciju gustoće vjerojatnosti slučajnog vektora (X, Y) :

$$f(x, y) = f_1(x) \cdot f_2(y)$$

$$f(x, y) = \begin{cases} 1, & 0 \leq x \leq 60, 0 \leq y \leq 60 \\ 0, & \text{inače.} \end{cases}$$

Računamo vjerojatnost događaja $A = \{(x, y) \in [0, 60] \times [0, 60] : |x - y| < 20\}$:
$P(A) = P(|X - Y| < 20) = \int\int_{D_A} f(x, y)dxdy = \frac{5}{9}$.
(vidite PRIMJER 2.20 u 2. poglavlju.)

9.3 KOVARIJANCA, KORELACIJA, PRAVCI REGRESIJE

MOTIV 9.4 *Promatramo slučajni pokus bacanje 2 igraće kocke i slučajnu varijablu X = suma brojeva koji su pali i slučajnu varijablu Y = broj 1 ako su pali jednaki brojevi, inače 0. (c) Jesu li X i Y nezavisne varijable? (d) Izračunajte kovarijancu. (e) Jesu li X i Y korelirane varijable?*

Definicija 9.16 *(KOVARIJANCA SLUČAJNOG VEKTORA - engl.covariance of X and Y)*

Neka je $(X, Y) : \Omega \to \mathbb{R}^2$ 2-dim slučajni vektor. Kovarijanca slučajnog vektora (X, Y) je

$$\mu_{XY} = E((X - E(X))(Y - E(Y))) = E(X \cdot Y) - E(X) \cdot E(Y).$$

Ako su X i Y nezavisne slučajne varijable onda je $\mu_{XY} = 0$.

PRIMJER 9.14 *(a) $Var(X + Y) = Var(X) + Var(Y) + 2\mu_{XY}$*
(b) Ako su X i Y nezavisne slučajne varijable onda je $Var(X + Y) = Var(X) + Var(Y)$.

Dokaz: *tko želi znati više*

Ako je $h(x, y) = x + y$ onda je $Z = X + Y$ slučajna varijabla, funkcija slučajnog vektora (X, Y).

$$\begin{aligned}
Var(Z) &= E(Z^2) - (E(Z))^2 = E((X + Y)^2) - (E(X + Y))^2 \\
&= E(X^2 + 2X \cdot Y + Y^2) - (E(X) + E(Y))^2 \\
&= E(X^2) + 2E(X \cdot Y) + E(Y^2) - ((E(X))^2 + 2E(X) \cdot E(Y) + (E(Y))^2) \\
&= E(X^2) - (E(X))^2 + E(Y^2) - (E(Y))^2 + 2(E(X \cdot Y) - E(X) \cdot E(Y))
\end{aligned}$$

Ako su X i Y nezavisne slučajne varijable onda je $\mu_{XY} = 0$ pa tvrdnja slijedi iz (a).

Definicija 9.17 *(KOEFICIJENT KORELACIJE, engl. correlation coefficient)*

Neka je $(X, Y) : \Omega \to \mathbb{R}^2$ 2-dim slučajni vektor Neka je μ_{XY} kovarijanca slučajnih varijabli X i Y, a σ_1 i σ_2 njihove standardne devijacije. Koeficijent korelacije komponenti X i Y slučajnog vektora je definiran s

$$\rho_{XY} = \frac{\mu_{XY}}{\sigma_1 \cdot \sigma_2}.$$

T: SVOJSTVA KOEFICIJENTA KORELACIJE ρ_{XY} :
(i) $\rho_{XY} = \rho_{YX}$,
(ii) $-1 \leq \rho_{XY} \leq 1$,
(iii) $\rho_{XY} = 1$ ako je $Y = aX + b,\ a > 0$, $\rho_{XY} = -1$ ako je $Y = aX + b,\ a < 0$,
(iv) $\rho_{UW} = \rho_{XY}$, ako je $U = aX + b,\ W = cY + d,\ a, c \neq 0$.

D: *tko želi znati više*

(ii) Za slučajnu varijablu $U = \sigma_2 X - \sigma_1 Y$, odredimo varijancu:

$$\begin{aligned}
Var(U) &= E(U^2) - (E(U))^2 \\
&= E(\sigma_2^2 X^2 - 2\sigma_1\sigma_2 XY + \sigma_1^2 Y^2) - (\sigma_2 E(X) - \sigma_1 E(Y))^2 \\
&= \sigma_2^2 E(X^2) - 2\sigma_1\sigma_2 E(XY) + \sigma_1^2 E(Y^2) \\
&\quad -(\sigma_2^2 (E(X))^2 - 2\sigma_1\sigma_2 E(X)E(Y) + \sigma_1^2 (E(Y))^2) \\
&= \sigma_2^2 Var(X) + \sigma_1^2 Var(Y) - 2\sigma_1\sigma_2 (E(XY) - E(X)E(Y)) \\
&= \sigma_2^2 \cdot \sigma_1^2 + \sigma_1^2 \cdot \sigma_2^2 - 2\sigma_1\sigma_2 \mu_{XY} \\
&= 2\sigma_1^2 \cdot \sigma_2^2 - 2\sigma_1\sigma_2 \mu_{XY}.
\end{aligned}$$

Budući je varijanca pozitivan broj zaključujemo da je $\mu_{XY} \leq \sigma_1\sigma_2,\ \ \rho_{XY} \leq 1$.
Analogno, promatrajući slučajnu varijablu $U = \sigma_2 X + \sigma_1 Y$ dobit ćemo nejednakost $\mu_{XY} \geq -\sigma_1\sigma_2,\ \ \rho_{XY} \geq -1$.
(iii)

$$\begin{aligned}
\mu_{XY} &= E(XY) - E(X)E(Y) = E(X(aX + b)) - E(X)E(aX + b) \\
&= aE(X^2) + bE(X) - a(E(X))^2 + bE(X) = aVar(X) = a\sigma_1^2,
\end{aligned}$$

$$\sigma_2^2 = Var(Y) = Var(aX + b) = a^2 Var(X) = a^2\sigma_1^2,$$

$$\rho_{XY} = \frac{\mu_{XY}}{\sigma_1 \cdot \sigma_2} = \frac{a\sigma_1^2}{\sigma_1 \cdot \sigma_1 |a|} = \frac{a}{|a|}.$$

Definicija 9.18 *(NEKORELIRANE SLUČAJNE VARIJABLE)*
Neka je $(X, Y) : \Omega \to \mathbb{R}^2$ 2-dim slučajni vektor Neka je μ_{XY} kovarijanca slučajnih varijabli X i Y, a σ_1 i σ_2 njihove standardne devijacije.
Za slučajne varijable X i Y kažemo da su nekorelirane ako je koeficijent korelacije $\rho_{XY} = 0$.
Ako je $\rho_{XY} \neq 0$ slučajne varijable su korelirane.

PRIMJER 9.15 *(a) Ako su X i Y nezavisne slučajne varijable onda su one nekorelirane. (b) Ako su X i Y nekorelirane slučajne varijable onda one ne moraju biti nezavisne.*

Dokaz:
(a) X i Y nezavisne $\Rightarrow \mu_{XY} = 0 \;\;\Rightarrow\;\; \rho_{XY} = 0 \;\;\Rightarrow\;\; X$ i Y nekorelirane.
(b) X i Y nekorelirane $\Rightarrow \rho_{XY} = 0$ ali ne moraju biti X i Y nezavisne (mogu se naći primjeri).

NAPOMENA 9.4 *Neka je zadan slučajni vektor (X, Y) s funkcijom vjerojatnosti $f(x, y)$, ili funkcija gustoće vjerojatnosti. Tražimo funkcijsku vezu između slučajnih varijabli X i Y npr. $Y = g(X)$ tako da $E((Y - g(X)^2)) \to \min$. Tada se krivulja dana jednadžbom $y = g(x)$ zove regresijska krivulja od Y po X (u smislu najmanjih kvadrata) i određuje se kao*

$$y = g(x) = E(Y|X = x) = \sum_{j=1} y_j \frac{f(x, y_j)}{f_1(x}$$

ili

$$y = g(x) = E(Y|X = x) = \int_{-\infty}^{\infty} y \frac{f(x, y)}{f_1(x)}$$

Ako pretpostavimo linearnu ovisnost Y od X tako da je $g(x) = ax + b$ onda kao regresijsku krivulju dobijemo pravac:

$$y - \mu_2 = \rho_{XY} \cdot \frac{\sigma_2}{\sigma_1}(x - \mu_1)$$

je pravac regresije Y po X.

Koeficijent $a = \rho_{XY} \cdot \dfrac{\sigma_2}{\sigma_1} = \dfrac{\mu_{XY}}{\sigma_1^2}$ je koeficijent regresije Y po X.

PRIMJER 9.16 *Neka je $(X, Y) : \Omega \to \mathbb{R}^2$ 2-dim slučajni vektor. Neka je μ_{XY} kovarijanca slučajnih varijabli X i Y, ρ_{XY} koeficijent korelacije, σ_1 i σ_2 njihove standardne devijacije, a μ_1, μ_2 njihova očekivanja. Pretostavimo da su X i Y zavisne slučajne varijable takve da je $Y = aX + b$. Parametre a i b možemo odrediti metodom najmanjih kvadrata tako da $E((Y - (aX + b))^2)$ ima minimalnu vrijednost.*

$a = \rho_{XY} \cdot \dfrac{\sigma_2}{\sigma_1} = \dfrac{\mu_{XY}}{\sigma_1^2}$ *je koeficijent regresije Y po X.*

$b = \mu_2 - \rho_{XY} \cdot \dfrac{\sigma_2}{\sigma_1}\mu_1 = \mu_2 - \dfrac{\mu_{XY}}{\sigma_1^2} \cdot \mu_1.$

$y - \mu_2 = \dfrac{\mu_{XY}}{\sigma_1^2}(x - \mu_1),$

$y-\mu_2=\rho_{XY}\cdot\frac{\sigma_2}{\sigma_1}(x-\mu_1)$ *je pravac regresije Y po X.*

Analogno, ako je X=aY+b

$a=\rho_{XY}\cdot\frac{\sigma_1}{\sigma_2}=\frac{\mu_{XY}}{\sigma_2^2}$ *je koeficijent regresije X po Y*

$x-\mu_1=\frac{\mu_{XY}}{\sigma_2^2}(Y-\mu_2),$

$x-\mu_1=\rho_{XY}\cdot\frac{\sigma_1}{\sigma_2}(y-\mu_2)$ *je pravac regresije X po Y.*

Dokaz: *tko želi znati više*
(metoda najmanjih kvadrata)
Označimo

$$\begin{aligned}
K(a,b) &= E((Y-aX-b)^2)=E(Y-\mu_2-a(X-\mu_1)-b+\mu_2-a\mu_1)^2 \\
&= \sigma_2^2+a^2\sigma_1^2-2a\mu_{XY}+(\mu_2-b-a\mu_1)^2
\end{aligned}$$

$K(a,b)$ je funkcija od 2 varijable i tražimo njen ekstrem. Nužni uvjeti su

$$\frac{\partial K}{\partial a}=2a\sigma_1^2-2\mu_{XY}-2(\mu_2-b-a\mu_1)=0,\quad \frac{\partial K}{\partial b}=-2(\mu_2-b-a\mu_1)=0.$$

Rješavanjem ovog sustava od 2 jed. s 2 nepoznanice dobit ćemo

$$a=\frac{\mu_{XY}}{\sigma_1^2},\quad b=\mu_2-a\mu_1=\mu_2-\frac{\mu_{XY}}{\sigma_1^2}\cdot\mu_1.$$

NAPOMENA 9.5 *Pravci regresije sijeku se u točki* (μ_1,μ_2) *koja se zove centar zajedničke distribucije X i Y.*

Kut između pravaca regresije $tg\varphi=\frac{1-\rho_{XY}^2}{\rho_{XY}}\cdot\frac{\sigma_1\cdot\sigma_2}{\sigma_1^2+\sigma_2^2}$.

Ako je $\rho_{XY}=1$ *ili* -1 *pravci regresije se poklapaju i tada postoji potpuna linearna zavisnost između X i Y.*
Ako je $\rho_{XY}=0$ *onda su pravci regresije* $y=\mu_2$, $x=\mu_1$ *okpmiti, a ne postoji linearna zavisnost slučajnih varijabli X i Y.*

PRIMJER 9.17 *Za primjer slučajnog vektora* (X,Y) *nađite kovarijancu, koeficijent korelacije i pravce regresije.*

$$\begin{pmatrix} X/Y & 1 & 2 & 3 & f_1(x) & \sum y_j p_{ij} & x_i \sum y_j p_{ij} \\ 1 & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{6}{18} & \frac{6}{9} & \frac{6}{9} \\ 2 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{6}{18} & \frac{5}{6} & \frac{10}{6} \\ 3 & 0 & 0 & \frac{1}{3} & \frac{6}{18} & \frac{3}{3} & \frac{9}{3} \\ f_2(y) & \frac{2}{18} & \frac{5}{18} & \frac{11}{18} & 1 & & \frac{96}{18} \\ \sum x_i p_{ij} & \frac{1}{9} & \frac{8}{18} & \frac{26}{18} & & & \\ y_j \sum x_i p_{ij} & \frac{1}{9} & \frac{16}{18} & \frac{78}{18} & \frac{96}{18} & & \end{pmatrix}$$

$$\mu_1 = E(X) = \sum x_i f_1(x_i) = 1 \cdot \frac{6}{18} + 2 \cdot \frac{6}{18} + 3 \cdot \frac{6}{18} = \frac{36}{18}$$

$$\mu_2 = E(Y) = \sum y_j f_2(y_j) = 1 \cdot \frac{2}{18} + 2 \cdot \frac{5}{18} + 3 \cdot \frac{11}{18} = \frac{45}{18}$$

$$E(X^2) = \sum x_i^2 f_1(x_i) = 1 \cdot \frac{6}{18} + 4 \cdot \frac{6}{18} + 9 \cdot \frac{6}{18} = \frac{84}{18}$$

$$E(Y^2) = \sum y_j^2 f_2(y_j) = 1 \cdot \frac{2}{18} + 4 \cdot \frac{5}{18} + 9 \cdot \frac{11}{18} = \frac{121}{18}$$

$$\sigma_1^2 = Var(X) = \frac{84}{18} - (\frac{36}{18})^2 = \frac{2}{3}$$

$$\sigma_2^2 = Var(Y) = \frac{121}{18} - (\frac{45}{18})^2 = \frac{17}{36}$$

Kovarijanca je $\mu_{XY} = E(XY) - E(X)E(Y) = \frac{96}{18} - \frac{36}{18} \cdot \frac{45}{18} = \frac{1}{3}$.

Pravac regresije Y po X je $y - \mu_2 = \frac{\mu_{XY}}{\sigma_1^2}(x - \mu_1)$, $y - \frac{45}{18} = \frac{\frac{1}{3}}{\frac{2}{3}}(x - \frac{36}{18})$,

$y = \frac{1}{2}x + \frac{3}{2}$.

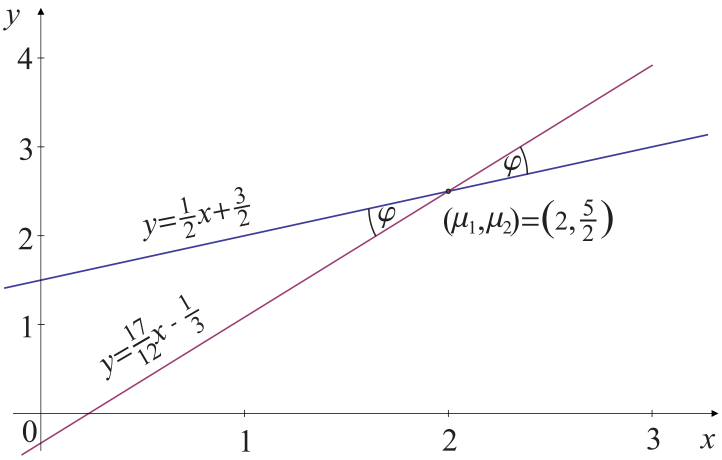
Pravac regresije X po Y je $x - \mu_1 = \frac{\mu_{XY}}{\sigma_2^2}(y - \mu_2)$, $x - \frac{36}{18} = \frac{\frac{1}{3}}{\frac{17}{36}}(y - \frac{45}{18})$,

$x = \frac{12}{17}y + \frac{4}{17}$, $y = \frac{17}{12}x - \frac{1}{3}$.

Pravci se sijeku u točki $(\mu_1, \mu_2) = (2, \frac{5}{2})$.

$tg\varphi = \frac{\frac{17}{12} - \frac{1}{2}}{1 + \frac{17}{12} \cdot \frac{1}{2}} = \frac{22}{41}$; Kut između pravaca regresije je $\varphi = 28.2^0$.

Koeficijent korelacije je $\rho_{XY} = \frac{\mu_{XY}}{\sigma_1 \cdot \sigma_2} = \frac{\frac{1}{3}}{\sqrt{\frac{2}{3} \cdot \frac{17}{36}}} = 0.594\,09$.



Slika 9.1: Pravci regresije slučajnog vektora (X, Y).

Prisjetimo se motivacijskog primjera:

PRIMJER 9.18 *motiv*

Promatramo slučajni pokus bacanje 2 igraće kocke i slučajnu varijablu X = suma brojeva koji su pali i slučajnu varijablu Y = broj 1 ako su pali jednaki brojevi, inače 0. (c) Jesu li X i Y nezavisne varijable? (d) Izračunajte kovarijancu. (e) Jesu li X i Y korelirane varijable?

Rješenje: (c) $\begin{pmatrix} X/Y & 0 & 1 & f_1(x) \\ 2 & 0 & \frac{1}{36} & \frac{1}{36} \\ 3 & \frac{2}{36} & 0 & \frac{2}{36} \\ 4 & \frac{2}{36} & \frac{1}{36} & \frac{3}{36} \\ 5 & \frac{4}{36} & 0 & \frac{4}{36} \\ 6 & \frac{4}{36} & \frac{1}{36} & \frac{5}{36} \\ 7 & \frac{6}{36} & 0 & \frac{6}{36} \\ 8 & \frac{4}{36} & \frac{1}{36} & \frac{5}{36} \\ 9 & \frac{4}{36} & 0 & \frac{4}{36} \\ 10 & \frac{2}{36} & \frac{1}{36} & \frac{3}{36} \\ 11 & \frac{2}{36} & 0 & \frac{2}{36} \\ 12 & 0 & \frac{1}{36} & \frac{1}{36} \\ f_2(y) & \frac{30}{36} & \frac{6}{36} & 1 \end{pmatrix}$

Varijable nisu nezavisne jer je npr. $f(6,1) \neq f_1(6) \cdot f_2(1)$,
$f(6,1) = \frac{1}{36}$, $f_1(6) = \frac{5}{36}$, $f_2(1) = \frac{6}{36}$.

(d) $E(X) = 7$, $E(Y) = \frac{1}{6}, E(XY) = \sum_i x_i \sum_j y_j f(x_i, y_j) = \frac{7}{6}$
Kovarijanca je $\mu_{XY} = E(XY) - E(X) \cdot E(Y) = 0$, pa je i koeficijnt korelacije jednak $\rho_{XY} = 0$.
(e) Varijable su nekorelirane jer je koeficijnt korelacije jednak $\rho_{XY} = 0$.

9.4 Ponovimo

DISKRETNI 2-dim SLUČAJNI VEKTOR

diskretni 2-dim sl. vek.	$(X,Y) : \Omega \to \mathbb{R}^2$
funkcija vjerojatnosti	$f(x_i,y_j) = P(X = x_i, Y = y_j)$, $(x_i,y_j) \in \mathcal{R}((X,Y))$
funkcija distribucije vj.	$F(x,y) = P(X \leq x, Y \leq y)$
marginalne funkcije vj.	
po X	$f_1(x) = \sum_j f(x,y_j)$
po Y	$f_2(y) = \sum_i f(x_i,y)$
nezavisne sl. var.	$f(x_i,y_j) = f_1(x_i) \cdot f_2(y_j)$, za sve $(x_i,y_j) \in \mathcal{R}((X,Y))$
funkcija 2-dim sl. vek.	$Z = h((X,Y))$, $h : \mathbb{R}^2 \to \mathbb{R}$
očekivanje od $Z = h((X,Y))$	$E(Z) = \sum_i \sum_j h(x_i,y_j) f(x_i,y_j)$
	$E(X \cdot Y) = \sum_i \sum_j x_i \cdot y_j \cdot f(x_i,y_j)$
	$E(X+Y) = E(X) + E(Y)$
ako su X i Y nezavisna	$E(X \cdot Y) = E(X) \cdot E(Y)$

KOVARIJANCA - KORELACIJA PRAVCI REGRESIJE

kovarijanca od (X,Y)	$\mu_{XY} = E(XY) - E(X) \cdot E(Y)$
	$Var(X+Y) = Var(X) + Var(Y) + 2\mu_{XY}$
koef. korelacije od (X,Y)	$\rho_{XY} = \frac{\mu_{XY}}{\sigma_1 \cdot \sigma_2}$
nekorelirane sl. var.	$\rho_{XY} = 0$
pravac regresije Y po X	$y - \mu_2 = \frac{\mu_{XY}}{\sigma_1^2}(x - \mu_1)$

Poglavlje 10

ZAKONI VELIKIH BROJEVA CENTRALNI GRANIČNI TEOREM *tko želi znati više*

Najvažniji teoremi koji će se koristiti u matematičkoj statistici označeni su s *važno*. Preporučujemo i motivirajuće primjere za Čebiševljevu nejednakost, za zakon velikih brojeva i centalni granični teorem.

Čebiševljeva nejednakost, zakoni velikih brojeva i cenralni granični teoremi su važan alati koji otkrivaju svojstva diskretnih ili kontinuiranih slučajnih varijabli koje imaju konačno očekivanje i varijancu ako nam i nije poznata njihova distribucija.

Čebiševljeva nejednakost daje ocjenu vjerojatnosti da se vrijednosti slučajne varijable razlikuje od očekivanja više od zadonog ε.

Zakoni velikih brojeva su skup teorema koji se odnosi na granične vrijednosti niza slučajnih varijabli.

Neka su $X_1, X_2, ..., X_n$ nezavisna mjerenja slučajne varijable X u ponovljenim pokusima. Slučajne varijable $X_1, X_2, ..X_n$ su nezavisne i sve imaju istu distribuciju kao i slučajna varijabla X.
Može se uočiti da njihova aritmetička sredina $\overline{X} = \frac{1}{n}(X_1 + X_2 + .. + X_n)$ ima svojstvo stabilnosti distribucije i da je vjerojatnost da se vrijednosti od $\overline{X}$ razlikuje od očekivanja više od zadonog ε jednaka nuli kad $n \to \infty$.

Centralni granični teoremi se odnose na granične zakone distribucije niza slučajnih varijabli. Suma velikog broja slučajnih varijabli ima standardnu normalnu distribuciju. Teoremi daju različite uvjete na pribrojnike u toj sumi.

10.1 ČEBIŠEVLJEVA NEJEDNAKOST

MOTIV 10.1 *Neka slučajna varijabla ima varijancu (disperziju)* $Var(X) = 0.001$. *Kolika je vjerojatnost da slučajna varijabla odstupa od očekivanja manje od* $\varepsilon = 0.1$?

TEOREM 10.1 *(MARKOVLJEVA NEJEDNAKOST)*

Neka je X slučajna varijabla s nenegativnim vrijednostima i konačnim očekivanjem $E(X)$.
Tada $\forall a > 0$ *vrijedi*

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Dokaz:
Možemo pokazati za kontinuiranu slučajnu varijablu s funkcijom gustoće $f(x)$ (analogno za diskretnu).
Kako X poprima samo nenegativne vrijednosti uočimo one $x \in R(X)$,
$0 \leq x < a$ i $x \geq a$.
Prema definiciji očekivanja: $E(X) = \int_0^\infty xf(x)dx = \int_0^a xf(x)dx + \int_a^\infty xf(x)dx$.
Budući su integrali pozitivni vrijedi nejednakost

$$E(X) \geq \int_a^\infty xf(x)dx.$$

Ako u podintegralnoj funkciji zamijenimo x s konstantom koja je uvijek manja od x, $a \leq x$, zadržat će se znak nejednakosti

$$E(X) \geq a \int_a^\infty f(x)dx.$$

Prema definiciji funkcije gustoće vjerojatnosti f(x) dobivamo traženu nejednakost:

$$E(X) \geq a \cdot P(X \geq a).$$

TEOREM 10.2 *(POOPĆENJE MARKOVLJEVE NEJEDNAKOSTI)*

Neka je X slučajna varijabla i $h : R \to R$ *nenegativna funkcija tako da postoji očekivanjem* $E(h(X))$.
Tada $\forall a > 0$ *vrijedi*

$$P(h(X) \geq a) \leq \frac{E(h(X))}{a}$$

Dokaz:
Kako h(X) poprima samo nenegativne vrijednosti uočimo one $x \in R(X)$,
$0 \leq h(x) < a$, za x$\in D_1 \subseteq R$, i $h(x) \geq a$, za $x \in D_2 \subseteq R$:
Prema definiciji očekivanja funkcije slučajne varijable X:

$$E(h(X)) = \int_0^\infty h(x)f(x)dx = \int_{D_1} h(x)f(x)dx + \int_{D_2} h(x)f(x)dx.$$

Budući su integrali pozitivni vrijedi nejednakost

$$\begin{aligned} E(h(X)) &\geq \int_{D_2} h(x)f(x)dx \geq a \int_{D_2} f(x)dx \\ &\geq a \cdot P(h(X) \geq a). \end{aligned}$$

TEOREM 10.3 *važno*
(ČEBIŠEVLJEVA NEJEDNAKOST, engl. Chebyshev's inequality)

Neka je X slučajna varijabla s konačnom varijancom $Var(X)$. Tada $\forall \varepsilon > 0$ vrijedi

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2},$$

$$P(|X - E(X)| < \varepsilon) \geq 1 - \frac{Var(X)}{\varepsilon^2}.$$

Ako označimo $Var(X) = \sigma^2, E(X) = \mu, \varepsilon = \lambda\sigma$, onda vrijedi

$$P(|X - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2},$$

$$P(|X - \mu| < \lambda\sigma) \geq 1 - \frac{1}{\lambda^2}.$$

Dokaz:
(a) Dokaz pomoću generalizirane Markovljeve nejednakosti.
Pretpostavka teorema je da X ima varijancu tj. ima očekivanje $E((X - E(X)^2) = Var(X)$ pa možemo primijeniti teorem za nenegativnu funkciju $h(x) = (x - E(X))^2$.

Vrijedi nejednakost $P((X - E(X))^2 \geq a) \leq \dfrac{E((X - E(X))^2)}{a}, \ \forall a > 0.$

Budući je $P((x - E(X))^2 \geq a) = P(|X - E(X)| \geq \sqrt{a})$, vrijedi nejednakost

$$P(|X - E(X)| \geq \sqrt{a}) \leq \frac{E((X - E(X))^2)}{a}, \ \forall a > 0.$$

Tada $\forall\varepsilon > 0$ vrijedi

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}.$$

(b) Dokaz za diskretnu slučajnu varijablu (bez generalizirane Markovljeve nejednakosti)

Neka je X diskretna slučajna varijabla sa slikom $R(X) = \{x_1, x_2, ...\}$ i neka je $f(x)$ njena funkcija vjerojatnosti.

Pretpostavka teorema je da X ima konačnu varijancu (i očekivanje).

$Var(X) = \sum_{x_i}(x_i - (E(X))^2 f(x_i)$.

Uočimo one $x_i \in R(X)$ za koje je $|x_i - (E(X)| < \varepsilon$ i one za koje je

$|x_i - (E(X)| \geq \varepsilon$.

$$\begin{aligned} Var(X) &\geq \sum_{x_i:\,|x_i-(E(X)|<\varepsilon} (x_i - (E(X))^2 f(x_i) + \sum_{x_i:\,|x_i-(E(X)|\geq\varepsilon} (x_i - (E(X))^2 f(x_i) \\ &\geq \sum_{x_i:\,|x_i-(E(X)|\geq\varepsilon} (x_i - (E(X))^2 f(x_i). \end{aligned}$$

Zamjenom svakog člana sume s manjom konstantom ε, nejednakost se zadržava pa vrijedi:

$$Var(X) \geq \sum_{x_i:\,|x_i-(E(X)|\geq\varepsilon} \varepsilon^2 f(x_i) \geq \varepsilon^2 \cdot \sum_{x_i:\,|x_i-(E(X)|\geq\varepsilon} f(x_i).$$

Suma $\displaystyle\sum_{x_i:\,|x_i-(E(X)|\geq\varepsilon} f(x_i)$ je prema definiciji funkcije distribuije jednaka $P(|X - (E(X)| \geq \varepsilon)$ i dobivamo konačnu nejednakost:

$$Var(X) \geq \varepsilon^2 \cdot P(|x_i - (E(X)| \geq \varepsilon).$$

Prisjetimo se motivacijskog primjera:

PRIMJER 10.1 *motiv*

Neka slučajna varijabla ima varijancu (disperziju)

$Var(X) = 0.001$. *Kolika je vjerojatnost da slučajna varijabla odstupa od očekivanja manje od* $\varepsilon = 0.1$*?*

Rješenje:

Trebamo izračunati $P(|X - E(X)| < \varepsilon)$.

Koristimo Čebiševljevu nejednakost u obliku $P(|X - E(X)| < \varepsilon) \geq 1 - \frac{Var(X)}{\varepsilon^2}$,

$$P(|X - E(X)| < 0.1) \geq 1 - \frac{0.001}{0.1^2} \geq 0.9.$$

PRIMJER 10.2 *Slučajna varijabla ima očekivanje* $\mu = 3$ *i standardnu devijaciju* $\sigma = 0.1$. *Ocijenite* $P(2.5 < X < 3.5)$.

Rješenje:
Zadano je očekivanje $\mu = 3$ i uočimo da je
$P(2.5 < X < 3.5) = P(\mu - 0.5 < X < \mu + 0.5) = P(|X - \mu| < 0.5)$.
Koristimo Čebiševljevu nejednakost u obliku: $P(|X - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}$,
$P(|X - \mu| < 0.5) \geq 1 - \frac{0.1^2}{0.5^2} = 0.96$,

$$P(2.5 < X < 3.5) \geq 0.96.$$

10.2 ZAKON VELIKIH BROJEVA

MOTIV 10.2 *Kontrolor uzima uzorak veličine $m = 1000$ iz skupa uređaja. Vjerojatnost da je uređaj neispravan je $p = 0.03$. U kojim granicama će biti broj neispravnih uređaja u uzorku s vjerojatnošću $\gamma = 0.99$.*

Definicija 10.1 *(KONVERGENCIJA PO VJEROJATNOSTI)*

Neka je (X_n), $n \in N$ niz slučajnih varijabli. Ako postoji slučajna varijabla X takva da

$$\forall \varepsilon > 0 \ \ \lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1$$

kažemo da niz (X_n) slučajnih varijabli konvergira slučajnoj varijabli X po vjerojatnosti i označavamo

$$(P) \lim_{n\to\infty} X_n = X.$$

Definicija 10.2 *Za niz slučajnih varijabli $(X_n), n \in N$ kažemo da zadovoljava zakon velikih brojeva ako postoji konstanta C takva da vrijedi*

$$\forall \varepsilon > 0 \ \ \lim_{n\to\infty} P(|X_n - C| < \varepsilon) = 1$$

i označavamo

$$(P) \lim_{n\to\infty} X_n = C.$$

TEOREM 10.4 *važno*
(ZAKON VELIKIH BROJEVA
(specijalan slučaj za aritmetičku sredinu))

Neka je $\{X_n\}, n \in N$ niz slučajnih varijabli takvih da za svaki n slučajne varijable $X_1, X_2, ..X_n$ su nezavisne, imaju ograničenu varijancu
$Var(X_i) = \sigma^2 \leq M > 0$ i $E(X_i) = \mu,\ i = 1, ..., n$.
Tada za aritmetičku sredinu

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + .. + X_n)$$

vrijedi

$$\forall \varepsilon > 0 \ \ \lim_{n\to\infty} P(|\overline{X} - \mu| < \varepsilon) = 1,$$

$$(P) \lim_{n\to\infty} \overline{X} = \mu.$$

Dokaz:

$\overline{X} = \dfrac{1}{n}(X_1 + X_2 + .. + X_n),\ E(\overline{X}) = \mu,\ Var(\overline{X}) = \dfrac{\sigma^2}{n}.$

Koristimo Čebiševljevu nejednakost za slučajnu varijablu $\overline{X}$ u obliku:

$P(|\overline{X} - E(\overline{X})| < \varepsilon) \geq 1 - \dfrac{Var(\overline{X})}{\varepsilon^2},$

$$P(|\overline{X} - \mu| < \varepsilon) \geq 1 - \frac{\sigma}{n \cdot \varepsilon^2},$$

$$\lim_{n\to\infty} P(|\frac{1}{n}(X_1 + X_2 + .. + X_n) - \mu| < \varepsilon) = 1.$$

TEOREM 10.5 *(BERNOULLIJEV SLABI ZAKON VELIKIH BROJEVA - za rel.frekv binomne sl. varijable)*

Neka je u Bernoullijevoj shemi slučajna varijabla X=broj uspjeha događaja A u m nezavisnih ponavljanja, $P(A) = p$. $X \sim B(m,p)$. Slučajna varijabla $Y = \frac{X}{m}$ zove se relativna frekvencija uspjeha događaja A u Bernoullijevoj shemi.
Tada vrijedi

$$\lim_{m\to\infty} \sum_{|\frac{k}{m}-p|<\varepsilon} \binom{m}{k} p^k (1-p)^{m-k} = \lim_{m\to\infty} P(|\frac{X}{m} - p| < \varepsilon) = 1,$$

$$(P) \lim_{m\to\infty} \frac{X}{m} = p.$$

Dokaz:

Za binomnu slučajnu varijablu $X \sim B(m,p)$, $E(X) = mp$,
$Var(X) = mp(1-p)$.
Za relativnu frekvenciju uspjeha $Y = \frac{X}{m}$ odredimo očekivanje i varijancu:
$E(\frac{X}{m}) = p,\ Var(\frac{X}{m}) = \frac{p(1-p)}{m}.$

Primijenimo Čebiševljevu nejednakost za slučajnu varijablu $Y = \frac{X}{m}$ nejednakost u obliku: $P(|Y - E(Y)| < \varepsilon) \geq 1 - \frac{Var(Y)}{\varepsilon^2}$,

$$P(|\frac{X}{m} - p| < \varepsilon) \geq 1 - \frac{p(1-p)}{m \cdot \varepsilon^2},$$

$$\lim_{m\to\infty} P(|\frac{X}{m} - p| < \varepsilon) = 1.$$

Kako je funkcija vjerojatnosti binomne slučajne varijable
$f(x) = \binom{m}{x} p^x (1-p)^{m-x}$

$$\lim_{m\to\infty} \sum_{|\frac{k}{m}-p|<\varepsilon} \binom{m}{k} p^k (1-p)^{m-k} = \lim_{m\to\infty} P(|\frac{X}{m} - p| < \varepsilon) = 1.$$

NAPOMENA 10.1 *važno*

Oblik Bernoullijevog slabog zakona velikih brojeva

$$P(|\frac{X}{m} - p| < \varepsilon) \geq 1 - \frac{p(1-p)}{m \cdot \varepsilon^2}$$

često se koristi u zadacima za određivanje
(a) minimalnog broja pokusa m
(b) odstupanja ε
da bi za zadani γ, $P(|\frac{X}{m} - p| < \varepsilon) \geq \gamma$.

Rješenje:

(a) $m \geq \dfrac{1}{1-\gamma} \cdot \dfrac{1}{\varepsilon^2} \cdot p(1-p)$.

(b) $\varepsilon^2 \geq \dfrac{1}{1-\gamma} \cdot \dfrac{1}{m} \cdot p(1-p)$.

Ako je p *nepoznato onda se procjenjuje da je* $p(1-p) \leq \frac{1}{4}$, *i*

(a) $m \geq \dfrac{1}{1-\gamma} \cdot \dfrac{1}{4 \cdot \varepsilon^2}$,

(b) $\varepsilon^2 \geq \dfrac{1}{1-\gamma} \cdot \dfrac{1}{4 \cdot m}$.

PRIMJER 10.3 *U Bernoullijevoj shemi vjerojatnost događaja* A *je* $p = P(A) = 1/3$. *Odredite minimalan broj ponavljanja tako da s vjerojatnošću ne manjom od* $\gamma = 0.99$ *apsolutno ostupanje relativne frekvencije od* p *bude najviše* $\varepsilon = 0.01$.

Rješenje:

Trebamo odrediti m tako da $P(|\frac{X}{m} - p| < \varepsilon) \geq 0.99$.
Koristimo Bernoullijev slabi zakon velikih brojeva: $P(|\frac{X}{m} - p| < \varepsilon) \geq 1 - \frac{p(1-p)}{m \cdot \varepsilon^2}$

$$P(|\frac{X}{m} - \frac{1}{3}| < 0.01) \geq 1 - \frac{\frac{2}{9}}{m \cdot 0.01^2}$$

Broj ponavljanja ćemo odrediti iz zadane vjerojatnosti γ i uvjeta
$1 - \frac{\frac{2}{9}}{m\cdot(0.01)^2} \geq 0.99$

$$m \geq \frac{1}{1-\gamma} \cdot \frac{1}{\varepsilon^2} \cdot p(1-p) = \frac{2}{9} \cdot \frac{1}{1-0.99} \cdot \frac{1}{(0.01)^2} \quad \Rightarrow \quad m \geq 222222.$$

Napomena (zadatak ćemo riješiti i koristeći Moivre-Laplaceov teorem-poslije).

Prisjetimo se motivacijskog primjera:

PRIMJER 10.4 *motiv*

Kontrolor uzima uzorak veličine $m = 1000$ iz skupa uređaja. Vjerojatnost da je uređaj neispravan je $p = 0.03$. U kojim granicama će biti broj neispravnih uređaja u uzorku s vjerojatnošću $\gamma = 0.99$.

Rješenje:

Treba oderditi a i b takve da za X = broj neispravnih uređaja u uzorku veličine m, $X \sim B(m, p)$, $P(a < X < b) \geq 0.99$.
Koristimo Bernoullijev slabi zakon velikih brojeva:
$P(|\frac{X}{m} - p| < \varepsilon) \geq 1 - \frac{p(1-p)}{m\cdot\varepsilon^2}$,

$$P(|\frac{X}{1000} - 0.03| < \varepsilon) \geq 1 - \frac{0.03(1-0.03)}{1000 \cdot \varepsilon^2}.$$

Odstupanje ε ćemo odrediti iz zadane vjerojatnosti γ i uvjeta
$1 - \frac{0.03(1-0.03)}{1000\cdot\varepsilon^2} \geq 0.99$.

$$\varepsilon^2 \geq \frac{1}{1-\gamma} \cdot \frac{1}{m} \cdot p(1-p) = \frac{1}{1-0.99} \cdot \frac{1}{1000} \cdot 0.03 \cdot (1-0.03) = 0.003.$$

$$\varepsilon \geq \sqrt{0.003} = 0.054 \quad \Rightarrow \quad |\frac{X}{1000} - 0.03| < 0.054 \quad \Rightarrow \quad 0 < X < 80.$$

Napomena (zadatak ćemo riješiti i koristeći Moivre-Laplaceov teorem-poslije).

10.3 CENTRALNI GRANIČNI TEOREM

Motivirajući primjer je isti kao i poglavlju Zakon velikih brojeva, ali će se sad riješiti primjenom centralnog graničnog teorema (Moivre - Laplaceov teorem).

MOTIV 10.3 *Kontrolor uzima uzorak veličine $m = 1000$ iz skupa uređaja. Vjerojatnost da je uređaj neispravan je $p = 0.03$. U kojim granicama će biti broj neispravnih uređaja u uzorku s vjerojatnošću $\gamma = 0.99$*

TEOREM 10.6 *(CENTRALNI GRANIČNI TEOREM-CGT; specijalni slučaj)*
Neka je $\{X_n\}, n \in N$ niz slučajnih varijabli takvih da za svaki n slučajne varijable $X_1, X_2, ..., X_n$ su nezavisne, imaju ograničenu varijancu
$Var(X_i) = \sigma^2 \leq M > 0$ i $E(X_i) = \mu,\ i = 1, ..., n$.
Tada

$$\lim_{n\to\infty} P(a < \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\cdot\sigma} < b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx,$$

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\cdot\sigma} \sim N(0,1), \quad n \to \infty.$$

Za velike n vrijedi

$$P(a < \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\cdot\sigma} < b) \approx F^*(b) - F^*(a).$$

($F^*(x)$ funkcija distribucije standardne normale distribucije).

Dokaz: (literatura)

TEOREM 10.7 *važno*
(CGT za aritmetičku sredinu)
Neka je $\{X_n\}, n \in N$ niz slučajnih varijabli takvih da za svaki n slučajne varijable $X_1, X_2, ..X_n$ su nezavisne, imaju ograničenu varijancu
$Var(X_i) = \sigma^2 \leq M > 0$ i $E(X_i) = \mu,\ i = 1, ..., n$.
Tada za aritmetičku sredinu $\overline{X} = \frac{1}{n}(X_1 + X_2 + .. + X_n)$ vrijedi

$$\lim_{n\to\infty} P(a < \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx,$$

$$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1), \qquad n \to \infty,$$

odnosno

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}), \qquad n \to \infty.$$

Za velike n vrijedi

$$P(a < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < b) \approx F^*(b) - F^*(a).$$

Dokaz:
Primijenimo CGT (specijalni slučaj) za $\overline{X} = \frac{1}{n}(X_1 + X_2 + .. + X_n)$.

PRIMJER 10.5 *Neka su $X_1, X_2, ..., X_n$ nezavisne slučajne varijable, imaju ograničenu varijancu $Var(X_i) = \sigma^2 = 2$ i $E(X_i) = \mu = 3,\ i = 1, ..., n = 3200$. Za aritmetičku sredinu $\overline{X} = \frac{1}{n}(X_1 + X_2 + .. + X_n)$ odredite*
$P(2.95 < \overline{X} < 3.075)$.

Rješenje:
Prema CGT za aritmetičku sredinu n slučajnih varijabli
$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1), \qquad n \to \infty,$

$$\lim_{n\to\infty} P(a < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx.$$

$$\begin{aligned} P(2.95 < \overline{X} < 3.075) &= P(\frac{2.95-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{3.075-\mu}{\frac{\sigma}{\sqrt{n}}}) \\ &= P(\frac{2.95-3}{\frac{\sqrt{2}}{\sqrt{3200}}} < \frac{\overline{X}-3}{\frac{\sqrt{2}}{\sqrt{3200}}} < \frac{3.075-3}{\frac{\sqrt{2}}{\sqrt{3200}}}) \\ &= P(-2 < \frac{\overline{X}-3}{\frac{\sqrt{2}}{\sqrt{3200}}} < 3) = F^*(3) - F^*(-2) = 0.975. \end{aligned}$$

TEOREM 10.8 *(integralni MOIVRE - LAPLACEOV TEOREM, CGT za binomnu sl. varijablu)*

Neka je u Bernoullijevoj shemi slučajna varijabla X=broj uspjeha događaja A u m nezavisnih ponavljanja $P(A) = p$. $X \sim B(m, p)$.
Tada vrijedi

$$\lim_{m\to\infty} P(a < \frac{X - m\cdot p}{\sqrt{mp(1-p)}} < b) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}x^2}dx,$$

$$\frac{X - m\cdot p}{\sqrt{mp(1-p)}} \sim N(0,1), \qquad m\to\infty,$$

odnosno

$$X \sim N(mp, mp(1-p)), \qquad m\to\infty.$$

Za velike n vrijedi

$$P(a < \frac{X - m\cdot p}{\sqrt{mp(1-p)}} < b) \approx F^*(b) - F^*(a),$$

odnosno

$$P(a < X < b) \approx F^*(\frac{b-mp}{\sqrt{mp(1-p)}}) - F^*(\frac{a-mp}{\sqrt{mp(1-p)}}).$$

Dokaz:
Promatrajmo slučajne varijable $X_i \sim B(1,p),\ i = 1,...,m$, kad $m\to\infty$.
$Var(X_i) = p(1-p)$ i $E(X_i) = p,\ i = 1,...,m$.
Integralni Moivre-Laplaceov teorem je specijalan slučaj CGT za niz slučajnih varijabli $X_i = 1,...,m$, kad $m\to\infty$

$$\lim_{m\to\infty} P(a < \frac{\sum_{i=1}^{m} X_i - m\mu}{\sqrt{m}\cdot\sigma} < b) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}x^2}dx,$$

$$\lim_{m\to\infty} P(a < \frac{\sum_{i=1}^{m} X_i - mp}{\sqrt{m\cdot p(1-p)}} < b) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}x^2}dx.$$

Slučajna varijabla X=broj uspjeha događaja A u m nezavisnih ponavljanja $P(A) = p$. $X \sim B(m,p)$. Tada je $X = \sum_{i=1}^{m} X_i$ pa vrijedi

$$\lim_{m\to\infty} P(a < \frac{X - mp}{\sqrt{m\cdot p(1-p)}} < b) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}x^2}dx.$$

PRIMJER 10.6 *Neka je X binomna slučajna varijabla* $X \sim B(m,p)$, $m = 3200,\ p = \frac{1}{2}$. *Izračunajte vjerojatnost da slučajna varijabla poprimi vrijednosti u intervalu* $(1550, 1650)$.

Rješenje:
Trebamo izračunati $P(1550 < X < 1650)$.
Prema integralnom Moivre-Laplaceovom teoremu $\frac{X-m\cdot p}{\sqrt{mp(1-p)}} \sim N(0,1)$,
$m \to \infty$,
i vrijedi aproksimacija: $P(a < X < b) \approx F^*(\frac{b-mp}{\sqrt{mp(1-p)}}) - F^*(\frac{a-mp}{\sqrt{mp(1-p)}})$.

$$\begin{aligned}
P(1550 < X < 1650) &\approx F^*(\frac{1650 - 3200 \cdot 0.5}{\sqrt{3200 \cdot 0.25}}) - F^*(\frac{1550 - 3200 \cdot 0.5}{\sqrt{3200 \cdot 0.25}}) \\
&= F^*(1.767) - F^*(-1.767) = 2F^*(1.767) - 1 \\
&= 2 \cdot 0.961 - 1 = 0.922.
\end{aligned}$$

PRIMJER 10.7 *Vjerojatnost da novorođenče bude muško ili žensko je 1/2. Kolika je vjerojatnost da među 1000 novorođenčadi bude barem 490 muških?*

Rješenje: $X \sim B(m,p)$, $m = 1000$, $p = \frac{1}{2}$
Trebamo izračunati $P(X \geq 490) = 1 - P(X \leq 490)$.
Prema integralnom Moivre-Laplaceovom teoremu $\frac{X-m\cdot p}{\sqrt{mp(1-p)}} \sim N(0,1)$,
$m \to \infty$,
i vrijedi aproksimacija: $P(a < X < b) \approx F^*(\frac{b-mp}{\sqrt{mp(1-p)}}) - F^*(\frac{a-mp}{\sqrt{mp(1-p)}})$.

$$\begin{aligned}
P(X \leq 490) &\approx F^*(\frac{490 - 1000 \cdot 0.5}{\sqrt{1000 \cdot 0.25}}) = F^*(-\frac{10}{\sqrt{250}}) = 1 - F^*(\frac{10}{\sqrt{250}}) \\
&= 1 - F^*(0.63) = 1 - 0.7357 = 0.2643.
\end{aligned}$$

$$P(X \geq 490) = 1 - P(X \leq 490) \approx 1 - 0.2643 = 0.7357.$$

TEOREM 10.9 *(integralni MOIVRE - LAPLACEOV TEOREM za rel. frekv. binomne sl. varijable) CGT za binomnu =Bernoullijev slabi ZVB za rel. frekvencije binomne*

Neka je u Bernoullijevoj shemi slučajna varijabla X=broj uspjeha događaja A u m nezavisnih ponavljanja, $P(A) = p$. $X \sim B(m,p)$. Slučajna varijabla $Y = \frac{X}{m}$ zove se relativna frekvencija uspjeha događaja A u Bernoullijevoj shemi.
Tada vrijedi

$$\lim_{m\to\infty} P(|\frac{X}{m} - p| < \varepsilon) = \lim_{m\to\infty} 2F^*(\varepsilon\sqrt{\frac{m}{p(1-p)}}) - 1 = 1.$$

Za velike m vrijedi

$$P(|\frac{X}{m} - p| < \varepsilon) \approx 2F^*(\varepsilon\sqrt{\frac{m}{p(1-p)}}) - 1.$$

Dokaz: $P(|\frac{X}{m}-p|<\varepsilon)=P(-\varepsilon\sqrt{\frac{m}{p(1-p)}}<\frac{X-mp}{\sqrt{mp(1-p)}}<\varepsilon\sqrt{\frac{m}{p(1-p)}})$

Koristimo Integralni Moivre-Laplaceov teorem; za $X\sim B(m,p)$ je
$P(a<\frac{X-m\cdot p}{\sqrt{mp(1-p)}}<b)\approx F^*(b)-F^*(a)$,

$$P(-\varepsilon\sqrt{\frac{m}{p(1-p)}}<\frac{X-mp}{\sqrt{mp(1-p)}}<\varepsilon\sqrt{\frac{m}{p(1-p)}})\approx F^*(b)-F^*(a),$$

gdje je $b=-a=\varepsilon\sqrt{\frac{m}{p(1-p)}}$.
Koristimo svojstvo $F^*(x)=1-F^*(-x)$

$$P(-\varepsilon\sqrt{\frac{m}{p(1-p)}}<\frac{X-mp}{\sqrt{mp(1-p)}}<\varepsilon\sqrt{\frac{m}{p(1-p)}})\approx 2F^*(b)-1,$$

i dobivamo željenu tvrdnju

$$P(|\frac{X}{m}-p|<\varepsilon)\approx 2F^*(\varepsilon\sqrt{\frac{m}{p(1-p)}})-1.$$

PRIMJER 10.8 *Kolika je vjerojatnost da se prilikom bacanja simetričnog novčića* $m=3600$ *puta relativna frekvencija pojavljivanja pisma po apsolutnoj vrijednosti razlikuje od* $p=1/2$ *za* $\varepsilon=0.01$?

Rješenje:
X=broj pojavljivanja pisma u Bernoullijevoj shemi bacanja novčića
$X\sim B(m,p)$, $m=3600$, $p=\frac{1}{2}$.
Trebamo izračunati $P(|\frac{X}{3600}-\frac{1}{2}|<0.01)$.
Primijenit ćemo integralni Moivre-Laplaceov teorem za rel. frekv. binomne sl. varijable u obliku $P(|\frac{X}{m}-p|<\varepsilon)\approx 2F^*(\varepsilon\sqrt{\frac{m}{p(1-p)}})-1$.

$$\begin{aligned} P(|\frac{X}{3600}-\frac{1}{2}|<0.01) &\approx 2F^*(\varepsilon\sqrt{\frac{m}{p(1-p)}})-1=2F^*(0.01\sqrt{\frac{3600}{\frac{1}{2}\cdot\frac{1}{2}}})-1\\ &= 2F^*(1.2)-1=2\cdot 0.8849-1=0.7698. \end{aligned}$$

NAPOMENA 10.2 *važno*

Oblik integralnog Moivre-Laplaceovog teorema
$P(|\frac{X}{m}-p|<\varepsilon)\approx 2F^*(\varepsilon\sqrt{\frac{m}{p(1-p)}})-1$ *često se koristi u zadacima za određivanje*
(a) minimalnog broja pokusa m i

(b) odstupanja ε
da bi za zadani γ, $P(|\frac{X}{m}-p|<\varepsilon)\geq\gamma$.

Rješenje:
(a) $m\geq\frac{(z_{\frac{1+\gamma}{2}})^2}{\varepsilon^2}\cdot p(1-p)$, $F^*(z_{\frac{1+\gamma}{2}})=\frac{1+\gamma}{2}$,
(b) $\varepsilon^2\geq\frac{(z_{\frac{1+\gamma}{2}})^2}{m}\cdot p(1-p)$.
Ako je p nepoznato onda se procjenjuje da je $p(1-p)\leq\frac{1}{4}$:
(a) $m\geq\frac{(z_{\frac{1+\gamma}{2}})^2}{4\varepsilon^2}$, $\quad F^*(z_{\frac{1+\gamma}{2}})=\frac{1+\gamma}{2}$,
(b) $\varepsilon^2\geq\frac{(z_{\frac{1+\gamma}{2}})^2}{4m}$.

PRIMJER 10.9 *U Bernoullijevoj shemi vjerojatnost događaja A je $p=P(A)=1/3$. Odredite minimalan broj ponavljanja tako da s vjerojatnošću ne manjom od $\gamma=0.99$ apsolutno ostupanje relativne frekvencije od p bude najviše $\varepsilon=0.01$.*

Rješenje:
Trebamo odrediti m tako da $P(|\frac{X}{m}-p|<\varepsilon)\geq 0.99$.
Koristimo Integralni Moivre-Laplaceov teorem za relativnu frekvenciju $\frac{X}{m}$ u Bernoullijevoj shemi.
$P(|\frac{X}{m}-p|<\varepsilon)\approx 2F^*(\varepsilon\sqrt{\frac{m}{p(1-p)}})-1$,

$$P(|\frac{X}{m}-\frac{1}{3}|<0.01)\approx 2F^*(0.01\sqrt{\frac{m}{\frac{2}{9}}})-1.$$

Broj ponavljanja ćemo odrediti iz zadane vjerojatnosti γ i uvjeta
$2F^*(0.01\sqrt{\frac{m}{\frac{2}{9}}})-1\geq 0.99 \quad\Rightarrow\quad F^*(0.01\sqrt{\frac{m}{\frac{2}{9}}})\geq 0.995$,
$F^*(z)=0.995 \quad\Rightarrow\quad z=2.6 \quad\Rightarrow\quad 0.01\sqrt{\frac{m}{\frac{2}{9}}}\geq 2.6$,
$m\geq\frac{(z_{\frac{1+\gamma}{2}})^2}{\varepsilon^2}\cdot p(1-p)=\frac{2.6^2}{0.01^2}\cdot\frac{2}{9} \quad\Rightarrow\quad m\geq 15022.$
(Bernoullijev SZVB za rel. frekv. dao je ocjenu $m\geq 222222$.)

PRIMJER 10.10 *Koliko puta treba baciti simetričnu kocku da bi relativna frekvencija pojavljivanja broja 6 bila između $\frac{19}{120}$ i $\frac{21}{120}$ s vjerojatnošću $\gamma=0.95$.*

Rješenje:
X=broj pojavljivanja 6 u Bernoullijevoj shemi bacanja kocke
$X \sim B(m, p)$, m, $p = \frac{1}{6}$.
Trebamo odrediti m tako da $P(\frac{19}{120} < \frac{X}{m} < \frac{21}{120}) \geq 0.95$.

$$\begin{aligned} P(\frac{19}{120} < \frac{X}{m} < \frac{21}{120}) &= P(\frac{19}{120} - \frac{1}{6} < \frac{X}{m} - \frac{1}{6} < \frac{21}{120} - \frac{1}{6}) \\ &= P(-\frac{1}{120} < \frac{X}{m} - \frac{1}{6} < \frac{1}{120}) = P(|\frac{X}{m} - \frac{1}{6}| < \frac{1}{120}). \end{aligned}$$

Koristimo Integralni Moivre-Laplaceov teorem za relativnu frekvenciju $\frac{X}{m}$ u Bernoullijevoj shemi.
$P(|\frac{X}{m} - p| < \varepsilon) \approx 2F^*(\varepsilon\sqrt{\frac{m}{p(1-p)}}) - 1$,

$$P(|\frac{X}{m} - \frac{1}{6}| < \frac{1}{120}) \approx 2F^*(\frac{1}{120}\sqrt{\frac{m}{\frac{5}{36}}}) - 1$$

Broj ponavljanja ćemo odrediti iz zadane vjerojatnosti γ i uvjeta
$2F^*(\frac{1}{120}\sqrt{\frac{m}{\frac{5}{36}}}) - 1 \geq 0.95 \quad \Rightarrow \quad F^*(\frac{1}{120}\sqrt{\frac{m}{\frac{5}{36}}}) \geq 0.975$
$F^*(z) = 0.975 \quad \Rightarrow \quad z = 1.96 \quad \Rightarrow \quad \frac{1}{120}\sqrt{\frac{m}{\frac{5}{36}}} \geq 1.96$,
$m \geq \frac{(z_{\frac{1+\gamma}{2}})^2}{\varepsilon^2} \cdot p(1-p) = \frac{1.96^2}{(\frac{1}{120})^2} \cdot \frac{5}{36} = 7684 \quad \Rightarrow \quad m \geq 7684$.

PRIMJER 10.11 *Simetričnu kocku bacamo* $m = 4500$ *puta. U kojim granicama s vjerojatnošću* $\gamma = 0.9$ *treba očekivati relativne frekvencije pojavljivanja boja 6?*

Rješenje:
X =broj pojavljivanja 6 u Bernoullijevoj shemi bacanja kocke,
$X \sim B(m, p)$, $m, p = \frac{1}{6}$.
Treba oderditi a i b takve da za X =broj 6 u m bacanja $P(a < \frac{X}{m} < b) \geq 0.9$.
Koristimo Integralni Moivre-Laplaceov teorem za relativnu frekvenciju $\frac{X}{m}$ u Bernoullijevoj shemi $P(|\frac{X}{m} - p| < \varepsilon) \approx 2F^*(\varepsilon\sqrt{\frac{m}{p(1-p)}}) - 1$.

$$P(|\frac{X}{4500} - \frac{1}{6}| < \varepsilon) \approx 2F^*(\varepsilon\sqrt{\frac{4500}{\frac{1}{6}(1-\frac{1}{6})}}) - 1$$

Odstupanje ε ćemo odrediti iz zadane vjerojatnosti γ i uvjeta
$2F^*(\varepsilon\sqrt{\frac{4500}{\frac{5}{36}}}) - 1 \geq 0.9 \quad \Rightarrow \quad F^*(\varepsilon\sqrt{36 \cdot 900}) \geq 0.95$.
$F^*(z) = 0.9 \Rightarrow z = 1.65 \quad \Rightarrow \quad \varepsilon\sqrt{36 \cdot 900} \geq 1.65 \Rightarrow \varepsilon \geq 0.0091$

Prema formuli $\varepsilon^2 \geq \frac{(z_{\frac{1+\gamma}{2}})^2}{m} \cdot p(1-p) = \frac{1.65^2}{4500} \cdot \frac{5}{36} = 7.566 \times 10^{-5}$

$$\varepsilon \geq \frac{1.65}{30} \cdot \frac{1}{6} = 9.1667 \times 10^{-3} \quad \Rightarrow \quad \varepsilon \geq 9.1667 \times 10^{-3}.$$

Odredili smo ε tako da $P(|\frac{X}{4500} - \frac{1}{6}| < 0.00916) \geq 0.9$.

$$|\frac{X}{4500} - \frac{1}{6}| < 0.00916 \quad \Rightarrow \quad \frac{1}{6} - 0.00916 < \frac{X}{4500} < \frac{1}{6} + 0.00916$$

$$\Rightarrow \quad 0.15751 < \frac{X}{4500} < 0.17583.$$

Prisjetimo se motivacijskog primjera:

PRIMJER 10.12 *motiv*

Kontrolor uzima uzorak veličine $m = 1000$ iz skupa uređaja. Vjerojatnost da je uređaj neispravan je $p = 0.03$. U kojim granicama će biti broj neispravnih uređaja u uzorku s vjerojatnošću $\gamma = 0.99$

Rješenje:

Treba oderditi a i b takve da za X =broj neispravnih uređaja u uzorku veličine m, $X \sim B(m, p)$, $P(a < X < b) \geq 0.99$.

Koristimo Integralni Moivre-Laplaceov teorem za relativnu frekvenciju $\frac{X}{m}$ u Bernoullijevoj shemi $P(|\frac{X}{m} - p| < \varepsilon) \approx 2F^*(\varepsilon \sqrt{\frac{m}{p(1-p)}}) - 1$,

$$P(|\frac{X}{1000} - 0.03| < \varepsilon) \approx 2F^*(\varepsilon \sqrt{\frac{1000}{0.03(1-0.03)}}) - 1.$$

Odstupanje ε ćemo odrediti iz zadane vjerojatnosti γ i uvjeta

$2F^*(\varepsilon \sqrt{\frac{1000}{0.03(1-0.003)}}) - 1 \geq 0.99 \quad \Rightarrow \quad F^*(\varepsilon \sqrt{\frac{1000}{0.03(1-0.003)}}) \geq 0.995.$

$F^*(z) = 0.995 \;\Rightarrow\; z = 2.6 \quad \Rightarrow \quad \varepsilon \sqrt{\frac{1000}{0.03(1-0.003)}} \geq 2.6.$

Prema formuli $\varepsilon^2 \geq \frac{(z_{\frac{1+\gamma}{2}})^2}{m} \cdot p(1-p) = \frac{2.6^2}{1000} \cdot 0.03 \cdot (1 - 0.03) = 1.9672 \times 10^{-4}$

$$\varepsilon \geq 0.014$$

Odredili smo ε tako da $P(|\frac{X}{1000} - 0.03| < 0.014) \geq 0.99$.

$$|\frac{X}{1000} - 0.03| < 0.014 \quad \Rightarrow \quad 16 < X < 44.$$

(Bernoullijev SZVB za rel. frekv. dao je ocjenu $0 < X < 80$.)

10.4 Ponovimo

ČEBIŠEVLJEVA NEJEDNAKOST

za svaki $\varepsilon > 0$	$P(\lvert X - E(X)\rvert \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$,
za svaki λ	$P(\lvert X - \mu\rvert \geq \lambda\sigma) \leq \frac{1}{\lambda^2}$,

ZAKONI VELIKIH BROJEVA (ZVB)

ZVB za $\overline{X}$	$\lim_{n\to\infty} P(\lvert\overline{X} - \mu\rvert < \varepsilon) = 1$,
ZVB za $X \sim B(m, p)$	$\lim_{n\to\infty} P(\lvert\frac{X}{m} - p\rvert < \varepsilon) = 1$,
Bernoull. slabi ZVB	$P(\lvert\frac{X}{m} - p\rvert < \varepsilon) \geq 1 - \frac{p(1-p)}{m\cdot\varepsilon^2}$

CENTRALNI GRANIČNI TEOREMI (CGT)

CGT za niz nez. sl, var. X_i	$n \to \infty$,
	$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\cdot\sigma} \sim N(0,1)$
CTG za $\overline{X}$	$n \to \infty$,
	$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$
CTG za $X \sim B(m,p)$, $m \to \infty$	= Moivre-Laplace
$\frac{X-m\cdot p}{\sqrt{mp(1-p)}} \sim N(0,1)$,	$P(a < X < b) \approx F^*(\frac{b-mp}{\sqrt{mp(1-p)}}) - F^*(\frac{a-mp}{\sqrt{mp(1-p)}})$
CGT za rel. frekv. od $X \sim B(m,p)$	= Moivre-Laplace za rel. frekv.
$\frac{\frac{X}{m}-p}{\sqrt{\frac{p(1-p)}{m}}} \sim N(0,1)$,	$P(\lvert\frac{X}{m} - p\rvert < \varepsilon) \approx 2F^*(\varepsilon\sqrt{\frac{m}{p(1-p)}}) - 1$

Dio III

Matematička statistika

Poglavlje 11

MATEMATIČKA STATISTIKA

Matematička statistika je znanstvena disciplina koja provjerava matematičke modele slučajnog pokusa u realnosti. Proučava svojstva slučajnog uzoraka i donosi zaključke o populaciji iz koje je uzet slučajni uzorak. Statističke metode daju zaključke s nekom vjerojatnošću pa se temelje na teoriji vjerojatnosti.

Deskriptivna statistika bavi se uređivanjem prikupljenih, empirijskih podataka, njihovim grafičkim prikazivanjem i opisivanjem pomoću numeričkih vrijednosti: prosjek, standardna devijacija, korelacijski koeficijent,...

Induktivna statistika (Inferencijalna statistika)bavi se metodama koje se zasnivaju na teoriji vjerojatnosti i koje omogućavaju da se donose zaključci o populaciji pomoću uzoraka iz populacije.

Tri pravca u matematčkoj statistici (induktivnoj statististici) su:
teorija procjene,
teorija testiranja statističkih hipoteza,
teorija planiranja eksperimenata.

U teoriji procjene osvnut ćemo se na:
točkaste procjenitelje,
metodu max vjerojatnosti za određivanje procjenitelja,*tko želi znati više*
intervale povjerenja za procjenitelje za parametre normalne razdiobe.

U teoriji testiranja osvnut ćemo se na:
test hipoteze o parametrima normalne razdiobe,

Teorija planiranja eksperimenta razvija metodu sekvencijalne analize, broj promatranja je slučajan, pa se provjera statstičkih hipoteza ovom metodom izvodi postepeno, u etapama. Hipoteza se može prihhvatiti, odbiti ili produžiti eksperiment.

11.1 DESKRIPTIVNA STATISTIKA

Definicija 11.1 *(POPULACIJA)*

Populacija (osnovni skup, statistički skup) je skup svih elemenata od kojih bismo mogli uzeti podatke o određenim veličinama.
Populacija može biti konačna ili beskonačna.

PRIMJER 11.1 *Populacija - sve obitelji u jednoj zgradi.*
Veličine koje možemo razmatrati: broj djece, mjesečni dohodak..

Definicija 11.2 *(STATISTIČKA VARIJABLA-OBILJEŽJE)*

Statističko obilježje (vaijabla) je numeričko svojstvo elemenata statističkog skupa. Ako je skup vrijednosti R(X) statističkog obilježja diskretan onda za X kažemo da je diskretno obilježje, a ako je R(X) ⊆ R kažemo da je kontinuirano obilježje.
Uzorak je podskup populacije koji uzimamo na unaprijed određen način.

Definicija 11.3 *(FREKVENCIJA, RELATIVNA FREKVENCIJA, KUMULATIVNA RELATIVNA FREKVENCIJA, ARITMETIČKA SREDINA, VARIJANCA, STANDARDNA DEVIJACIJA)*

Neka je X statističko obilježje i neka se mjerenje ponovi n, konačno mnogo puta (nezavisno) i dobije n statističkih podataka x_i, $i = 1, .., n$. Slika R(X)=$\{x_k^*,\ k = 1, .., r\}$ sadrži r različitih statističkih podataka. Ako se x_k^* pojavi f_k puta onda kažemo da x_k^* pripada frekvencija f_k i relativna frekvencija $\frac{f_k}{n}$, za $k = 1, .., r$.
Vrijedi: $\sum\limits_{k=1}^{r} f_k = n$, $\sum\limits_{k=1}^{r} \frac{f_k}{n} = 1$.
Za $x \in R$ kažemo da ima kumulativnu relativnu frekvenciju $F_n(x)$= $\sum\limits_{k,\, x_k \leq x} \frac{f_k}{n}$.
Aritmetička redina n statističkih podataka x_i, $i = 1, .., n$:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \overline{x} = \frac{1}{n}\sum_{k=1}^{r} x_k^* f_k.$$

Varijanca n statističkih podataka x_i, $i = 1, .., n$:

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2,$$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{r}(x_k^* - \overline{x})^2 f_k = \frac{1}{n}\sum_{k=1}^{r}(x_k^*)^2 f_k - \overline{x}^2.$$

Standardna devijacija je $\widehat{\sigma}$.

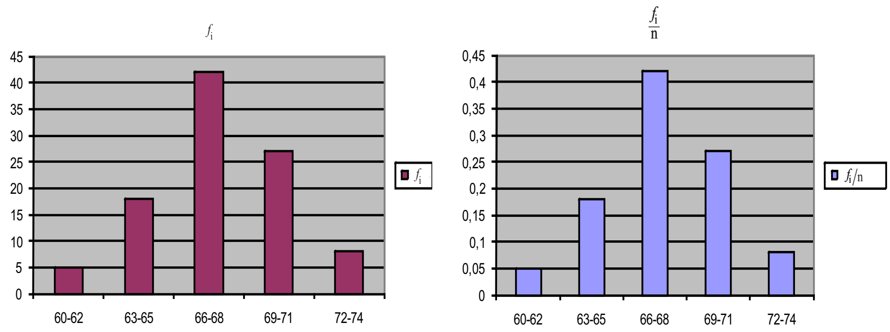
Statističke podatke koji se dobiju mjerenjem statističkog obilježja X možemo prikazati:
tablično: tablicom frekvencija i tablicom relativnih frekvencija,
grafički: grafikonom frekvencija, relativnih frekvencija, kumulativnih frekvencija, histogramom (nad dobivenim podacima x_k^* nacrtani su pravokutnici visine jednake frekvenciji $f_k$ili relativnoj frekvenciji $\frac{f_k}{n}$),
poligonom (izlomljena linija koja spaja točke (x_k^*, f_k)).
Ako je n veliki i skup vrijednosti ima veliki broj elemenata (posebno kod kontinuirane slučajne varijable-statističkog obilježja) formiramo r razreda. Prilikom tabličnog i grafičkog prikazivanja vrijednosti slučajnog uzorka na apscisu nanosimo r podintervala (razreda), sa sredinama razreda x_{ksr}^*, a na ordinatu sumu frekvencija f_k elemenata iz tog razreda.
Broj razreda r ponekad se računa po formulama: $r = \sqrt{n}$, $r = 2\sqrt[3]{n}$.
U praksi se koristi slijedeća shema za izbor broja razreda:

n	r
40-60	6-8
60-100	7-10
100-200	8-12
200-500	12-17
> 500	21

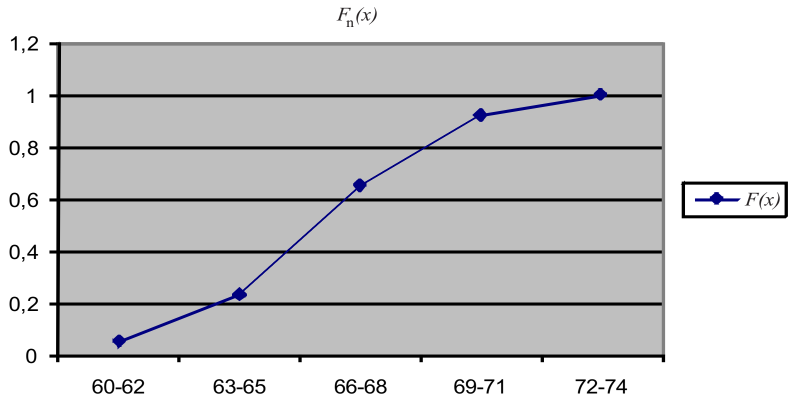
PRIMJER 11.2 *Mjerenjem kontinuirane slučajne varijable X= prosječne težine studenata jednog turnusa na uzorku veličine 100 dobivena je vrijednost slučajnog uzorka* $(x_1, x_2, ..., x_{100})$ *dana u tablici:*

razred	x_{ksr}^*	f_k	$\frac{f_k}{n}$	$F_n(x)$
60-62	*61*	*5*	*0,05*	*0,05*
63-65	*64*	*18*	*0,18*	*0,23*
66-68	*67*	*42*	*0,42*	*0,65*
69-71	*70*	*27*	*0,27*	*0,92*
72-74	*73*	*8*	*0,08*	*1,00*
ukupno		*n=100*	*1,00*	

Relativne frekvencije odgovaraju pojmu statističke vjerojatnosti.
$P(66 < X < 68) \approx 0,42$



Slika 11.1: Histogrami frekvencija i relativnih frekvencija iz primjera 11.2.



Slika 11.2: Graf kumulativnih relativnih frekvencija iz primjera 11.2.

Definicija 11.4 *(STATISTIČKA RAZDIOBA)*

Statističko obilježje (slučajna varijabla) X sa skupom vrijednosti R(X) opisano grafom relativnih frekvencija ili grafom kumulativnih relativnih frekvencija ima statističku funkciju distribucije $F_n(x)$. Slučajna varijabla X ima i teorijsku funkciju distribucije $F(x)$.

TEOREM 11.1 *(GLIVENKO)*

Ako su vrijednosti u uzorku slučajne varijable X (statističkog obilježja) nezavisni, onda je

$$P(\sup_{-\infty<x<\infty} \|F_n(x) - F(x)\| \to 0) = 1, \text{ kad } n \to \infty.$$

Kad je uzorak dovoljno velik, onda se s vjerojatnošću skoro 1 statistička razdioba malo razlikuje od teorijske razdiobe.

Definicija 11.5 *(Kvantil, medijan, prvi kvartil, treći kvartil)*

Ako je F funkcija distribucije slučajne varijable X onda se rješenje jednadžbe $F(x_p) = p$ *zove kvantil reda p.*

Medijan $Me = x_{0.5}$; $F(Me) = 0.5$ *tj.* $P(X \leq Me) = 0.5$

Prvi kvartil $Q_1 = x_{0.25}$; $F(Q_1) = 0.25$ *tj.* $P(X \leq Q_1) = 0.25$

Drugi kvartil $Q_2 = x_{0.5} = Me$

Treći kvartil $Q_3 = x_{0.75}$; $F(Q_3) = 0.75$ *tj.* $P(X \leq Q_3) = 0.75$

PRIMJER 11.3 *Računanje medijana statističkog obilježja X:*

(A) Ako je niz statističkih podataka, vrijednosti nekog statističkog obilježja X rastući $x_1 \leq x_2 \leq ... \leq x_n$, *onda je*

$$Me = \begin{cases} x_{\frac{n+1}{2}}, & \text{za } n \text{ neparan;} \\ \dfrac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}, & \text{za } n \text{ paran.} \end{cases}$$

PRIMJER 11.4 *Odredite medijan za zadani niz statističkih podataka*

3 4 4 5 6 8 8 8 10, n=9, neparan.

$Me = x_{\frac{n+1}{2}} = x_5 = 6.$

PRIMJER 11.5 *Računanje medijana statističkog obilježja X:*

(B) Ako su vrijednosti statističkog obilježja date u razredima s odgovarajućim frekvencijama f_i *onda je*

$$Me = L_{Me} + d \cdot \frac{\frac{n}{2} - (f_1 + f_2 + ... + f_k)}{f_{k+1}},$$

gdje je k izabran tako da je

$$F_k' = (f_1 + f_2 + ... + f_k) \leq \frac{n}{2} \leq f_1 + f_2 + ... + f_k + f_{k+1} = F_{k+1}',$$

L_{Me} *je lijevi rub* $k + 1$ *razreda, d je širina razreda.*

PRIMJER 11.6 *Računanje prvog kvartila statističkog obilježja X:*
(A) Ako je niz statističkih podataka, vrijednosti nekog statističkog obilježja X rastući $x_1 \leq x_2 \leq ... \leq x_n$, *onda je*

$$Q_1 = \begin{cases} x_{cijelo(\frac{n}{4}+1)}, & za\ n\ nije\ djeljiv\ s\ 4; \\ \dfrac{x_{\frac{n}{4}} + x_{\frac{n}{4}+1}}{2}, & za\ n\ djeljiv\ s\ 4. \end{cases}$$

PRIMJER 11.7 *Odredite prvi kvartil za niz statističkih podataka*
3 4 4 5 6 8 8 8 10, $n = 9$, *nije djeljiv s 4.*
$Q_1 = x_{cijelo(\frac{n}{4}+1)} = x_3 = 4.$

PRIMJER 11.8 *Računanje prvog kvartila statističkog obilježja X:*
(B) Ako su vrijednosti statističkog obilježja date u razredima s odgovarajućim frekvencijama f_i *onda je*

$$Q_1 = L_{Q_1} + d \cdot \frac{\frac{n}{4} - (f_1 + f_2 + ... + f_k)}{f_{k+1}},$$

gdje je k izabran tako da je

$$F_k' = (f_1 + f_2 + ... + f_k) \leq \frac{n}{4} \leq f_1 + f_2 + ... + f_k + f_{k+1} = F_{k+1}',$$

L_{Q_1} *je lijevi rub* $k+1$ *razreda, d je širina razreda.*

PRIMJER 11.9 *Računanje trećeg kvartila:*
(A) Ako je niz statističkih podataka, vrijednosti nekog statističkog obilježja X rastući $x_1 \leq x_2 \leq ... \leq x_n$, *onda je*

$$Q_3 = \begin{cases} x_{cijelo(\frac{3n}{4}+1)}, & za\ n\ nije\ djeljiv\ s\ 4; \\ \dfrac{x_{\frac{3n}{4}} + x_{\frac{3n}{4}+1}}{2}, & za\ n\ djeljiv\ s\ 4. \end{cases}$$

PRIMJER 11.10 *Odredite treći kvartil niz statističkih podataka*
3 4 4 5 6 8 8 8 10, $n = 9$, *nije djeljiv s 4.*
$Q_3 = x_{cijelo(\frac{3n}{4}+1)} = x_7 = 8.$

PRIMJER 11.11 *Računanje trećeg kvartila:*
(B) Ako su vrijednosti statističkog obilježja date u razredima s odgovarajućim frekvencijama f_i *onda je*

$$Q_3 = L_{Q_3} + d \cdot \frac{\frac{3n}{4} - (f_1 + f_2 + ... + f_k)}{f_{k+1}},$$

gdje je k izabran tako da je

$$F'_k = (f_1 + f_2 + ... + f_k) \leq \frac{3n}{4} \leq f_1 + f_2 + ... + f_k + f_{k+1} = F'_{k+1},$$

L_{Q_3} *je lijevi rub* $k+1$ *razreda, d je širina razreda.*

Definicija 11.6 *(MOD)*
Mod je vrijednost statističkog obilježja koja ima najveću frekvenciju. Može se dogoditi da mod ne postoji ili da postoji više modova.

PRIMJER 11.12 *Odredite mod niza statističkih podataka*
3 4 4 5 6 8 8 8 10.

x_i	f_i
3	1
4	2
5	1
6	1
8	3
10	1

$x_i = 8$ *ima maksimalnu frekvenciju* $f_i = 3$, $Mo = 8$.

PRIMJER 11.13 *Računanje moda:*
Ako su vrijednosti statističkog obilježja date u razredima s odgovarajućim frekvencijama f_i *onda je*

$$Mo = L_{Mo} + d \cdot \frac{\Delta_1}{\Delta_1 + \Delta_2},$$

gdje je k izabran tako da je f_k *maksimalan,* L_{Mo} *je lijevi rub* $k - tog$ *razreda, d je širina razreda,* $\Delta_1 = f_k - f_{k-1}$, $\Delta_2 = f_k - f_{k+1}$.

Definicija 11.7 *(koeficijent varijacije)*
Koeficijent varijacije je relativna mjera standardne devijacija i računa se na dva načina $K_V = \frac{\sigma}{\overline{x}} \cdot 100$ *ili pomoću kvartila* $K_V = \frac{Q_3 - Q_1}{Q_3 + Q_1}$.

Definicija 11.8 *(koeficijent asimetrije-eng. skewness)*
Koeficijent asimetrije za slučajnu varijablu X je broj K_A *koji karakterizira simetriju razdiobe i definira se kao kvocijent trećeg centralnog momenta i kuba standardne devijacije* σ :

$$K_A = \frac{\mu_3}{\sigma^3}.$$

Definicija 11.9 *Koeficijent asimetrije statističkog obilježja X, ako su vrijednosti statističkog obilježja date kao niz x_i^* s frekvencijama f_i, $i = 1, ..., r$, definira se kao*

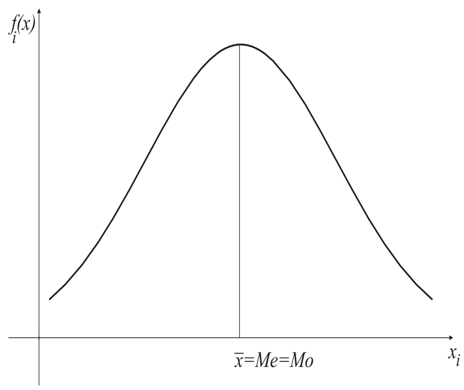
$$K_A = \frac{\widehat{\mu_3}}{\widehat{\sigma}^3},$$

gdje je

$$\widehat{\mu_3} = \frac{1}{n}\sum_{i=1}^{r}(x_{isr}^* - \overline{x})^3 f_i, \quad (\sum_{i=1}^{r} f_i = n);$$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{r}(x_{isr}^* - \overline{x})^2 f_i, \quad (\sum_{i=1}^{r} f_i = n).$$

NAPOMENA 11.1 *Ako je $K_A = 0$ onda je razdioba frekvencija simetrična u odnosu na pravac $x = \overline{x}$ onda se poklapaju $\overline{x} = Me = Mo$. (Normalna razdioba ima $K_A = 0$)*



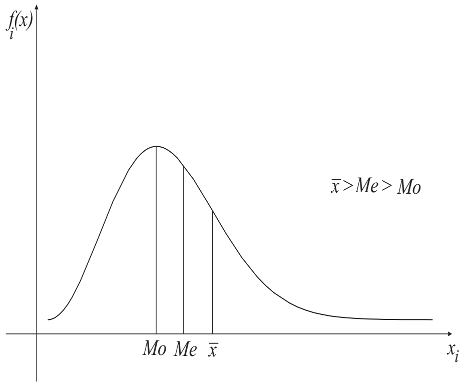
Ako je $K_A > 0$ onda je razdioba frekvencija asimetrična u odnosu na pravac $x = \overline{x}$, asimetrija je pozitivna i vrijedi $\overline{x} > Me > Mo$.
Ako je $K_A < 0$ onda je razdioba frekvencija asimetrična u odnosu na pravac $x = \overline{x}$, asimetrija je negativna i vrijedi $\overline{x} < Me < Mo$.

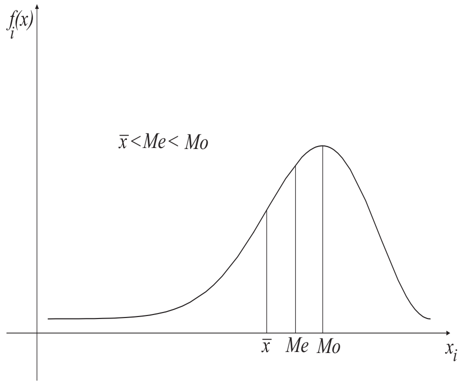
Definicija 11.10 *(koeficijent spljoštenosti (eksces)-engl. kurtosis)*
Koeficijent spljoštenosti slučajne varijable X je broj K_E koji karakterizira zaobljenost razdiobe i definira se kao pomoću kvocijenta četvrtog centralnog momenta i četvrte potencije standardne devijacije σ :

$$K_E = \frac{\mu_4}{\sigma^4} - 3.$$

Definicija 11.11 *Koeficijent spljoštenosti statističkog obilježja X, ako su vrijednosti statističkog obilježja date kao niz x_i^* s frekvencijama f_i, $i = 1, ..., r$, definira se kao*

$$K_E = \frac{\widehat{\mu_4}}{\widehat{\sigma}^4} - 3,$$



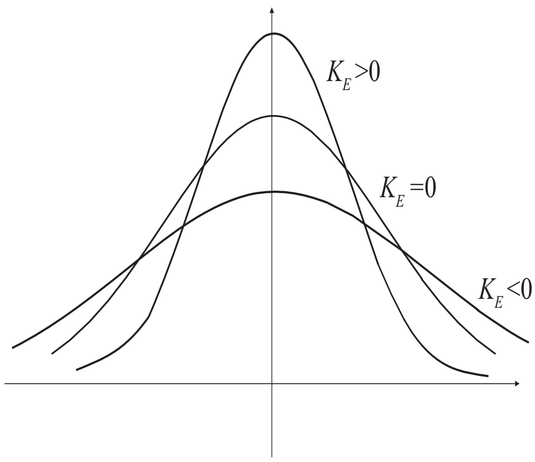


gdje je

$$\widehat{\mu}_4 = \frac{1}{n}\sum_{i=1}^{r}(x^*_{isr} - \overline{x})^4 f_i, \quad (\sum_{i=1}^{r} f_i = n);$$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{r}(x^*_{isr} - \overline{x})^2 f_i, \quad (\sum_{i=1}^{r} f_i = n).$$

NAPOMENA 11.2 *Ako je $K_E = 0$ onda je razdioba frekvencija normalna razdioba. (Normalna razdioba ima $K_E = 0$)*
Ako je $K_E > 0$ onda je graf funkcije razdiobe frekvencija uži od grafa normalne razdiobe (spljoštenost je manja).
Ako je $K_E < 0$ onda je graf funkcije razdioba frekvencija širi od normalne razdiobe (spljoštenost je veća).



Slika 11.3: Spljoštenost

PRIMJER 11.14 *Mjerenjem kontinuirane slučajne varijable X= prosječne težine studenata jednog turnusa na uzorku veličine 100 dobivena je vrijednost slučajnog uzorka* $(x_1, x_2, ..., x_{100})$ *dana u tablici:*

razred	x^*_{isr}	f_i	F'_i	$\frac{f_i}{n}$	$F_n(x)$
60-62	*61*	*5*	*5*	*0,05*	*0,05*
63-65	*64*	*18*	*23*	*0,18*	*0,23*
66-68	*67*	*42*	*65*	*0,42*	*0,65*
69-71	*70*	*27*	*92*	*0,27*	*0,92*
72-74	*73*	*8*	*100*	*0,08*	*1,00*
ukupno		*n=100*		*1,00*	

Odrediti očekivanje, varjancu, standardnu devijaciju, mod, medijan, prvi kvartil, treći kvartil, koeficijent varijacije, koeficijent asimetrije, koeficijent spljoštenosti.
NAPOMENA: Razredi su u tablici dati smbolično npr. razred $60-62$ *je razred* $59.5-62.5$ *tako da je širina razreda* $d = 3$.

Rješenje:

očekivanje $\overline{x} = \frac{1}{100}\sum_{i=1}^{5} x^*_{isr} f_i = 67.45$

varijanca $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{r}(x^*_{isr} - \overline{x})^2 f_i = 8.527$

medijan $Me = L_{Me} + d \cdot \frac{\frac{n}{2} - (f_1 + f_2 + ... + f_k)}{f_{k+1}}$, gdje je $k + 1 = 3$ izabran tako da je $F'_k = f_1 + f_2 = 23 \leq \frac{n}{2} = 50 \leq f_1 + f_2 + f_3 = 65 = F'_{k+1}$,
$L_{Me} = 65.5$ je lijevi rub $k + 1 = 3$. razreda, $d = 3$ je širina razreda.

$$Me = L_{Me} + d \cdot \frac{\frac{n}{2} - (f_1 + f_2 + ... + f_k)}{f_{k+1}} = 65.5 + 3 \cdot \frac{50 - 23}{42} = 67.4.$$

prvi kvartil $Q_1 = L_{Q_1} + d \cdot \frac{\frac{n}{4} - (f_1 + f_2 + ... + f_k)}{f_{k+1}}$, gdje je $k + 1 = 3$ izabran tako da je $F'_k = f_1 + f_2 = 23 \leq \frac{n}{4} = 25 \leq f_1 + f_2 + f_3 = 65 = F'_{k+1}$,
$L_{Q_1} = 65.5$ je lijevi rub $k + 1 = 3$. razreda, $d = 3$ je širina razreda.

$$Q_1 = L_{Q_1} + d \cdot \frac{\frac{n}{4} - (f_1 + f_2 + ... + f_k)}{f_{k+1}} = 65.5 + 3 \cdot \frac{25 - 23}{42} = 65.643$$

treći kvartil $Q_3 = L_{Q_3} + d \cdot \frac{\frac{3n}{4} - (f_1 + f_2 + ... + f_k)}{f_{k+1}}$, gdje je $k + 1 = 4$ izabran tako da

je $F_k^{'} = f_1 + f_2 + f_3 = 65 \leq \dfrac{3n}{4} = 75 \leq f_1 + f_2 + f + f_4 = 92 = F_{k+1}^{'}$,
$L_{Q_3} = 68.5$ je lijevi rub $k + 1 = 4$. razreda, $d = 3$ je širina razreda.

$$Q_3 = L_{Q_3} + d \cdot \frac{\frac{3n}{4} - (f_1 + f_2 + ... + f_k)}{f_{k+1}} = 68.5 + 3 \cdot \frac{75 - 65}{27} = 69.611.$$

mod $Mo = L_{Mo} + d \cdot \dfrac{\Delta_1}{\Delta_1 + \Delta_2}$, gdje je $k = 3$ izabran tako da je $f_k = 42$ maksimalan, $L_{Mo} = 65.5$ je lijevi rub $k = 3$. razreda, $d = 3$ je širina razreda, $\Delta_1 = f_k - f_{k-1} = f_3 - f_2 = 42 - 18$, $\Delta_2 = f_k - f_{k+1} = f_3 - f_4 = 42 - 27$.

$$Mo = L_{Mo} + d \cdot \frac{\Delta_1}{\Delta_1 + \Delta_2} = 65.5 + 3 \cdot \frac{24}{24 + 15} = 67.346.$$

koeficijent varijacije $K_V = \dfrac{\widehat{\sigma}}{\overline{x}} \cdot 100\% = \dfrac{\sqrt{8.527}}{67.45} \cdot 100\% = 4.32\%$
$K_V = \dfrac{Q_3 - Q_1}{Q_3 + Q_1} = \dfrac{69.611 - 65.643}{69.611 + 65.643} = 2.9337 \times 10^{-2}$

koeficijet asimetrije i spljoštenosti $\widehat{\mu}_3 = \dfrac{1}{n} \sum_{i=1}^{r} (x_{isr}^{*} - \overline{x})^3 f_i = -2.293$

$\widehat{\mu}_4 = \dfrac{1}{n} \sum_{i=1}^{r} (x_{isr}^{*} - \overline{x})^4 f_i = 199.37$

$$K_A = \frac{\widehat{\mu}_3}{\widehat{\sigma}^3} = -0.14, \quad K_E = \frac{\widehat{\mu}_4}{\widehat{\sigma}^4} - 3 = -0.26$$

11.2 Ponovimo

STATISTIČKO OBILJEŽJE

statističko obilježje	X
vrijednosti stat. obilježja	$x_1, x_2, ..., x_n$
r različitih vrijednosti	$x_1^*, x_2^*, ..., x_r^*$
frkvencije	$f_1^*, f_2^*, ..., f_r^*$
relativne frkvencije	$\frac{f_1^*}{n}, \frac{f_2^*}{n}, ..., \frac{f_r^*}{n}$
kumulativne rel. frkvencije	$F_n(x) = \frac{f_1^*}{n} + \frac{f_2^*}{n} + ... + \frac{f_j^*}{n}$
	j takav da je $x \leq x_j^*$
statistička razdioba za X st. obilj.	$F_n(x)$
teorijska razdioba za sl. var.X	$F(x)$
$F_n(x) \to F(x)$	po vjerojatnosti kad $n \to \infty$

MJERE POLOŽAJA

vrijednosti stat. obilj.	ARITMETIČKA SREDINA
$x_1, x_2, ..., x_n$	$\overline{x} = \frac{1}{n}(x_1 + x_2 + ... + x_n)$
$x_1^*, x_2^*, ..., x_r^*$	$\overline{x} = \frac{1}{n}\sum_{j=1}^{r} x_j^* \cdot f_j^*$
r razreda $[L_j, D_j]$, $j = 1, ...r$; x_{jsr}	$\overline{x} = \frac{1}{n}\sum_{j=1}^{r} x_{jsr}^* \cdot f_j^*$

vrijednosti stat. obilj.	MOD
$x_1, x_2, ..., x_n$	x_k ako je je $f_k max$
$x_1^*, x_2^*, ..., x_r^*$	x_k^* ako je je $f_k^* max$
$x_{1sr}, x_{2sr}, ..., x_{rsr}$	x_{ksr}^* ako je je $f_k^* max$
r razreda $[L_j, D_j]$, $j = 1, ...r$	$Mo = L_{Mo} + d \cdot \frac{\Delta_1}{\Delta_1 + \Delta_2}$,
k	ako je f_k max
L_{Mo}	lijevi rub k razreda L_k,
d	širina razreda
Δ_1	$f_k - f_{k-1}$
Δ_2	$f_k - f_{k+1}$

vrijednosti stat. obilj.	MEDIJAN
$x_1, x_2, ..., x_n$ uzlazan niz	$Me = x_{\frac{n+1}{2}}$, za n neparan;
$x_1, x_2, ..., x_n$ uzlazan niz	$Me = \frac{1}{2} \cdot (x_{\frac{n}{2}} + x_{\frac{n}{2}+1})$, za n paran;
r razreda $[L_j, D_j],\ j = 1, ...r$	$Me = L_{Me} + d \cdot \frac{1}{f_{k+1}} \cdot \left(\frac{n}{2} - (f_1 + f_2 + ... + f_k)\right)$
k	$f_1 + f_2 + ... + f_k \leq \frac{n}{2} \leq f_1 + f_2 + ... + f_k + f_{k+1}$
L_{Me}	lijevi rub $k+1$ razreda L_{k+1},
d	širina razreda.

KVANTILI

kvantil reda p	x_p takav da je $F(x_p) = p$
prvi kvartil	$Q_1 = x_{\frac{1}{4}}$
drugi kvartil= MEDIJAN	$Me = x_{\frac{1}{2}}$
treći kvartil	$Q_3 = x_{\frac{3}{4}}$

vrijednosti stat. obilj.	Q_1
$x_1, x_2, ..., x_n$ uzlazan niz	$Q_1 = x_{cijelo(\frac{n}{4}+1)}$, za n nije djeljiv s 4;
$x_1, x_2, ..., x_n$ uzlazan niz	$Me = \frac{1}{2} \cdot (x_{\frac{n}{n}} + x_{\frac{n}{4}+1})$, za n djeljiv s 4;
r razreda $[L_j, D_j],\ j = 1, ...r$	$Q_1 = L_{Q_1} + d \cdot \frac{1}{f_{k+1}} \cdot \left(\frac{n}{4} - (f_1 + f_2 + ... + f_k)\right)$
k	$f_1 + f_2 + ... + f_k \leq \frac{n}{4} \leq f_1 + f_2 + ... + f_k + f_{k+1}$
L_{Q_1}	lijevi rub $k+1$ razreda L_{k+1},
d	širina razreda.

vrijednosti stat. obilj.	Q_3
$x_1, x_2, ..., x_n$ uzlazan niz	$Q_3 = x_{cijelo(\frac{3n}{4}+1)}$, za n nije djeljiv s 4;
$x_1, x_2, ..., x_n$ uzlazan niz	$Me = \frac{1}{2} \cdot (x_{\frac{3n}{n}} + x_{\frac{3n}{4}+1})$, za n djeljiv s 4;
r razreda $[L_j, D_j],\ j = 1, ...r$	$Q_3 = L_{Q_3} + d \cdot \frac{1}{f_{k+1}} \cdot \left(\frac{3n}{4} - (f_1 + f_2 + ... + f_k)\right)$
k	$f_1 + f_2 + ... + f_k \leq \frac{3n}{4} \leq f_1 + f_2 + ... + f_k + f_{k+1}$
L_{Q_3}	lijevi rub $k+1$ razreda L_{k+1},
d	širina razreda.

MJERE RASIPANJA

vrijednosti stat. obilj.	VARIJANCA
$x_1, x_2, ..., x_n$	$\widehat{\sigma}^2 = \frac{1}{n}(x_1^2 + x_2^2 + ... + x_n^2) - \overline{x}^2$
$x_1^*, x_2^*, ..., x_r^*$	$\widehat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{r} (x_j^*)^2 \cdot f_j^* - \overline{x}^2$
r razreda $[L_j, D_j],\ j = 1, ...r; x_{jsr}$	$\widehat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{r} (x_{jsr}^*)^2 \cdot f_j^* - \overline{x}^2$

vrijednosti stat. obilj.	$\widehat{\mu_3}$
$x_1^*, x_2^*, ..., x_r^*$	$\widehat{\mu_3} = \frac{1}{n}\sum_{j=1}^{r}(x_j^* - \overline{x}^2)^3 \cdot f_j^*$
r razreda $[L_j, D_j],\ j = 1, ...r; x_{jsr}$	$\widehat{\mu_3} = \frac{1}{n}\sum_{j=1}^{r}(x_{jsr}^* - \overline{x}^2)^3 \cdot f_j^*$

vrijednosti stat. obilj.	$\widehat{\mu_4}$
$x_1^*, x_2^*, ..., x_r^*$	$\widehat{\mu_4} = \frac{1}{n}\sum_{j=1}^{r}(x_j^* - \overline{x}^2)^4 \cdot f_j^*$
r razreda $[L_j, D_j],\ j = 1, ...r; x_{jsr}$	$\widehat{\mu_4} = \frac{1}{n}\sum_{j=1}^{r}(x_{jsr}^* - \overline{x}^2)^4 \cdot f_j^*$

MJERE OBLIKA

koeficijent asimetrije	$K_A = \frac{\widehat{\mu_3}}{\widehat{\sigma^3}}$
koeficijent spljoštenosti	$K_E = \frac{\widehat{\mu_4}}{\widehat{\sigma^4}} - 3$

Poglavlje 12

TEORIJA PROCJENA

Teorija procjene sastoji se u konstrukciji metoda za ocjenu vrijednosti jednog ili više parametara poznate distribucije slučajne varijable.

U prethodnom poglavlju smo za slučajnu varijablu X (statističko obilježje) promatrali n vrijednosti $x_1, x_2, ..., x_n$ kao uzorak veličina n.
U ovom poglavlju ćemo vrijednosti $x_1, x_2, ..., x_n$ promatrati kao pojedinačne vrijednosti niza od n nezavisnih slučajnih varijabli $X_1, X_2, ..., X_n$ koje imaju istu distribuciju kao i slučajna varijable X.

Definicija 12.1 *(SLUČAJNI UZORAK veličine n)*

Neka je X slučajna varijabla (statističko obilježje populacije) s funkcijom distribucije F(x). Slučajni uzorak veličine n za slučajnu varijablu X je slučajni vektor $(X_1, X_2, ..., X_n)$*, gdje su sve slučajne varijable* X_i*, i=1,...,n, nezavisne sa zajedničkom funkcijom distribucije vjerojatnosti F(x).*
Vrijednost slučajnog uzorka je uređena n-torka $(x_1, x_2, ..., x_n)$ *ako je izmjerena vrijednost slučajnih varijabli* X_i *jednaka* $x_i \in R(X)$*, i=1,...,n.*
Ako je X diskretna slučajna varijabla ($R(X)$ *je konačan ili prebrojiv), onda je* $(X_1, X_2, ..., X_n)$ *diskretni slučajni uzorak , a ako je X kontinuirana slučajna varijabla (*$R(X) \subseteq R$*), onda je* $(X_1, X_2, ..., X_n)$ *kontinuirani slučajni uzorak.*

Definicija 12.2 *(STATISTIKA)*

Ako je $Y = h(X_1, X_2, ..., X_n)$*, gdje je h funkcija od n varijabli, onda se slučajna varijabla Y naziva statistika.*

NAPOMENA 12.1 *Odabrani elementi uzorka veličine n iz populacije trebaju biti izabrani slučajno. Trebamo koristiti tablicu slučajnih brojeva za izbor n slučajnih brojeva ili program za generiranje slučajnih brojeva.*

PRIMJER 12.1 *Zadana je diskretna slučajna varijabla X s funkcijom vjerojatnosti*

x_i	0	1	2
p_i	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

Što je uzorak veličine 2 za ovu slučajnu varijablu?
Odredi sve moguće vrijednosti slučajnog uzorka veličine 2 za X.

Rješenje:
Slučajni uzorak veličine 2 za slučajnu varijablu X je slučajni vektor (X_1, X_2), gdje su sve slučajne varijable X_1 i X_2 nezavisne i jednake funkcije distribucije kao i X. Slika slučajne varijable X je $R(X) = \{0, 1, 2\}$. Slučajne varijable X_1 i X_2 mogu poprimiti iste vrijednosti kao i X. Vrijednost slučajnog uzorka je uređena dvojka (x_1, x_2) elemenata iz $R(X)$, tj. to je varijacija s ponavljanjem $r = 2$-og razreda od $n = 3$ elemenata. Broj svih takvih varijacija je $V_3^{(2)} = 3^2 = 9$.
Sve moguće vrijednosti slučajnog uzorka veličine 2 za slučajnu varijablu X:
(0,0), (0,1), (0,2),
(1,0), (1,1), (1,2),
(2,0), (2,1), (2,2).

12.1 TOČKASTE PROCJENE PARAMETARA

MOTIV 12.1 *Koliki uzorak iz normalne razdiobe s varijancom 81 treba biti da bi s vjerojatnošću 0.9544 apsolutna razlika uzoračke aritmetičke sredine i očekivanja bila manja od 5.5?*

Slučajna varijabla je određena svojom funkcijom distribucije. Mnoga statistička obilježja imaju zajedničku teorijsku funkciju distribucije pa govorimo o poznatim distribucijama (razdiobama): binomna, uniformna, normalna, Poissonova,....
Svaka razdioba karakterizirana je svojim parametrima $n, p, a, b, \mu, \sigma^2, \lambda, ...$:
X $\sim B(n, p)$,
X $\sim U(a, b)$,
X $\sim N(\mu, \sigma^2)$,
X $\sim Po(\lambda)$, ...
Ako želimo odrediti vezu između teorijske i statističke razdiobe postavljaju se dva zadatka:
1. parametarske procjene, kada pretpostavimo teorijsku razdiobu i moramo odrediti (procijeniti) parametre te razdiobe.
2. neparametaske procjene, kada moramo odabrati razdiobu.

Definicija 12.3 *(PROCJENITELJ ILI ESTIMATOR)*
Procjenitelj nepznatog parametra t je funkcija slučajnog uzorka
$\widehat{T} = h(X_1, X_2, ..., X_n)$.
Procjenitelj je statistika.

Zadatak je odrediti procjenitelj $\widehat{T}$ za parametar t koji će "najbolje" procijeniti t.
Za procjenu jednog parametra možemo izabirati razne procjenitelje (funkcije h).

Definicija 12.4 *(NEPRISTRANI PROCJENITELJ)*
Procjenitelj $\widehat{T}$ je nepristran za parametar t ako je očekivanje od $\widehat{T}$ jednako vrijednosti parametra t: $E(\widehat{T}) = t$.

Definicija 12.5 *(ASIMPTOTSKI NORMALAN PROCJENITELJ)*
Procjenitelj $\widehat{T}$ je asimptotski normalan za parametar t ako slučajnoj varijabli $\dfrac{\widehat{T} - t}{\sqrt{Var(\widehat{T})}}$ asimptotski, kad $n \rightarrow \infty$, pripada standardna normalna razdioba (distribucija) N(0,1).

Definicija 12.6 *(UZORAČKA ARITMETIČKA SREDINA)*

Statistika $\overline{X} = h(X_1, X_2, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$ *zove se uzoračka aritmetička sredina. Vrijednost uzoračke aritmetičke sredine računa pomoću*

$\overline{x} = h(x_1, x_2, ..., x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\overline{x} = \frac{1}{n}\sum_{k=1}^{r} x_k^* f_k$.

TEOREM 12.1 *(Svojstva uzoračke aritmetičke sredine)*

(i) *Neka je X slučajna varijabla s teorijskim očekivanjem* μ, *i varijancom* σ^2, $0 < \sigma^2 < \infty$, *koju ispitujemo pomoću slučajnog uzorka* $(X_1, X_2, ..., X_n)$. *Uzoračka aritmetička sredina* $\overline{X}$ *je pouzdan procjenitelj za* μ*:*
$E(\overline{X}) = \mu$.

(ii) $Var(\overline{X}) = \frac{1}{n}\sigma^2$.

(iii) $\overline{X} \sim N(\mu, \frac{1}{n}\sigma^2)$ *ako je normalna distribucija.*

(iv) *Uzoračka aritmetička sredina*$\overline{X}$ *je asimptotski normalan procjenitelj za* μ*:* $\overline{X}^* = \frac{\overline{X}-\mu}{\sqrt{Var(\overline{X})}} \sim N(0, 1)$.

Dokaz:*tko želi znati više*

(i) $E(\overline{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$.

(ii)

$$\begin{aligned} Var(\overline{X}) &= Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}Var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}Var(\sum_{i=1}^{n} X_i) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}n \cdot Var(X_i) = \frac{1}{n}\sigma^2. \end{aligned}$$

(iii) prema (i) i (ii).

(iv) Prisjetimo se centralnog graničnog teorema:

Neka je $S_n = X_1 + X_2 + ... + X_n$,tada slučajna varijabla $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ konvergira k $N(0, 1)$.

$$\overline{X}^* = \frac{\overline{X}-\mu}{\sqrt{Var(\overline{X})}} = \frac{\frac{S_n}{n}-\mu}{\sqrt{\frac{1}{n}\sigma^2}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \quad \text{konvergira } (n \to \infty) \text{ k } N(0, 1).$$

PRIMJER 12.2 *Izračunati* $P(69 < \overline{X} < 75)$, *ako je* $\overline{X}$ *uzoračka aritmetička sredina uzorka veličine n=36 iz normalne razdiobe* $X \sim N(70, 144)$.

Rješenje:
Ako je $X \sim N(70, 144)$ i $n = 36$, onda je $\overline{X} \sim N(70, 4)$,
$\overline{X}^* = \sqrt{n}\frac{\overline{X}-\mu}{\sigma} = \frac{\overline{X}-70}{2} \sim N(0, 1)$.

$$\begin{aligned} P(69 < \overline{X} < 75) &= F^*(\sqrt{n}\frac{75-\mu}{\sigma}) - F^*(\sqrt{n}\frac{59-\mu}{\sigma}) \\ &= F^*(\frac{75-70}{2}) - F^*(\frac{69-70}{2}) = F^*(2.5) - F^*(-0.5) = 0.68. \end{aligned}$$

Prisjetimo se motivacijskog primjera:

PRIMJER 12.3 *motiv*

Koliki uzorak iz normalne razdiobe s varijancom 81 treba biti da bi s vjerojatnošću 0.9544 apsolutna razlika uzoračke aritmetičke sredine i očekivanja bila manja od 5.5?

Rješenje:
Neka je $X \sim N(\mu, 81)$ i $P(|\overline{X} - \mu| < 5.5) \geq 0.9544$.
Trebamo odrediti veličinu uzorka n.
Znamo da je $\overline{X} \sim N(\mu, \frac{81}{n})$, a $\overline{X}^* = \sqrt{n}\frac{\overline{X}-\mu}{\sigma} = \sqrt{n}\frac{\overline{X}-\mu}{9} \sim N(0, 1)$.

$$\begin{aligned} P(|\overline{X} - \mu| < 5.5) &= P(|\sqrt{n}\frac{\overline{X}-\mu}{\sigma}| < \sqrt{n}\frac{5.5}{\sigma}) = P(|\overline{X}^*| \leq \sqrt{n}\frac{5.5}{\sigma}) \\ &= 2F^*(\sqrt{n}\frac{5.5}{\sigma}) - 1. \end{aligned}$$

Iz zadane vjerojatnosti dobivamo: $2F^*(\sqrt{n}\frac{5.5}{\sigma}) - 1 \geq 0.9544$,

$$F^*(\sqrt{n}\frac{5.5}{\sigma}) \geq 0.9772 \quad \Rightarrow \quad \sqrt{n}\frac{5.5}{\sigma} \geq 2 \quad \Rightarrow \quad n \geq 11.$$

Definicija 12.7 *(UZORAČKA VARIJANCA)*

Statistika $\widehat{\Sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ *zove se uzoračka varijanca.*

$$\widehat{\Sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$$

Vrijednost uzoračke varijance računa se formulom

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2.$$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{r}(x_k^* - \overline{x})^2 f_k = \frac{1}{n}\sum_{k=1}^{r}(x_k^*)^2 f_k - \overline{x}^2.$$

TEOREM 12.2 *(Svojstva uzoračke varijance)*

Neka je X slučajna varijabla s teorijskim očekivanjem μ i varijancom σ^2, koju ispitujemo pomoću slučajnog uzorka $(X_1, X_2, ..., X_n)$. Uzoračka varijanca $\widehat{\Sigma}^2$ nije pouzdan procjenitelj za σ^2: $E(\widehat{\Sigma}^2) = \frac{n-1}{n}\sigma^2$.

Dokaz:*tko želi znati više*
Prisjetimo se da je $Var(X)$=$E(X^2) - E(X)^2$ i $Var(\overline{X}) = \frac{1}{n}\sigma^2$.

$$\begin{aligned}
E(\widehat{\Sigma}^2) &= E(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - E(\overline{X}^2) \\
&= \frac{1}{n}\sum_{i=1}^{n}[Var(X_i) + E(X_i)^2] - [Var(\overline{X}) + E(\overline{X})^2] \\
&= [Var(X_i) + E(X_i)^2] - [Var(\overline{X}) + E(\overline{X})^2] \\
&= [\sigma^2 + \mu^2] - [\frac{1}{n}\sigma^2 + \mu^2] = \frac{n-1}{n}\sigma^2.
\end{aligned}$$

Definicija 12.8 *(KORIGIRANA UZORAČKA VARIJANCA)*

Statistika $\widehat{S}^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ zove se korigirana uzoračka varijanca.

$$\widehat{S}^2 = \frac{n}{n-1}\widehat{\Sigma}^2 = \frac{1}{n-1}(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2).$$

Vrijednost korigirane uzoračke varijance računa se formulom

$$\widehat{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2).$$

$$\widehat{s}^2 = \frac{1}{n-1}\sum_{k=1}^{r}(x_k^* - \overline{x})^2 f_k = \frac{1}{n-1}(\sum_{k=1}^{r}(x_k^*)^2 f_k - n\overline{x}^2).$$

TEOREM 12.3 *(Svojstva korigirane uzoračke varijance)*

Neka je X slučajna varijabla s teorijskim očekivanjem μ i varijancom σ^2, koju ispitujemo pomoću slučajnog uzorka $(X_1, X_2, ..., X_n)$.
Korigirana uzoračka varijanca $\widehat{S}^2$ je pouzdan procjenitelj za σ^2: $E(\widehat{S}^2) = \sigma^2$.

Dokaz: Prisjetimo se da je $E(\widehat{\Sigma}^2) = \frac{n-1}{n}\sigma^2$.

$$E(\widehat{S}^2) = E(\frac{n}{n-1}\widehat{\Sigma}^2) = \frac{n}{n-1}E(\widehat{\Sigma}^2) = \frac{n}{n-1} \cdot \frac{n-1}{n}\sigma^2 = \sigma^2.$$

TEOREM 12.4 *(O VEZI $\widehat{S}^2, \widehat{\Sigma}^2$ I DISTRIBUCIJA $\chi^2(n-1), t(n-1)$)*

Neka su $\overline{X}$, $\widehat{S}^2$, $\widehat{\Sigma}^2$ statistike slučajnog uzorka $(X_1, X_2, ..., X_n)$ iz normalne razdiobe $X \sim N(\mu, \sigma^2)$.
Tada vrijedi:
(i) Statistika $\frac{n-1}{\sigma^2}\widehat{S}^2 = \frac{n}{\sigma^2}\widehat{\Sigma}^2 \sim \chi^2(n-1)$,
(ii) Statistika $\sqrt{n}\frac{\overline{X}-\mu}{\widehat{S}} = \sqrt{n-1}\frac{\overline{X}-\mu}{\widehat{\Sigma}} \sim t(n-1)$.

Dokaz: *tko želi znati više*
(i) Dokaz je složen i koristi svojstvo $\chi^2(n)$ distribucije: $Y \sim \chi^2(n)$ ako je
$Y = Y_1^2 + Y_2^2 + ... + Y_n^2,\ Y_i \sim N(0,1)$.
(ii) Koristimo svojstvo Studentove distribucije s n stupnjeva slobode $t(n)$:
$Z \sim t(n)$ ako je $Z = Y\sqrt{\frac{n}{U}}$, za $Y \sim N(0,1),\ U \sim \chi^2(n)$.
Računamo za $X \sim N(\mu, \sigma^2),\ \overline{X} \sim N(\mu, \frac{\sigma^2}{n}),\ \overline{X}^* = \sqrt{n}\frac{\overline{X}-\mu}{\sigma} \sim N(0,1)$:

$$\begin{aligned}\sqrt{n}\frac{\overline{X}-\mu}{\widehat{S}} &= \sqrt{n}\frac{\overline{X}-\mu}{\sigma} \cdot \frac{\sigma\sqrt{n-1}}{\sqrt{n-1}} \cdot \frac{1}{\sqrt{\widehat{S}^2}} = \sqrt{n}\frac{\overline{X}-\mu}{\sigma} \cdot \frac{\sqrt{n-1}}{\sqrt{\frac{n-1}{\sigma^2}}} \cdot \frac{1}{\sqrt{\widehat{S}^2}} \\ &= \sqrt{n}\frac{\overline{X}-\mu}{\sigma} \cdot \frac{\sqrt{n-1}}{\sqrt{\frac{n-1}{\sigma^2} \cdot \widehat{S}^2}}.\end{aligned}$$

Prema tvrdnji (i) zaključujemo $\sqrt{n}\frac{\overline{X}-\mu}{\widehat{S}} \sim t(n-1)$.
Koristeći $\widehat{S}^2 = \frac{n}{n-1}\widehat{\Sigma}^2$ možemo dobiti i tvrdnju $\sqrt{n-1}\frac{\overline{X}-\mu}{\widehat{\Sigma}} \sim t(n-1)$.

PRIMJER 12.4 *Izračunati uzoračku aritmetičku sredinu, uzoračku varijancu i korigiranu uzoračku varijancu u primjeru "težina studenata".*

x_{ksr}	f_k
61	5
64	18
67	42
70	27
73	8
	n=100

Rješenje: $\overline{x} = \frac{1}{n}\sum\limits_{k=1}^{r} x_k^* f_k$.

$$\begin{aligned}\overline{x} &= \frac{1}{n}\sum_{k=1}^{r} x_{ksr}^* f_k = \frac{1}{100}(61*5+64*18+67*42+70*27+73*8)\\ &= \frac{1349}{20} = 67,45\end{aligned}$$

$\widehat{s}^2 = \frac{1}{n-1}(\sum\limits_{k=1}^{r}(x_k^*)^2 f_k - n\overline{x}^2)$

$$\widehat{s}^2 = \frac{1}{n-1}(\sum_{k=1}^{r}(x_{ksr}^*)^2 f_k - n\overline{x}^2) = \frac{379}{44} = 8,6136$$

$\widehat{\sigma}^2 = \frac{1}{n}\sum\limits_{k=1}^{r}(x_k^*)^2 f_k - \overline{x}^2$.

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{r}(x_{ksr}^*)^2 f_k - \overline{x}^2 = \frac{3411}{400} = 8,5275$$

12.2 REGRESIJSKA ANALIZA

MOTIV 12.2 *Deformacije* x [mm] *i Brinellova tvrdoća* y [$\frac{kg}{mm^2}$] *za neki tip čelika dani su tablicom*

x	06	09	11	11	13	22	26	28	33	35
y	68	67	65	53	44	40	37	28	34	32

Odredite pravce regresije, uzorački koeficijent regresije i uzorački koeficijent korelacoje. Jesu li deformacija i Brinellova tvrdoća jako korelirane?

Definicija 12.9 *(UZORAČKA KOVARIJANCA. UZORAČKI KOEFICIJENT KORELACIJE)*
Neka je za zadani slučajni 2-dim vektor (X, Y) *dobiven slučajni uzorak* $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$. *Statistika*

$$\widehat{\mu}_{XY} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$$

zove se uzoračka kovarijanca. Vrijednost korigirane uzoračke kovarijance računa se formulom

$$\widehat{\mu}_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

Neka su $\widehat{\sigma}_1$ *i* $\widehat{\sigma}_2$ *uzoračke standardne devijacije od* X *i* Y. *Uzorački koeficijent korelacije komponenti* X *i* Y *slučajnog vektora je definiran s*

$$\widehat{\rho}_{xy} = \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1 \cdot \widehat{\sigma}_2}.$$

$$\widehat{\rho}_{xy} = \frac{n\sum x \cdot y - (\sum x)\cdot(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\cdot\sqrt{n\sum y^2 - (\sum y)^2}}.$$

NAPOMENA 12.2 *(regresijska analiza)*

Regresijska analiza (engl. regression analysis) je statistička metoda za određivanje veze među slučajnim varijablama. Promatramo u slučajnom vektoru (X, Y) *jednu slučajnu varijablu (npr.* X*) kao nezavisnu-kontroliranu (njene vrijednosti zadajemo). Druga varijabla* Y *je slučajna varijabla i zanima nas kako ona ovisi o* X.
Prema Napomeni 9.4 u poglavlju Dvodimenzionalni slučajni vektor računamo uzoračke pravce regresije.
Ako je X *nezavisna varijabla i* $Y = aX + b$, *parametre* a *i* b *možemo odrediti metodom najmanjih kvadrata tako da* $E((Y - (aX + b))^2)$ *ima minimalnu vrijednost.*

$a = \widehat{\rho}_{xy} \cdot \dfrac{\widehat{\sigma}_2}{\widehat{\sigma}_1} = \dfrac{\widehat{\mu}_{XY}}{\widehat{\sigma}_1^2}$ *je uzorački koeficijent regresije Y po X.*

$b = \overline{y} - \widehat{\rho}_{xy} \cdot \dfrac{\widehat{\sigma}_2}{\widehat{\sigma}_1}\overline{x} = \overline{y} - \dfrac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1^2} \cdot \overline{x}$

$y - \overline{y} = \dfrac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1^2}(x - \overline{x}),$

$y - \overline{y} = \rho_{XY} \cdot \dfrac{\sigma_2}{\sigma_1}(x - \overline{x})$ *je pravac regresije Y po X.*

Analogno, ako je X=aY+b

$a = \widehat{\rho}_{xy} \cdot \dfrac{\widehat{\sigma}_1}{\widehat{\sigma}_2} = \dfrac{\widehat{\mu}_{XY}}{\widehat{\sigma}_2^2}$ *je uzorački koeficijent regresije X po Y*

$x - \overline{x} = \dfrac{\widehat{\mu}_{xy}}{\widehat{\sigma}_2^2}(Y - \overline{y})$

$x - \overline{x} = \widehat{\rho}_{xy} \cdot \dfrac{\widehat{\sigma}_1}{\widehat{\sigma}_2}(y - \overline{y})$ *je pravac regresije X po Y.*

PRIMJER 12.5 *U tablici su zapisani uzorci visina x i y od 12 mama i njihovih kćeri.*

x	165	160	170	163	173	158	178	168	173	170	175	180
y	173	168	173	165	175	168	173	165	180	170	173	178

Oderedite uzorački pravac regresije Y u odnosu na X i uzorački pravac regresije X u odnosu na Y. Odredite uzorački koeficijent korelacije i uzorački koeficijent regresije Y po X. Jesu li visine mama i kćeri jako korelirane?

Rješenje: Trebamo odrediti a i b u jednadžbi $y = ax + b$:

$$a = \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1^2} = \frac{n\sum x \cdot y - (\sum x) \cdot (\sum y)}{n\sum x^2 - (\sum x)^2}$$

$$b = \overline{y} - \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1^2} \cdot \overline{x} = \frac{(\sum y) \cdot (\sum x^2) - (\sum x) \cdot (\sum x \cdot y)}{n\sum x^2 - (\sum x)^2}$$

Računamo: $\sum x = 2033$, $\sum y = 2061$, $\sum x^2 = 344.95$, $\sum x \cdot y = 349.42$, $\sum y^2 = 354.22$.

Uzorački koeficijent regresije Y po X je $a = 0.48$, $b = 90.9$ pa je pravac regresije Y

po X $y = 0.48x + 90.9$.

Trebamo odrediti a' i b' u jednadžbi $x = a'y + b'$:

$$a' = \frac{\widehat{\mu_{xy}}}{\widehat{\sigma_2^2}} = \frac{n\sum x\cdot y - (\sum x)\cdot(\sum y)}{n\sum y^2 - (\sum y)^2}$$

$$b' = \overline{x} - \frac{\widehat{\mu_{xy}}}{\widehat{\sigma_1^2}}\cdot\overline{y} = \frac{(\sum x)\cdot(\sum y^2) - (\sum y)\cdot(\sum x\cdot y)}{n\sum y^2 - (\sum y)^2}$$

Uzorački koeficijent regresije X po Y je $a' = 1.02$, $b' = -5.12$ pa je pravac regresije X po Y $x = 1.02x - 5.12$.

Uzorački koeficijent korelacije je

$$\widehat{\rho_{xy}} = \frac{\widehat{\mu_{xy}}}{\widehat{\sigma_1}\cdot\widehat{\sigma_2}}.$$

$$\widehat{\rho_{xy}} = \frac{n\sum x\cdot y - (\sum x)\cdot(\sum y)}{\sqrt{n - \sum x^2 - (\sum x)^2}\cdot\sqrt{n - \sum y^2 - (\sum y)^2}} = 0.69.$$

Prisjetimo se motivacijskog primjera:

PRIMJER 12.6 *motiv*

Deformacije x [mm] i Brinellova tvrdoća y $[\frac{kg}{mm^2}]$ za neki tip čelika dani su tablicom

x	06	09	11	11	13	22	26	28	33	35
y	68	67	65	53	44	40	37	28	34	32

Odredite pravce regresije , uzorački koeficijent korelacije i uzorački koeficijent regresije. Jesu li deformacije i Brienellova tvrdoća jako korelirane ?

Rješenje: Trebamo odrediti a i b u jednadžbi $y = ax + b$:

$$a = \frac{\widehat{\mu_{xy}}}{\widehat{\sigma_1^2}} = \frac{n\sum x\cdot y - (\sum x)\cdot(\sum y)}{n\sum x^2 - (\sum x)^2}$$

$$b = \overline{y} - \frac{\widehat{\mu_{xy}}}{\widehat{\sigma_1^2}}\cdot\overline{x} = \frac{(\sum y)\cdot(\sum x^2) - (\sum x)\cdot(\sum x\cdot y)}{n\sum x^2 - (\sum x)^2}$$

Računamo: $\sum x = 183$, $\sum y = 440$, $\sum x^2 = 4665$, $\sum x\cdot y = 7701$, $\sum y^2 = 23232$. Uzorački koeficijent regresije Y po X je $a = -1.32$, $b = 75.72$ pa je pravac regresije

Y po X $y = -1.32x + 75.72$.

Trebamo odrediti a' i b' u jednadžbi $x = a'y + b'$:

$$a' = \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_2^2} = \frac{n \sum x \cdot y - (\sum x) \cdot (\sum y)}{n \sum y^2 - (\sum y)^2}$$

$$b' = \overline{x} - \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1^2} \cdot \overline{y} = \frac{(\sum x) \cdot (\sum y^2) - (\sum y) \cdot (\sum x \cdot y)}{n \sum y^2 - (\sum y)^2}$$

Uzorački koeficijent regresije X po Y je $a' = -0.72$, $b' = 55.72$ pa je pravac regresije X po Y $x = -0.72x + 55.72$.

Uzorački koeficijent korelacije je

$$\widehat{\rho}_{xy} = \frac{n \sum x \cdot y - (\sum x) \cdot (\sum y)}{\sqrt{n - \sum x^2 - (\sum x)^2} \cdot \sqrt{n - \sum y^2 - (\sum y)^2}} = -0.97$$

Budući je $\widehat{\rho}_{xy} \approx -1$ slučajne varijable su linearnoj vezi, jako korelirane.

12.3 METODA NAJVEĆE VJEROJATNOSTI (ML) *tko želi znati više*

U ovom poglavlju istaknuli smo primjere s oznakom *važno* u kojima su dani procjenitelji za parametre osnovnih distribucija u smislu najveće vjerojatnosti.

MOTIV 12.3 *U četiri mjerenja Rockwellove tvrdoće jedne ploče radnici du dobili sljedeće vrijednosti:*
$64.9, 64.1, 63.8, 64.0.$
(a) Izračunajte vrijednost procjenitelja (u smislu najveće vjerojatnosti) za očekivanje i varijancu tvrdoće ako pretpostavimo normalnu distribuciju.
(b) Izračunajte vrijednost nepristranih procjenitelja za očekivanje i varijancu tvrdoće ako pretpostavimo normalnu distribuciju.

Definicija 12.10 *(FUNKCIJA VJERODOSTOJNOSTI) Neka je X slučajna varijabla (statističko obilježje) sa teorijskom funkcijom distribucije F(x,t) s nepoznatim parametrom t i sa funkcijom vjerojatnosti za diskretnu razdiobu i funkcijom gustoće vjerojatnosti za kontinuiranu f(x,t). Neka je* $(x_1, x_2, ..., x_n)$ *vrijednost slučajnog uzorka* $(X_1, X_2, ..., X_n)$ *za promatranu varijablu.*
Za diskretnu razdiobu funkcija vjerodostojnosti $L(t)$ *definira se kao funkcija vjerojatnosti slučajnog uzorka (slučajnog vektora):*

$$L(t) = P(X_1 = x_1, t) \cdot P(X_1 = x_2, t) \cdot ... \cdot P(X_1 = x_n, t).$$

Za kontinuiranu razdiobu funkcija vjerodostojnosti $L(t)$ *definira se kao funkcija gustoće vjerojatnosti slučajnog uzorka (slučajnog vektora):*

$$L(t) = f(x_1, t) \cdot f(x_2, t) \cdot ... \cdot f(x_n, t).$$

Metoda najveće vjerojatnosti (ML = maximum likelihood method), za određivanje procjenitelja $\widehat{T} = h(X_1, X_2, ..., X_n)$ za nepoznati parametar t sastoji se u izboru one funkcije h takve da funkcija vjerodostojnosti $L(t)$ (ili $lnL(t)$) dostiže najveću vrijednost za $t = h(x_1, x_2, ..., x_n)$.

$$\frac{d}{dt}L(t) = 0 \quad \Rightarrow \quad t = h(x_1, x_2, ..., x_n) \quad \Rightarrow \quad \widehat{T} = h(X_1, X_2, ..., X_n).$$

PRIMJER 12.7 *važno*

Metodom najveće vjerojatnosti pokažite da je procjenitelj za parametar λ *u populaciji s Poisonovom razdiobom* $Po(\lambda)$ *jednak* $\widehat{T} = \overline{X}$.

Rješenje: Neka je $X \sim Po(\lambda)$.
Teorijska funkcija vjerojatnosti je $f(x,\lambda)$=$P(X=x,\lambda)=\frac{\lambda^x}{x!}e^{\lambda}$.
Trebamo naći ML-procjenitelj za λ.
Funkcija vjerodostojnosti je

$$\begin{aligned} L(\lambda) &= P(X_1=x_1,\lambda)\cdot P(X_1=x_2,\lambda)\cdot ...\cdot P(X_1=x_n,\lambda) \\ &= \frac{\lambda^{x_1}}{x_1!}e^{\lambda}\cdot \frac{\lambda^{x_2}}{x_2!}e^{\lambda}\cdot ...\cdot \frac{\lambda^{x_n}}{x_n!}e^{\lambda} = \frac{\lambda^{\sum x_i}}{x_1!x_2!...x_n!}e^{-n\lambda}. \end{aligned}$$

Tražimo maksimum funkcije vjerodostojnosti $\ln L(\lambda)$:

$$\ln L(\lambda) = -n\lambda + \sum x_i \ln\lambda - \sum \ln(x_i!),$$

$$\frac{d}{d\lambda}\ln L(\lambda) = -n + \frac{1}{\lambda}\sum x_i.$$

$$\frac{d}{d\lambda}\ln L(\lambda) = 0 \quad\Rightarrow\quad \lambda = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}.$$

$$\lambda = h(x_1,x_2,...,x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i \quad\Rightarrow\quad \widehat{T} = h(X_1,X_2,...,X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

ML-procjenitelj za λ u Poisonovoj razdiobi je $\widehat{T}=\overline{X}$.
Možemo pokazati da je $\widehat{T}$ nepristrani procjenitelj $E(\widehat{T})=\lambda$.
Prisjetimo se da je $E(X)$=λ i da je $\overline{X}$ nepristrani procjenitelj za očekivanje.
$E(\widehat{T})=E(\overline{X})=E(X)=\lambda$.

PRIMJER 12.8 *važno*

Metodom najveće vjerojatnosti pokažite da je procjenitelj
(a) za parametar μ *u populaciji s Normalnom razdiobom* $N(\mu,\sigma^2)$ *ako je* σ^2 *poznato jednak* $\widehat{T}=\overline{X}$
(b) za parametar σ^2 *u populaciji s Normalnom razdiobom* $N(\mu,\sigma^2)$ *ako je* μ *poznato jednak* $\widehat{T}=\widehat{\Sigma}^2$.

Rješenje: Neka je $X \sim N(\mu,\sigma^2)$.
Teorijska funkcija gustoće vjerojatnosti je $f(x,\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.
Trebamo naći ML-procjenitelje za μ i σ^2.
Funkcija vjerodostojnosti je:

$$L(\mu,\sigma^2) = f(x_1,t)\cdot f(x_2,t)\cdot ...\cdot f(x_n,t) = (\frac{1}{\sigma\sqrt{2\pi}})^n \cdot e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2}.$$

$$\ln L(\mu, \sigma^2) = -\frac{1}{2\sigma^2} \sum (x_i - \mu)^2 - n \ln \sigma - n \ln \sqrt{2\pi}.$$

(a) $\frac{d}{d\mu} \ln L(\mu, \sigma^2) = -\frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0,$

$$\frac{d}{d\mu} \ln L(\mu, \sigma^2) = 0 \quad \Rightarrow \quad \mu = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x},$$

ML-procjenitelj za očekivanje u normalnoj razdiobi je $\widehat{T} = \overline{X}$. To je nepristrani procjenitelj za μ jer je $E(\widehat{T})=\mu$.

(b) $\frac{d}{d\sigma} \ln L(\mu, \sigma^2) = -\frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i - \mu)^2 - \frac{n}{\sigma} = 0,$

$$\frac{d}{d\sigma} \ln L(\mu, \sigma^2) = 0 \quad \Rightarrow \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 = \widehat{\sigma}, \text{ uzoračka varijanca.}$$

ML-procjenitelj za varijancu u normalnoj razdiobi je $\widehat{T} = \widehat{\Sigma}^2$. To nije nepristrani procjenitelj za σ^2 jer je $E(\widehat{T})=\frac{n-1}{n}\sigma^2$.

Prisjetimo se motivacijskog primjera:

PRIMJER 12.9 *motiv*

U četiri mjerenja Rockwellove tvrdoće jedne ploče radnici du dobili sljedeće vrijednosti: 64.9, 64.1, 63.8, 64.0. *(a) Izračunajte vrijednost procjenitelja (u smislu najveće vjerojatnosti-ML-procjenitelj) za očekivanje i varijancu tvrdoće ako pretpostavimo normalnu distribuciju.*
(b) Izračunajte vrijednost nepristranih procjenitelja za očekivanje i varijancu tvrdoće ako pretpostavimo normalnu distribuciju.

Rješenje: (a) ML-procjenitelj za očekivanje μ je uzoračka aritmetička sredina $\overline{X}$.
$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = 64.2$

ML-procjenitelj za varijancu σ^2 je uzoračka varijanca $\widehat{\Sigma}^2$.
$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 = 0.175$

(b)Nepristrani procjenitelj za očekivanje μ je uzoračka aritmetička sredina $\overline{X}$.
$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = 64.2$

Nepristrani procjenitelj za varijancu σ^2 je korigirana uzoračka varijanca $\widehat{S}^2$.
$\widehat{s}^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} (x_i - \overline{x})^2 = 0.233.$

PRIMJER 12.10 *važno*

Metodom najveće vjerojatnosti pokažite da je procjenitelj za parametar p u populaciji s Binomnom razdiobom $B(m,p)$ uz pretpostavku da je m poznato jednak $\widehat{T} = \frac{\overline{X}}{m}$

Rješenje: Neka je $X \sim B(m,p)$. Teorijska funkcija vjerojatnosti je $f(x,m,p)$=$P(X = x, m, p) = \binom{m}{x}(1-p)^{m-x}p^x$. Trebamo naći ML-procjenitelj za p. Funkcija vjerodostojnosti je

$$\begin{aligned} L(p) &= P(X_1 = x_1, m, p) \cdot P(X_1 = x_2, m, p) \cdot \ldots \cdot P(X_1 = x_n, m, p) \\ &= \prod_{i=1}^{n} \binom{m}{x_i}(1-p)^{m-x_i}p^{x_i} \end{aligned}$$

Tražimo maksimum funkcije vjerodostojnosti $\ln L(p)$:

$$\ln L(p) = \sum [\ln \binom{m}{x_i} + (m - x_i)\ln(1-p) + x_i \ln p],$$

$$\frac{d}{dp}\ln L(p) = -\frac{1}{1-p}\sum (m - x_i) + \frac{1}{p}\sum x_i$$

$$\frac{d}{d\lambda}\ln L(\lambda) = 0 \quad \Rightarrow \quad p = \frac{1}{m}\cdot\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{m}\overline{x}.$$

$p = h(x_1, x_2, ..., x_n) = \frac{1}{m}\cdot\frac{1}{n}\sum\limits_{i=1}^{n} x_i = \frac{1}{m}\overline{x}$

$\widehat{T} = h(X_1, X_2, ..., X_n) = \frac{1}{m}\cdot\frac{1}{n}\sum\limits_{i=1}^{n} X_i = \frac{1}{m}\overline{X}$.

ML-procjenitelj za p u Binomnoj razdiobi $B(m,p)$ s poznatim m je $\widehat{T} = \frac{1}{m}\overline{X}$.
Možemo pokazati da je $\widehat{T}$ nepristrani procjenitelj $E(\widehat{T}) = p$.
Prisjetimo se da je $E(X)$=mp i da je $\overline{X}$ nepristrani procjenitelj za očekivanje.

$$E(\widehat{T}) = E(\frac{1}{m}\overline{X}) = \frac{1}{m}E(\overline{X}) = \frac{1}{m}E(X) = \frac{1}{m}mp = p.$$

12.4 Ponovimo

PROCJENITELJI PARAMETARA ZADANE DISTRIBUCIJE s očekivanjem μ i varijancom σ^2

slučajni uzorak	$(X_1, X_2, ..., X_n) : \Omega \to \mathbb{R}^n$
statistika	$T = h(X_1, X_2, ..., X_n)\ ,\ T : \Omega \to \mathbb{R}$
procjenitelj	$\widehat{T} = h(X_1, X_2, ..., X_n)$
nepristrani procjenitelj za parametar t	$E(\widehat{T}) = t$
uzoračka aritm. sredina	$\overline{X} = \frac{1}{n}(X_1 + X_2 + ... + X_n)$
	$E(\overline{X}) = \mu,$
	$Var(\overline{X}) = \frac{\sigma^2}{n}$
$n \to \infty$	$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$
uzoračka varijanca	$\widehat{\Sigma}^2 = \frac{1}{n}\sum_i (X_i - \overline{X})^2$
	$\widehat{\sigma}^2 = \frac{1}{n}\sum_i (x_i - \overline{x})^2$
uzoračka stand. devijacija	$\widehat{\sigma}$
	$E(\widehat{\Sigma}^2) = \frac{n-1}{n}\sigma^2$
korigirana uzoračka varijanca	$\widehat{S}^2 = \frac{1}{n-1}\sum_i (X_i - \overline{X})^2$
	$\widehat{s}^2 = \frac{1}{n-1}\sum_i (x_i - \overline{x})^2$
korigirana uz. stand. devijacija	$\widehat{s}$
	$E(\widehat{S}^2) = \sigma^2$

PROCJENITELJI PARAMETARA ZADANE DISTRIBUCIJE (ML)

distribucija ; parametar	procjenitelj
$Po(\lambda); \lambda$	$\widehat{T} = \overline{X}$
$N(\mu, \sigma^2);\ \ \mu$	$\widehat{T} = \overline{X}$
$N(\mu, \sigma^2);\ \ \sigma^2$	$\widehat{T} = \widehat{\Sigma}^2$
$(B(m, p);\ \ p$	$\widehat{T} = \frac{1}{m}\overline{X}$

STATISTIKE PARAMETARA NORMALNE DISTRIBUCIJE $N(\mu, \sigma^2)$

slučajni uzorak za X	$(X_1, X_2, ..., X_n) : \Omega \to \mathbb{R}^n$
statistika $T = h(X_1, X_2, ..., X_n)$	distribucija
$\overline{X}$	$\sim N(\mu, \frac{\sigma}{\sqrt{n}})$
$\sqrt{n}\frac{\overline{X}-\mu}{\sigma}$	$\sim N(0,1)$
$\sqrt{n}\frac{\overline{X}-\mu}{\widehat{S}}$	$\sim t(n-1)$
$\frac{n-1}{\sigma^2}\widehat{S}^2$	$\sim \chi^2(n-1)$

REGRESIJSKA ANALIZA

slučajni uzorak za (X, Y)	$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
uzoračka kovarijanca	$\widehat{\mu}_{xy} = \frac{1}{n}\sum_i (x_i - \overline{x})(y_i - \overline{y})$
uzorački koef. korelacije	$\widehat{\rho}_{xy} = \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1 \cdot \widehat{\sigma}_2}$
	$\frac{n\sum x\cdot y - (\sum x)\cdot(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\cdot\sqrt{n\sum y^2 - (\sum y)^2}}$
$Y = aX + b$	a- uzorački koef. regresije
a	$\frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1^2}$
	$= \frac{n\sum x\cdot y - (\sum x)\cdot(\sum y)}{n\sum x^2 - (\sum x)^2}$
b	$\overline{y} - \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1^2} \cdot \overline{x}$
	$= \frac{(\sum y)\cdot(\sum x^2) - (\sum x)\cdot(\sum x\cdot y)}{n\sum x^2 - (\sum x)^2}$
pravac regresije Y po X	$y - \overline{y} = \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1^2}(x - \overline{x})$
pravac regresije X po Y	$x - \overline{x} = \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_2^2}(y - \overline{y})$

Poglavlje 13

INTERVAL POVJERENJA ZA OČEKIVANJE

Definicija 13.1 *(KVANTIL)*

Neka je X slučajna varijabla s funkcijom distribucije $F(x)$i neka je zadan $q \in (0, 1)$. Broj z_q zove se kvantil distribucije F ako vrijedi $F(z_q) = q$.

Definicija 13.2 *(INTERVAL POVJERENJA POUZDANOSTI γ)*

Neka je $(X_1, X_2, ..X_n)$ slučajni uzorak slučajne varijable X koja ima poznatu distribuciju s nepoznatim parametrom t i neka je zadana pouzdanost $\gamma \in (0, 1)$.
Za procjenitelje $G_1 = h_1(X_1, X_2, ..X_n)$ i $G_2 = h_2(X_1, X_2, ..X_n)$ za parametar t kažemo da čine interval povjerenja (G_1, G_2) za parametar t s pouzdanošću γ ako vrijedi: $P(G_1 < t < G_2) \geq \gamma$.
Parametar t poprimit će vrijednosti unutar intervala (g_1, g_2) s puzdanošću γ, gdje je $g_1 = h_1(x_1, x_2, ..x_n)$, $g_2 = h_2(x_1, x_2, ..x_n)$.

13.1 INTERVAL POVJERENJA ZA OČEKIVANJE ako je veliki uzorak $n \to \infty$

MOTIV 13.1 *Slučajni uzorak od 50 studenata za broj bodova (max 100 bodova) iz VIS-a od ukupno 200 studenata ove generacije pokazuje uzoraču aritmetičku sredinu 75 i korigiranu uzoračku standardnu devijaciju 10.*
(a) Odredite interval povjerenja za očekivani broj bodova iz VIS-a za ovu genaraciju studenata s pouzdanošću 95%.
(b) Kolika je pouzdanost da će interval povjerenja za očekivani broj bodova biti $[74, 76]$*?*
(c) Odredite interval povjerenja za očekivani broj bodova iz VIS-a s pouzdanošću 95% *ako je ovo bio uzorak iz podataka za sve generacije studenata koje je nastavnik vodio.*

TEOREM 13.1 *Neka je* $(X_1, X_2, ..X_n)$ *slučajni uzorak slučajne varijable X koja ima poznatu distribuciju s nepoznatim parametrom očekivanje* μ *i poznatom varijancom* σ^2 *(ili poznatom korigiranom uzoračkom varijancom* $\widehat{s}^2$*).*
Ako je veliki uzorak $(n \to \infty)$ *onda inerval povjerenja* (G_1, G_2) *za parametar očekivanje* μ *s pouzdanošću* γ *čine procjenitelji*

$$G_1 = \overline{X} - \lambda \frac{\sigma}{\sqrt{n}} \ \ i \ \ G_2 = \overline{X} + \lambda \frac{\sigma}{\sqrt{n}},$$

ili

$$G_1 = \overline{X} - \lambda \frac{\widehat{s}}{\sqrt{n}} \ \ i \ \ G_2 = \overline{X} + \lambda \frac{\widehat{s}}{\sqrt{n}},$$

gdje je $\overline{X}$ *uzoračka aritmetička sredina, a* $\lambda = z_{\frac{1+\gamma}{2}}$ *kvantil standardne normalne distribucije* $F^*(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.

Dokaz: *tko želi znati više*

Prema Centralnom graničnom teormu za aritmetčku sredinu

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1), \qquad n \to \infty.$$

Primijenimo CGT na simetrični interval $(-\lambda, \lambda)$:
$P(-\lambda < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \lambda) \approx F^*(\lambda) - F^*(-\lambda) = 2F^*(\lambda) - 1$
tj.
$P(\overline{X} - \lambda \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + \lambda \frac{\sigma}{\sqrt{n}}) \approx 2F^*(\lambda) - 1.$

Ako je zadana pozdanost γ, $P(G_1 < \mu < G_2) = \gamma$, onda možemo odrediti λ tako da vrijedi $F^*(\lambda) = \frac{1+\gamma}{2}$ tj. $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil standardne normalne distribucije $F^*(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.
Zaključujemmo da je za velike n $P(\overline{X} - \lambda\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + \lambda\frac{\sigma}{\sqrt{n}}) = \gamma$.
Procjenitelji $G_1 = \overline{X} - \lambda\frac{\sigma}{\sqrt{n}}$, $G_2 = \overline{X} + \lambda\frac{\sigma}{\sqrt{n}}$ čine interval povjerenja

$$(\overline{X} - \lambda\frac{\sigma}{\sqrt{n}}, \overline{X} + \lambda\frac{\sigma}{\sqrt{n}})$$

za parametar očekivanja μ slučajne varijable X s puzdanošću γ ako je poznata varijanca σ^2.
Parametar očekivanja μ s pozdanošću γ poprimit će vrijednosti u u intervalu $(\overline{x} - \lambda\frac{\sigma}{\sqrt{n}}, \overline{x} + \lambda\frac{\sigma}{\sqrt{n}})$, gdje je $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil standardne normalne distribucije $F^*(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.

NAPOMENA 13.1 *Ova procjena parametra očekivanja slučajne varijable može se koristi u zadacima za određivanje*
(a) $\delta = 2\lambda\frac{\sigma}{\sqrt{n}}$ *širine inervala*
(b) $n = 4\lambda^2\frac{\sigma^2}{\delta^2}$ *minimalne veličine uzorka*
uz zadanu pozdanost γ *za interval povjerenja* $(\overline{X} - \lambda\frac{\sigma}{\sqrt{n}}, \overline{X} + \lambda\frac{\sigma}{\sqrt{n}})$, *gdje je* $\lambda = z_{\frac{1+\gamma}{2}}$ *kvantil standardne normalne distribucije* $F^*(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.

NAPOMENA 13.2 *(uzorak bez vračanja u konačnoj populaciji)*

Neka je $(X_1, X_2, ..X_n)$ *slučajni uzorak slučajne varijable X koja ima poznatu distribuciju s nepoznatim parametrom očekivanje* μ *i poznatom varijancom* σ^2 *(ili poznatom korigiranom uzoračkom varijancom* $\widehat{s^2}$*).*
Neka je uzet uzorak bez vraćanja iz konačne populacije veličine N. Ako je veliki uzorak $(n \to \infty)$, $(\ n \leq 30)$ *onda interval povjerenja* (G_1, G_2) *za parametar očekivanja* μ *s puzdanošću* γ *čine procjenitelji*

$$G_1 = \overline{X} - \lambda\frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}},$$

$$G_2 = \overline{X} + \lambda\frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}},$$

ili

$$G_1 = \overline{X} - \lambda\frac{\widehat{s}}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}},$$

$$G_2 = \overline{X} + \lambda \frac{\widehat{s}}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}},$$

gdje je $\overline{X}$ *uzoračka aritmetička sredina, a* $\lambda = z_{\frac{1+\gamma}{2}}$ *kvantil standardne normalne distribucije* $F^*(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.

Prisjetimo se motivacijskog primjera:

PRIMJER 13.1 *motiv*

Slučajni uzorak od 50 studenata za broj bodova (max 100 bodova) iz VIS-a od ukupno 200 studenata ove generacije pokazuje uzoraču aritmetičku sredinu 75 i korigiranu uzoračku standardnu devijaciju 10.
(a) Odredite interval povjerenja za očekivani broj bodova iz VIS-a za ovu genaraciju studenata s pouzdanošću 95%.
(b) Kolika je pouzdanost da će interval povjerenja za očekivani broj bodova biti $[74, 76]$*?*
(c) Odredite interval povjerenja za očekivani broj bodova iz VIS-a s pouzdanošću 95% *ako je ovo bio uzorak iz podataka za sve generacije studenata koje je nastavnik vodio.*

Rješenje: Pretpostavljamo da je X slučajna varijabla (statističko obilježe) broj bodova na ispitu iz VIS-a.
Prema prethodnoj napomeni odredit ćemo interval povjerenja za očekivanje ako je veličina uzorka velika $n > 30$ za konačnu populaciju $N = 200$ pri uzorku bez vraćanja.

Iz zadatka iščitavamo podatke: $\widehat{s} = 10$, $N = 200$, $n = 50$, $\overline{x} = 75$.

(a) Za pouzdanost $\gamma = 0.95$ koeficijent $\lambda = 1.96$.

$$g_1 = \overline{x} - \lambda \frac{\widehat{s}}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = 75 - 2.4,$$

$$g_2 = \overline{x} + \lambda \frac{\widehat{s}}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = 75 + 2.4$$

Uz pouzdanost 95% broj bodova za kolegij VIS za studente ove generacije je unutar intervala $[75 - 2.4, 75 + 2.4]$.

(b) Ako je širina intervala $\delta = 2$ onda je

$$\delta = 2\lambda \cdot \frac{\widehat{s}}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = 2$$

pa je $2\lambda \cdot 1.23 = 2$ Dobili smo vrijednost $\lambda = 0.81$.
Veza λ i γ dana izrazom $\lambda = z_{\frac{1+\gamma}{2}}$. Kako je kvantil $z_{\frac{1+\gamma}{2}} = 0.81$ iz tablice za normalnu razdiobu očitamo da je $F(0.81) = 0.791$.
Pouzdanost γ određujemo iz izraza $\frac{1+\gamma}{2} = 0.791$
Tražena puzdanost je $\gamma = 0.582$.

Uz pouzdanost 58.2% broj bodova za kolegij VIS u populaciji ove generacije je unutar intervala $[75-1, 75+1]$.

(c) U ovom slučaju imamo veliku populaciju pa je interval povjerenja (G_1, G_2) za parametar očekivanja μ s puzdanošću γ

$$G_1 = \overline{X} - \lambda \frac{\widehat{s}}{\sqrt{n}}, \;\; G_2 = \overline{X} + \lambda \frac{\sigma}{\sqrt{n}},$$

gdje je $\overline{X}$ uzoračka aritmetička sredina, a $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil standardne normalne distribucije $F^*(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.

Za pouzdanost $\gamma = 0.95$ koeficijent $\lambda = 1.96$.

$$g_1 = \overline{x} - \lambda \frac{\widehat{s}}{\sqrt{n}} = 75 - 2.77$$

$$g_2 = \overline{X} + \lambda \frac{\widehat{s}}{\sqrt{n}} = 75 + 2.77$$

Uz pouzdanost 95% broj bodova za kolegij VIS u cjelokupnoj populaciji je unutar intervala $[75 - 2.77, 75 + 2.77]$.

13.2 INTERVAL POVJERENJA ZA OČEKIVANJE NORMALNE distribucije ako je varijanca poznata

MOTIV 13.2 *Neka slučajna varijabla ima normalnu distribuciju $X \sim N(\mu, \sigma^2)$ nepoznatog očekivanja i poznate varijance $\sigma^2 = 0.64$. Koliki minimalni uzorak treba uzeti da bi greška procjene očekivanja μ bila najviše jednaka 0.5, uz pouzdanost $\gamma = 0.95$?*

TEOREM 13.2 *Neka je $(X_1, X_2, ..X_n)$ slučajni uzorak slučajne varijable $X \sim N(\mu, \sigma^2)$ s nepoznatim parametrom očekivanje μ i poznatom varijancom σ^2. Interval povjerenja (G_1, G_2) za parametar očekivanja μ s puzdanošću γ čine procjenitelji*

$$G_1 = \overline{X} - \lambda \frac{\sigma}{\sqrt{n}}, \;\; G_2 = \overline{X} + \lambda \frac{\sigma}{\sqrt{n}},$$

gdje je $\overline{X}$ uzoračka aritmetička sredina, a $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil standardne normalne distribucije $F^*(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.

Dokaz: *tko želi znati više*

Neka je $X \sim N(\mu, \sigma^2)$, onda je $\overline{X} \sim N(\mu, \frac{1}{n}\sigma^2)$, a $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$.
Na simetričnom intervalu $(-\lambda, \lambda)$:
$P(-\lambda < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \lambda) = F^*(\lambda) - F^*(-\lambda) = 2F^*(\lambda) - 1$
tj.
$P(\overline{X} - \lambda \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + \lambda \frac{\sigma}{\sqrt{n}}) = 2F^*(\lambda) - 1.$
Ako je zadana pozdanost γ, $P(G_1 < \mu < G_2) = \gamma$, onda možemo odrediti λ tako da vrijedi $F^*(\lambda) = \frac{1+\gamma}{2}$ tj. $\lambda = z_{\frac{1+\gamma}{2}}$ je kvantil standardne normalne distribucije $F^*(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.
Procjenitelji $G_1 = \overline{X} - \lambda \frac{\sigma}{\sqrt{n}}$, $G_2 = \overline{X} + \lambda \frac{\sigma}{\sqrt{n}}$ čine interval povjerenja

$$(\overline{X} - \lambda \frac{\sigma}{\sqrt{n}}, \; \overline{X} + \lambda \frac{\sigma}{\sqrt{n}})$$

za parametar očekivanja μ slučajne varijable $X \sim N(\mu, \sigma^2)$ s pouzdanošću γ ako je poznata varijanca σ^2.
Parametar očekivanja μ s pozdanošću γ poprimit će vrijednosti u u intervalu

$(\overline{x} - \lambda \frac{\sigma}{\sqrt{n}}, \overline{x} + \lambda \frac{\sigma}{\sqrt{n}})$, gdje je $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil standardne normalne distribucije $F^*(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.

NAPOMENA 13.3 *Ova procjena parametra očekivanja slučajne varijable $X \sim N(\mu, \sigma^2)$ poznate varijance, može se koristi u zadacima za određivanje*
(a) $\delta = 2\lambda \frac{\sigma}{\sqrt{n}}$ širine inervala
(b) $n = 4\lambda^2 \frac{\sigma^2}{\delta^2}$ minimalne veličine uzorka
uz zadanu pozdanost γ za interval povjerenja $(\overline{X} - \lambda \frac{\sigma}{\sqrt{n}}, \overline{X} + \lambda \frac{\sigma}{\sqrt{n}})$, gdje je $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil standardne normalne distribucije $F^*(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.

NAPOMENA 13.4 *Kvantili za standardnu normalnu distribuciju*
$F^*(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$:

γ	0.90	0.95	0.99
$\frac{1+\gamma}{2}$	0.95	0.975	0.995
$z_{\frac{1+\gamma}{2}}$	1.65	1.96	2.58

PRIMJER 13.2 *Neka slučajna varijabla ima normalnu distribuciju*
$X \sim N(\mu, \sigma^2 = 0.64)$. *Uzet je uzorak veličine $n = 5$ i dobivena je vrijednost uzoračke aritmetičke sredine $\overline{x} = 10.2$. Odredite interval povjerenja za očekivanje slučajne varijable s pouzdanošću $\gamma = 0.95$.*

Rješenje: $P(\overline{X} - \lambda \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + \lambda \frac{\sigma}{\sqrt{n}}) = \gamma$.
Za očekivanje μ interval povjerenja pouzdanosti γ je $(\overline{X} - \lambda \frac{\sigma}{\sqrt{n}}, \overline{X} + \lambda \frac{\sigma}{\sqrt{n}})$, gdje je $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil standardne normalne distribucije.
Iz tablice očitavamo za $\gamma = 0.95$, $\lambda = z_{\frac{1+\gamma}{2}} = 1.96$.
Koristeći date vrijednosti iz uzorka $n = 5$, $\overline{x} = 10.2$ dobivamo interval povjerenja za μ :

$$(\overline{x} - \lambda \frac{\sigma}{\sqrt{n}}, \overline{x} + \lambda \frac{\sigma}{\sqrt{n}}) = (10.2 - 1.96 * \frac{0.8}{\sqrt{5}}, 10.2 + 1.96 * \frac{0.8}{\sqrt{5}}) = (9.49, 10.90).$$

Prisjetimo se motivacijskog primjera:

PRIMJER 13.3 *motiv*

Neka slučajna varijabla ima normalnu distribuciju
$X \sim N(\mu, \sigma^2)$ *nepoznatog očekivanja i poznate varijance* $\sigma^2 = 0.64$. *Koliki minimalni uzorak treba uzeti da bi greška procjene očekivanja μ bila najviše jednaka 0.5, uz pouzdanost* $\gamma = 0.95$?

Rješenje: $n = 4\lambda^2 \frac{\sigma^2}{\delta^2} = \lambda^2 \frac{\sigma^2}{greska^2}$, gdje je $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil standardne normalne, $F^*(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.
Iz tablice očitavamo za $\gamma = 0.95$, $\lambda = z_{\frac{1+\gamma}{2}} = 1.96$.

$$n = 1.96^2 * \frac{1}{0.5^2} * 0.64 = 9.8345.$$

13.3 INTERVAL POVJERENJA ZA OČEKIVANJE NORMALNE distribucije ako je varijanca nepoznata

MOTIV 13.3 *U četiri mjerenja Rockwellove tvrdoće jedne ploče radnici du dobili sljedeće vrijednosti:*
$64.9, 64.1, 63.8, 64.0.$
Odredite interval povjerenja za očekivanu vrijednost tvrdoće s pouzdanošću 99%.

TEOREM 13.3 *Neka je* $(X_1, X_2, ..X_n)$ *slučajni uzorak slučajne varijable* $X \sim N(\mu, \sigma^2)$ *s nepoznatim parametrom očekivanje* μ *i nepoznatom varijancom* σ^2.
(i) Interval povjerenja (G_1, G_2) *za parametar očekivanja* μ *s pouzdanošću* γ *čine procjenitelji :*

$$G_1 = \overline{X} - \lambda\frac{\widehat{S}}{\sqrt{n}} \;\; i \;\; G_2 = \overline{X} + \lambda\frac{\widehat{S}}{\sqrt{n}},$$

gdje je $\overline{X}$ *uzoračka aritmetička sredina,* $\widehat{S}^2$ *korigirana uzoračka varijanca, a* $\lambda = z_{\frac{1+\gamma}{2}}$ *kvantil sudentove distribucije* $t(n-1)$, $F(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.
(ii) Interval povjerenja (G_1, G_2) *za parametar očekivanja* μ *s pouzdanošću* γ *čine procjenitelji:*

$$G_1 = \overline{X} - \lambda\frac{\widehat{\Sigma}}{\sqrt{n-1}}, \;\; G_2 = \overline{X} + \lambda\frac{\widehat{\Sigma}}{\sqrt{n-1}},$$

gdje je $\overline{X}$ *uzoračka aritmetička sredina,* $\widehat{\Sigma}^2$ *uzoračka varijanca, a* $\lambda = z_{\frac{1+\gamma}{2}}$ *kvantil studentove distribucije* $t(n-1)$.

Dokaz: *tko želi znati više*

(i) Neka je $X \sim N(\mu, \sigma^2)$, onda je $Y = \sqrt{n}\frac{\overline{X}-\mu}{\widehat{S}} \sim t(n-1)$.
Na simetričnom intervalu $(-\lambda, \lambda)$:
$P(-\lambda \leq \sqrt{n}\frac{\overline{X}-\mu}{\widehat{S}} \leq \lambda) = F(\lambda) - F(-\lambda) = 2F(\lambda) - 1$
tj.
$P(\overline{X} - \lambda\frac{\widehat{S}}{\sqrt{n}} < \mu < \overline{X} + \lambda\frac{\widehat{S}}{\sqrt{n}}) = 2F(\lambda) - 1.$
Ako je zadana pozdanost γ, $P(G_1 < \mu < G_2) = \gamma$, onda možemo odrediti λ tako da

vrijedi $F(\lambda) = \frac{1+\gamma}{2}$ tj. $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil studentove distribucije $t(n-1)$, $F(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.
Procjenitelji

$$G_1 = \overline{X} - \lambda \frac{\widehat{S}}{\sqrt{n}}, G_2 = \overline{X} + \lambda \frac{\widehat{S}}{\sqrt{n}}$$

čine interval povjerenja

$$(\overline{X} - \lambda \frac{\widehat{S}}{\sqrt{n}},\ \overline{X} + \lambda \frac{\widehat{S}}{\sqrt{n}})$$

za parametar očekivanja μ slučajne varijable $X \sim N(\mu, \sigma^2)$ s puzdanošću γ ako je nepoznata varijanca σ^2.
Parametar očekivanja μ s pozdanošću γ poprimit će vrijednosti u intervalu

$$(\overline{x} - \lambda \frac{\widehat{s}}{\sqrt{n}},\ \overline{x} + \lambda \frac{\widehat{s}}{\sqrt{n}}),$$

gdje je $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil studentove distribucije
$F(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.
Ako je tablica studentove distribucije $Y \sim t(n-1,)$ dana u obliku
$P(|Y| > \varepsilon) = p$, onda λ tražimo tako da je $P(|\sqrt{n}\frac{\overline{X}-\mu}{\widehat{S}}| < \lambda) = \frac{1-\gamma}{2}$.
(ii) Koristimo vezu $\widehat{S}^2 = \frac{n}{n-1}\widehat{\Sigma}^2$.

NAPOMENA 13.5 *Širina intervala povjerenja za očekivanje slučajne varijable $X \sim N(\mu, \sigma^2)$ nepoznate varijance je*

$$\delta = 2\lambda \frac{\widehat{s}}{\sqrt{n}}$$

uz zadanu pozdanost γ, gdje je $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil studentove distribucije $t(n-1)$, $F(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.

NAPOMENA 13.6 *Kvantili za studentovu distribuciju za $n = 5$, $t(4)$,*
$F(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$:

γ	0.95	0.99
$\frac{1+\gamma}{2}$	0.975	0.995
$z_{\frac{1+\gamma}{2}}$	2.78	4.60

PRIMJER 13.4 *Neka slučajna varijabla ima normalnu distribuciju $X \sim N(\mu, \sigma^2)$ nepoznate varijance σ^2. Uzet je utorak veličine $n = 5$ i dobivena je vrijednost uzoračke aritmetičke sredine $\overline{x} = 10.2$, i vrijednost korigirane uzoračke varijance $\widehat{s^2} = 0.64$. Odredite interval povjerenja za očekivanje slučajne varijable s pouzdanošću $\gamma = 0.95$.*

Rješenje: $P(\overline{X} - \lambda \frac{\widehat{S}}{\sqrt{n}} < \mu < \overline{X} + \lambda \frac{\widehat{S}}{\sqrt{n}}) = \gamma$.
Za očekivanje μ interval povjerenja pouzdanosti γ je $(\overline{X} - \lambda \frac{\widehat{S}}{\sqrt{n}}, \overline{X} + \lambda \frac{\widehat{S}}{\sqrt{n}})$, gdje je $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil studentove distribucije $t(n-1)$.
Za n=5, $\lambda = z_{\frac{1+\gamma}{2}}$ je kvantil studentove distribucije $t(4)$.
Iz tablice očitavamo za $\gamma = 0.95$, $\lambda = z_{\frac{1+\gamma}{2}} = 2.78$.
Koristeći date vrijednosti iz uzorka $\overline{x} = 10.2$, i $\widehat{s}^2 = 0.64$ dobivamo interval povjerenja za μ :

$$(\overline{x} - \lambda \frac{\widehat{s}}{\sqrt{n}}, \overline{x} + \lambda \frac{\widehat{s}}{\sqrt{n}}) = (10.2 - 2.78 \cdot \frac{0.8}{\sqrt{5}}, 10.2 + 2.78 \cdot \frac{0.8}{\sqrt{5}}) = (9.20, 11.19).$$

Prisjetimo se motivacijskog primjera:

PRIMJER 13.5 *motiv*

U četiri mjerenja Rockwellove tvrdoće jedne ploče radnici du dobili sljedeće vrijednosti: $64.9, 64.1, 63.8, 64.0$.
Odredite interval povjerenja za očekivanu vrijednost tvrdoće s pouzdanošću 99%.

Rješenje: Za zadanu pouzdanost $\gamma = 0.99$ interval povjerenja za očekivanje određujemo iz
$P(\overline{X} - \lambda \frac{\widehat{S}}{\sqrt{n}} \leq \mu \leq \overline{X} + \lambda \frac{\widehat{S}}{\sqrt{n}}) = \gamma$.
Za očekivanje μ interval povjerenja pouzdanosti γ je $(\overline{X} - \lambda \frac{\widehat{S}}{\sqrt{n}}, \overline{X} + \lambda \frac{\widehat{S}}{\sqrt{n}})$, gdje je $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil studentove distribucije $t(n-1)$.
Za $n = 4$, $\lambda = z_{\frac{1+\gamma}{2}}$ je kvantil studentove distribucije $t(3)$.
Iz tablice očitavamo za $\gamma = 0.99$, vrijednost $\lambda = z_{\frac{1+\gamma}{2}} = 5.84$.
Koristeći date vrijednosti iz uzorka $\overline{x} = 64.2$, i $\widehat{s}^2 = 0.233$ dobivamo interval povjerenja za μ :

$$(\overline{x} - \lambda \frac{\widehat{s}}{\sqrt{n}}, \overline{x} + \lambda \frac{\widehat{s}}{\sqrt{n}}) = (64.2 - 5.84 \cdot \frac{0.482}{\sqrt{4}}, 64.2 + 5.84 \cdot \frac{0.482}{\sqrt{4}}) = (64.2 - 1.4, 64.2 + 1.4).$$

S pouzdanošću 99% očekivana vrijednost tvrdoće po Rockwellu će se biti u intervalu $[62.8, 65.6]$.

13.4 INTERVAL POVJERENJA ZA VJEROJATNOST BINOMNE DISTRIBUCIJE n ($n \to \infty$)

MOTIV 13.4 *U anketi za izbore dobiveni su podatci za kandidata A:*
u uzorku od n=2500 glasača kandidat je dobio 1000 glasova.
Odredite interval povjerenja za postotak glasova koji će dobiti kandidat A na izborima s pouzdanošću 0.95.
(Pretpostavimo da je izbor binomna distribucija)

U Bernoullijevoj shemi s $X_i \sim B(1,p)$, $i = 1, ..., n$ slučajna varijabla $X = \sum_{i=1}^{n} X_i$ broj uspjeha u Bernoullijevoj shemi ima binomnu distribuciju $X \sim B(n,p)$.
Relativna frekvencija uspjeha u Bernoullijevoj shemi je slučajna varijabla $\frac{X}{n}$ koja odgovara $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ uzoračkoj aritmetičkoj sredini slučajnog uzorka $(X_1, X_2, ..X_n)$.
Prisjetimo se da je $\overline{X} = \frac{X}{n}$ je procjenitelj za vjerojatnost p u Binomnoj distibuciji.

TEOREM 13.4 *Ako je broj ponavljanja u Bernoullijevoj shemi velik ($n \to \infty$), onda interval povjerenja (G_1, G_2) za parametar p, vjerojatnost događaja A u slučajnom pokusu s pouzdanošću γ čine procjenitelji*

$$G_1 = \overline{X} - \lambda \frac{1}{\sqrt{n}} \sqrt{\overline{X}(1-\overline{X})} \;\; i \;\; G_2 = \overline{X} + \lambda \frac{1}{\sqrt{n}} \sqrt{\overline{X}(1-\overline{X})},$$

gdje je $\overline{X}$ uzoračka aritmetička sredina ($\overline{X} = \frac{X}{n}$ relativna frekvencija uspjeha u Bernoullijevoj shemi), a
(a) $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil standardne normalne distribucije $F^*(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$,
(b) $\lambda = \frac{1}{\sqrt{1-\gamma}}$.

Dokaz: *tko želi znati više*

Neka je $(X_1, X_2, ..X_n)$ slučajni uzorak iz Bernoullijeve sheme, $X_i \sim B(1,p)$.
Prisjetimo se da vrijedi $E(X_i) = p$, $Var(X_i) = p(1-p)$ i za uzoračku aritmetičku sredinu $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ vrijedi $E(\overline{X}) = p$, $Var(\overline{X}) = \frac{1}{n}p(1-p)$.

(a) Prema CGT za ($n \to \infty$) za uzoračku aritmetičku sredinu
$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ vrijedi $\overline{X} \sim N(p, \frac{1}{n}p(1-p))$.
Za simetrični interval $(-\lambda, \lambda)$ možemo približno odrediti
$P(-\lambda \leq \frac{\overline{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq \lambda) \approx 2F^*(\lambda) - 1$.
Nejednakost $-\lambda \leq \frac{\overline{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq \lambda$ ekvivalentna je nejednakosti: $(\frac{\overline{X}-p}{\sqrt{\frac{p(1-p)}{n}}})^2 \leq \lambda^2$, odnosno
$(n+\lambda^2)p^2 - (2\overline{X}n + \lambda^2)p + n\overline{X}^2 \leq 0$
Trebamo riješiti nejednakost po p.
Približna rješenja p_1, p_2 kvadratne jednadžbe (veliko n) su :

$$p_1 \approx \overline{X} - \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})},\ \ p_2 \approx \overline{X} + \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})}.$$

Kako je $(n+\lambda^2) > 0, p \in (p_1, p_2)$.
Zaključujemo da je

$$P(\overline{X} - \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})} \leq p \leq \overline{X} + \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})}) \approx 2F^*(\lambda) - 1.$$

Ako je zadana pouzdanost γ tako da je $P(G_1 < p < G_2) = \gamma$ onda možemo odrediti λ tako da vrijedi $F^*(\lambda) = \frac{1+\gamma}{2}$, tj. $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil standardne normalne distibucije $F^*(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.
Procjenitelji $G_1 = \overline{X} - \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})}$, $\ G_2 = \overline{X} + \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})}$ čine interval povjerenja (G_1, G_2) za paranetar vjerojatnost p uspjeha u Bernoullijevoj shemi.
(b) Prema Čebiševljevoj nejednakosti za slučajnu varijablu $\overline{X}$ koja ima
$E(\overline{X}) = p,\ Var(\overline{X}) = \frac{1}{n}p(1-p)$ u obliku
$P(p - \lambda\sqrt{\frac{p(1-p)}{n}} \leq \overline{X} \leq p + \lambda\sqrt{\frac{p(1-p)}{n}}) \geq 1 - \frac{1}{\lambda^2}$.
Nejednakost $p - \lambda\sqrt{\frac{p(1-p)}{n}} \leq \overline{X} \leq p + \lambda\sqrt{\frac{p(1-p)}{n}}$ ekvivalentna je nejednakosti (vidi pod (a)): $\overline{X} - \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})} \leq p \leq \overline{X} + \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})}$.
Zaključujemo da je

$$P(\overline{X} - \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})} \leq p \leq \overline{X} + \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})}) \geq 1 - \frac{1}{\lambda^2}.$$

Ako je zadana pouzdanost γ tako da je $P(G_1 < p < G_2) = \gamma$ onda možemo odrediti λ tako da vrijedi $1 - \frac{1}{\lambda^2} = \gamma$ tj. $\lambda = \frac{1}{\sqrt{1-\gamma}}$.

Procjenitelji $G_1 = \overline{X} - \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})}$, $G_2 = \overline{X} + \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})}$ čine interval povjerenja (G_1, G_2) za paranetar vjrojatnosti p uspjeha u Bernoullijevoj shemi.

NAPOMENA 13.7 *Možemo izvesti interval povjerenja i koristeći teorem Moivre-Laplacea (CGT) za relativnu frekvenciju u Bernullijevoj shemi*
$P(|\frac{X}{n} - p| < \varepsilon) \approx F^*(\varepsilon\sqrt{\frac{n}{p(p-1)}}) - 1,\ n \to \infty,$
tj.
$P(\frac{X}{n} - \varepsilon < p < \frac{X}{n} + \varepsilon) \approx F^*(\varepsilon\sqrt{\frac{n}{p(p-1)}}) - 1.$
Ako je zadana pozdanost γ, $P(G_1 < p < G_2) = \gamma$, *onda možemo odrediti* ε *tako da vrijedi* $F^*(\varepsilon\sqrt{\frac{n}{p(p-1)}}) = \frac{1+\gamma}{2}$, *tj.* $\varepsilon = z_{\frac{1+\gamma}{2}}\sqrt{\frac{p(1-p)}{n}}$, *odnosno* $\varepsilon = z_{\frac{1+\gamma}{2}}\frac{1}{2\sqrt{n}}$, *uz* $p(1-p) \leq \frac{1}{4}$.
Procjenitelji $G_1 = \frac{X}{n} - \varepsilon$, $G_2 = \frac{X}{n} - \varepsilon$ *čine interval povjerenja* (G_1, G_2) *za paranetar vjerojatnosti p uspjeha u Bernoullijevoj shemi. varijable*
$X \sim B(n, p)$ *s pouzdanošću* γ *ako je poznat parametar n broj ponavljanja pokusa.*
Parametar vjerojatnost p s pozdanošću γ *poprimit će vrijednosti u u intervalu*

$$(\frac{x}{n} - z_{\frac{1+\gamma}{2}}\frac{1}{2\sqrt{n}}, \frac{x}{n} + z_{\frac{1+\gamma}{2}}\frac{1}{2\sqrt{n}}),$$

gdje je $z_{\frac{1+\gamma}{2}}$ *kvantil standardne normalne distribucije* $F^*(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.

PRIMJER 13.6 *Odredite interval povjerenja za vjerojatnost p u Bernoullijevoj shemi s pouzdanošću* $\gamma = 0.95$ *ako se pokus ponovi n=100, a broj uspjeha je 32.*

Rješenje: Za slučajnu varijablu $X \sim B(n, p)$ vrijedi

$$P(\overline{X} - \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})} \leq p \leq \overline{X} + \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})}) = \gamma$$

gdje je $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil standardne normalne distibucije $F^*(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$, $\overline{X} = \frac{X}{n}$ relativna frekvencija uspjeha.
Procjenitelji $G_1 = \overline{X} - \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})}$, $G_2 = \overline{X} + \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})}$ čine interval povjerenja (G_1, G_2) za paranetar vjerojatnosti p.
Parametar vjerojatnost p s pouzdanošću γ poprimit će vrijednosti u intervalu $(\overline{x} - \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{x}(1-\overline{x})}, \overline{x} + \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{x}(1-\overline{x})})$ gdje je $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil standardne normalne distibucije.
Iz uzorka je uzeta relativna fekvencija $\overline{x} = \frac{x}{n} = \frac{32}{100}$, za pouzdanost $\gamma = 0.95$ iz

tablice očitamo $\lambda = z_{\frac{1+\gamma}{2}} = 1.96$ i odredimo interval

$$\begin{aligned}
&(\overline{x} - \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{x}(1-\overline{x})}, \overline{x} + \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{x}(1-\overline{x})}) \\
= \;& (0.32 - 1.96\frac{1}{\sqrt{100}}\sqrt{0.32(1-0.32)}, 0.32 + 1.96\frac{1}{\sqrt{100}}\sqrt{0.32(1-0.32)}) \\
= \;& (0.23, 0.41).
\end{aligned}$$

Parametar vjerojatnosti p s pouzdanošću $\gamma = 0.95$ poprimit će vrijednosti u intervalu $(0.23, 0.41)$.

PRIMJER 13.7 *Pravljena je anketa o dolasku na predavanja VIS. Na uzorku 163 studenta njih 62 je odgovorilo da je dolazilo na predavanja. Odredite interval povjerenja za vjerojatnost dolaska studenata prve godine na predavanja VIS s pouzdanošću 0.95. (Pretpostavimo da je izbor binomna distribucija)*

Rješenje: Parametar vjerojatnosti p s pozdanošću γ poprimit će vrijednosti u u intervalu $(\overline{x}-\lambda\frac{1}{\sqrt{n}}\sqrt{\overline{x}(1-\overline{x})}, \overline{x}+\lambda\frac{1}{\sqrt{n}}\sqrt{\overline{x}(1-\overline{x})})$ gdje je $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil standardne normalne distibucije.
Iz uzorka je uzeta relativna fekvencija $\overline{x} = \frac{x}{n} = \frac{62}{163} = 0.38$, za pouzdanost $\gamma = 0.95$ iz tablice očitamo $\lambda = z_{\frac{1+\gamma}{2}} = 1.96$ i odredimo interval

$$\begin{aligned}
&(\overline{x} - \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{x}(1-\overline{x})}, \overline{x} + \lambda\frac{1}{\sqrt{n}}\sqrt{\overline{x}(1-\overline{x})}) \\
= \;& (0.38 - 1.96\frac{1}{\sqrt{163}}\sqrt{0.38(1-0.38)}, 0.38 + 1.96\frac{1}{\sqrt{163}}\sqrt{0.38(1-0.38)}) \\
= \;& (0.30, 0.45).
\end{aligned}$$

Parametar vjerojatnosti p s pouzdanošću $\gamma = 0.95$ poprimit će vrijednosti u intervalu $(0.30, 0.45)$.
Vjerojatnost dolaska na predavanja VIS s pouzdanošću 0.95 je između 0.30 i 0.45.

Prisjetimo se motivacijskog primjera:

PRIMJER 13.8 *motiv*

U anketi za izbore dobiveni su podatci za kandidata A:
u uzorku od n=2500 glasača kandidat je dobio 1000 glasova.
Odredite interval povjerenja za postotak glasova koji će dobiti kandidat A na izborima s pouzdanošću 0.95.
(Pretpostavimo da je izbor binomna distribucija)

Rjšenje: Parametar vjerojatnost p s pozdanošću γ poprimit će vrijednosti u u intervalu $(\overline{x}-\lambda\frac{1}{\sqrt{n}}\sqrt{\overline{x}(1-\overline{x})}, \overline{x}+\lambda\frac{1}{\sqrt{n}}\sqrt{\overline{x}(1-\overline{x})})$ gdje je $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil standardne normalne distibucije.
Iz uzorka imamo relativnu fekvenciju $\overline{x} = \frac{x}{n} = \frac{1000}{2500} = 0.4$, za pouzdanost $\gamma = 0.95$ iz tablice očitamo $\lambda = z_{\frac{1+\gamma}{2}} = 1.96$ i odredimo interval

$$\begin{aligned}
&(\overline{x}-\lambda\frac{1}{\sqrt{n}}\sqrt{\overline{x}(1-\overline{x})}, \overline{x}+\lambda\frac{1}{\sqrt{n}}\sqrt{\overline{x}(1-\overline{x})}) \\
=\ &(0.4-1.96\frac{1}{\sqrt{2500}}\sqrt{0.4(1-0.4)}, 0.4+1.96\frac{1}{\sqrt{2500}}\sqrt{0.4(1-0.4)}) \\
=\ &(0.38, 0.42).
\end{aligned}$$

Parametar vjerojatnosti p s pozdanošću $\gamma = 0.95$ poprimit će vrijednosti u intervalu $(0.38, 0.42)$.
Postotak glasova koje će na izborima dobiti kandidat A s pouzdanošću 0.95 je između 38% i 42%.

PRIMJER 13.9 *Ako želimo odrediti postotak p% glasova koje će dobiti kandidat A na izborima pravimo anketu. Koliki uzorak treba uzeti da bi se za p odredio interval pouzdanosti 0.95 širine 0.04?*

Rješenje: Parametar vjerojatnost p s pozdanošću γ poprimit će vrijednosti u u intervalu $(\overline{x}-\lambda\frac{1}{\sqrt{n}}\sqrt{\overline{x}(1-\overline{x})}, \overline{x}+\lambda\frac{1}{\sqrt{n}}\sqrt{\overline{x}(1-\overline{x})})$ gdje je $\lambda = z_{\frac{1+\gamma}{2}}$ kvantil standardne normalne distibucije.
Širina intervala $\delta = 2\lambda\frac{1}{\sqrt{n}}\sqrt{\overline{x}(1-\overline{x})}$.

Kako je $\overline{x}(1-\overline{x}) \leq \frac{1}{4}$ možemo ocijeniti veličinu uzorka n: $n \leq \frac{(z_{\frac{1+\gamma}{2}})^2}{\delta^2}$.
Za zadane $\delta = 0.04$, $\gamma = 0.95$, dobivamo $\lambda = z_{\frac{1+\gamma}{2}} = 1.96$ i

$$n \leq \frac{(z_{\frac{1+\gamma}{2}})^2}{\delta^2} = \frac{1.96^2}{0.04^2} = 2401.$$

Uzorak mora imati bar 2401 glasača da bi s pouzdanošću 0.95 interval povjerenja za postotak bodova na izborima za kandidata A bio širok 0.04. (greška unutar 4%).

13.5 Ponovimo

INTERVAL POVJERENJA ZA parametar t

slučajni uzorak	$(X_1, X_2, ..., X_n)$
parametar	t
pouzdanost	γ
interval povjerenja za t	$P(G_1 \leq t \leq G_2) \geq \gamma$

INTERVAL POVJERENJA ZA OČEKIVANJE kad je $n \to \infty$

slučajni uzorak	$(X_1, X_2, ..., X_n), \;\; n \geq 30$
parametar	μ
pouzdanost	γ
interval povjerenja za t	$P(G_1 \leq \mu \leq G_2) \geq \gamma$
	$G_1 = \overline{X} - \lambda \frac{\sigma}{\sqrt{n}}$
	$G_2 = \overline{X} + \lambda \frac{\sigma}{\sqrt{n}}$
λ	$z_{\frac{1+\gamma}{2}}$, kvantil $N(0, 1)$

INTERVAL POVJERENJA ZA OČEKIVANJE NORMALNE distribucije
(varijanca poznata)

slučajni uzorak	$(X_1, X_2, ..., X_n), \;\;$ iz $N(\mu, \sigma^2)$
parametar	μ
pouzdanost	γ
interval povjerenja za μ	$P(G_1 \leq \mu \leq G_2) \geq \gamma$
	$G_1 = \overline{X} - \lambda \frac{\sigma}{\sqrt{n}}$
	$G_2 = \overline{X} + \lambda \frac{\sigma}{\sqrt{n}}$
λ	$z_{\frac{1+\gamma}{2}}$, kvantil $N(0, 1)$

INTERVAL POVJERENJA ZA OČEKIVANJE NORMALNE distribucije
(varijanca nepoznata)

slučajni uzorak	$(X_1, X_2, ..., X_n)$, iz $N(\mu, \sigma^2)$
parametar	μ
pouzdanost	γ
interval povjerenja za μ	$P(G_1 \leq \mu \leq G_2) \geq \gamma$
	$G_1 = \overline{X} - \lambda \frac{\widehat{s}}{\sqrt{n}}$
	$G_2 = \overline{X} + \lambda \frac{\widehat{s}}{\sqrt{n}}$
λ	$z_{\frac{1+\gamma}{2}}$, kvantil $t(n-1)$

INTERVAL POVJERENJA ZA VJEROJATNOST BINOMNE distribucije (veliki n)

slučajni uzorak	$(X_1, X_2, ..., X_n)$, iz $B(n, p)$
parametar	p
pouzdanost	γ
interval povjerenja za μ	$P(G_1 \leq p \leq G_2) \geq \gamma$
	$G_1 = \frac{X}{n} - \lambda \frac{1}{\sqrt{n}} \sqrt{\frac{X}{n}(1 - \frac{X}{n})}$
	$G_2 = \frac{X}{n} + \lambda \frac{1}{\sqrt{n}} \sqrt{\frac{X}{n}(1 - \frac{X}{n})}$
λ	$z_{\frac{1+\gamma}{2}}$, kvantil $N(0, 1)$

Poglavlje 14

INTERVAL POVJERENJA ZA VARIJANCU

14.1 INTERVAL POVJERENJA ZA VARIJANCU NORMALNU distribucije poznatog očekivanja

MOTIV 14.1 *Neka slučajna varijabla ima normalnu distribuciju $X \sim N(\mu, \sigma^2)$ nepoznate varijance σ^2 i poznatog očekivanja $\mu = 10$. Uzet je uzorak veličine $n = 5$ i dobivena je vrijednost uzorka $(7, 8, 10, 9, 9)$. Odredite interval povjerenja za varijancu σ^2 slučajne varijable s pouzdanošću $\gamma = 0.95$.*

TEOREM 14.1 *Neka je $(X_1, X_2, ..X_n)$ slučajni uzorak slučajne varijable $X \sim N(\mu, \sigma^2)$ s nepoznatim parametrom varijancom σ^2, i poznatog očekivanja μ. Interval povjerenja (G_1, G_2) za parametar varijance σ^2 s pouzdanošću γ čine procjenitelji:*

$$G_1 = \frac{1}{\lambda_2} \sum_{i=1}^{n} (X_i - \mu)^2 \;\; i \;\; G_2 = \frac{1}{\lambda_1} \sum_{i=1}^{n} (X_i - \mu)^2,$$

gdje su $\lambda_1 = z_{\frac{1-\gamma}{2}}$, $\lambda_2 = z_{\frac{1+\gamma}{2}}$, kvantili hi-kvadrat distribucije $\chi^2(n)$, $F(z_{\frac{1-\gamma}{2}}) = \frac{1-\gamma}{2}$, $F(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.

Dokaz: *tko želi znati više*

Neka je $X \sim N(\mu, \sigma^2)$, onda je $Y = \sum_{i=1}^{n} (\frac{X_i - \mu}{\sigma})^2 \sim \chi^2(n)$.
Na intervalu (λ_1, λ_2) :
$P(\lambda_1 < \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2 < \lambda_2) = F(\lambda_2) - F(\lambda_1)$, tj.

$$P(\frac{1}{\lambda_2} \sum_{i=1}^{n} (X_i - \mu)^2 < \sigma^2 < \frac{1}{\lambda_1} \sum_{i=1}^{n} (X_i - \mu)^2) = F(\lambda_2) - F(\lambda_1).$$

Ako je zadana pozdanost γ, $P(G_1 < \sigma^2 < G_2) = \gamma = \frac{1+\gamma}{2} - \frac{1-\gamma}{2}$, onda možemo odrediti λ_1, λ_2 tako da vrijedi $F(\lambda_1) = \frac{1-\gamma}{2}$, $F(\lambda_2) = \frac{1+\gamma}{2}$ tj. $\lambda_1 = z_{\frac{1-\gamma}{2}}$, $\lambda_2 = z_{\frac{1+\gamma}{2}}$ su kvantili hikvadrat distribucije $\chi^2(n)$.
Procjenitelji $G_1 = \frac{1}{\lambda_2} \sum_{i=1}^{n} (X_i - \mu)^2$, $G_2 = \frac{1}{\lambda_1} \sum_{i=1}^{n} (X_i - \mu)^2$ čine interval povjerenja

$$(\frac{1}{\lambda_2} \sum_{i=1}^{n} (X_i - \mu)^2, \frac{1}{\lambda_1} \sum_{i=1}^{n} (X_i - \mu)^2)$$

za parametar σ^2 slučajne varijable $X \sim N(\mu, \sigma^2)$ s pouzdanošću γ.
Parametar varijanca σ^2 s pouzdanošću γ poprimit će vrijednosti u intervalu

$$(\frac{1}{\lambda_2} \sum_{i=1}^{n} (x_i - \mu)^2, \frac{1}{\lambda_1} \sum_{i=1}^{n} (x_i - \mu)^2),$$

gdje su $\lambda_1 = z_{\frac{1-\gamma}{2}}$, $\lambda_2 = z_{\frac{1+\gamma}{2}}$ kvantili hikvadrat distribucije $\chi^2(n)$.
Ako je tablica hikvadrat distribucije $Y \sim \chi^2(n)$, dana u obliku
$P(Y > \varepsilon) = p$, onda $P(Y > \lambda_2) = \frac{1-\gamma}{2}$, $P(Y > \lambda_1) = \frac{1+\gamma}{2}$.

NAPOMENA 14.1 *Ova procjena parametra varijance slučajne varijable $X \sim N(\mu, \sigma^2)$ poznatog očekivanja može koristiti za određivanje širine inervala*

$$\delta = \sum_{i=1}^{n} (x_i - \mu)^2 (\frac{1}{\lambda_1} - \frac{1}{\lambda_2})$$

uz zadanu pozdanost γ za interval povjerenja

$$(\frac{1}{\lambda_2} \sum_{i=1}^{n} (x_i - \mu)^2, \frac{1}{\lambda_1} \sum_{i=1}^{n} (x_i - \mu)^2),$$

gdje su $\lambda_1 = z_{\frac{1-\gamma}{2}}$, $\lambda_2 = z_{\frac{1+\gamma}{2}}$ kvantili hikvadrat distribucije $\chi^2(n-1)$.

NAPOMENA 14.2 *Kvantili za hikvadrat distribuciju za $n = 5$, $\chi^2(5)$,
$F(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$:*

γ	0.95	0.99
$\frac{1+\gamma}{2}$	0.975	0.995
$z_{\frac{1+\gamma}{2}}$	12.83	16.75

γ	0.95	0.99
$\frac{1-\gamma}{2}$	0.025	0.01
$z_{\frac{1-\gamma}{2}}$	0.83	0.55

Prisjetimo se motivacijskog primjera:

PRIMJER 14.1 *motiv*

Neka slučajna varijabla ima normalnu distribuciju $X \sim N(\mu, \sigma^2)$ nepoznate varijance σ^2 i poznatog očekivanja $\mu = 10$. Uzet je uzorak veličine $n = 5$ i dobivena je vrijednost uzorka $(7, 8, 10, 9, 9)$. Odredite interval povjerenja za varijancu σ^2 slučajne varijable s pouzdanošću $\gamma = 0.95$.

Rješenje: $P(\frac{1}{\lambda_2}\sum_{i=1}^n (x_i - \mu)^2 < \sigma^2 < \frac{1}{\lambda_1}\sum_{i=1}^n (x_i - \mu)^2) = \gamma$.
Za varijancu σ^2, poznatog očekivvanja interval povjerenja pouzdanosti γ je $(\frac{1}{\lambda_2}\sum_{i=1}^n (x_i - \mu)^2, \frac{1}{\lambda_1}\sum_{i=1}^n (x_i - \mu)^2)$, gdje su $\lambda_1 = z_{\frac{1-\gamma}{2}}$, $\lambda_2 = z_{\frac{1+\gamma}{2}}$ kvantili hikvadrat distribucije $\chi^2(n)$.
Za $n = 5$, $\lambda_1 = z_{\frac{1-\gamma}{2}}$, $\lambda_2 = z_{\frac{1+\gamma}{2}}$.
Iz tablice očitavamo za $\gamma = 0.95$, $\lambda_1 = z_{\frac{1-\gamma}{2}} = 0.83$, $\lambda_2 = z_{\frac{1+\gamma}{2}} = 12.83$.
Koristeći date vrijednosti iz uzorka $\sum_{i=1}^{5}(x_i - 10)^2 = 3^2 + 2^2 + 1^2 + 1^2 = 15$
dobivamo interval povjerenja za σ^2 :

$$(\frac{1}{\lambda_2}\sum_{i=1}^{n}(x_i - \mu)^2, \frac{1}{\lambda_1}\sum_{i=1}^{n}(x_i - \mu)^2) = (\frac{1}{12.83} \cdot 15, \frac{1}{0.83} \cdot 15) = (1.17, 18.07).$$

14.2 INTERVAL POVJERENJA ZA VARIJANCU NORMALNE distribucije nepoznatog očekivanja

MOTIV 14.2 *U četiri mjerenja Rockwellove tvrdoće jedne ploče radnici du dobili sljedeće vrijednosti:*
$64.9, 64.1, 63.8, 64.0$.
Odredite interval povjerenja za varijancu tvrdoće s pouzdanošću 99%.

TEOREM 14.2 *Neka je* $(X_1, X_2, ..X_n)$ *slučajni uzorak slučajne varijable* $X \sim N(\mu, \sigma^2)$ *s nepoznatim parametrom varijancom* σ^2*, nepoznatog očekivanja* μ*. Interval povjerenja* (G_1, G_2) *za varijance* σ^2 *s pouzdanošću* γ *čine procjenitelji:*

$$G_1 = \frac{(n-1)}{\lambda_2}\widehat{S^2} \;\; i \;\; G_2 = \frac{(n-1)}{\lambda_1}\widehat{S^2},$$

gdje je $\widehat{S^2}$ *korigirana uzoračka varijanca, a* $\lambda_1 = z_{\frac{1-\gamma}{2}}$*,* $\lambda_2 = z_{\frac{1+\gamma}{2}}$*, kvantili hikvadrat distribucije* $\chi^2(n-1)$*,*
$F(z_{\frac{1-\gamma}{2}}) = \frac{1-\gamma}{2}$*,* $F(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$.

Dokaz: *tko želi znati še*

Neka je $X \sim N(\mu, \sigma^2)$, onda je $Y = \frac{n-1}{\sigma^2}\widehat{S^2} \sim \chi^2(n-1)$.
Na intervalu $(\lambda_1, \lambda_2) : P(\lambda_1 < \frac{n-1}{\sigma^2}\widehat{S^2} < \lambda_2) = F(\lambda_2) - F(\lambda_1)$ tj.

$$P(\frac{(n-1)}{\lambda_2}\widehat{S^2} < \sigma^2 < \frac{(n-1)}{\lambda_1}\widehat{S^2}) = F(\lambda_2) - F(\lambda_1).$$

Ako je zadana pozdanost γ, $P(G_1 < \sigma^2 < G_2) = \gamma = \frac{1+\gamma}{2} - \frac{1-\gamma}{2}$, i onda možemo odrediti λ_1, λ_2 tako da vrijedi $F(\lambda_1) = \frac{1-\gamma}{2}, F(\lambda_2) = \frac{1+\gamma}{2}$ tj. $\lambda_1 = z_{\frac{1-\gamma}{2}}$, $\lambda_2 = z_{\frac{1+\gamma}{2}}$ kvantili hikvadrat distribucije $\chi^2(n-1)$.
Procjenitelji $G_1 = \frac{(n-1)}{\lambda_2}\widehat{S^2}$, $G_2 = \frac{(n-1)}{\lambda_1}\widehat{S^2}$ čine interval povjerenja

$$(\frac{(n-1)}{\lambda_2}\widehat{S^2}, \frac{(n-1)}{\lambda_1}\widehat{S^2})$$

za parametar σ^2 slučajne varijable $X \sim N(\mu, \sigma^2)$ s pouzdanošću γ.
Parametar varijanca σ^2 s pouzdanošću γ poprimit će vrijednosti u u intervalu

$(\frac{(n-1)}{\lambda_2}\widehat{s^2}, \frac{(n-1)}{\lambda_1}\widehat{s^2})$, gdje su $\lambda_1 = z_{\frac{1-\gamma}{2}}$, $\lambda_2 = z_{\frac{1+\gamma}{2}}$ kvantili hikvadrat distribucije $\chi^2(n-1)$.
Ako je tablica hikvadrat distribucije $Y \sim \chi^2(n-1)$, dana u obliku
$P(Y > \varepsilon) = p$, onda $P(Y > \lambda_2) = \frac{1-\gamma}{2}$, $P(Y > \lambda_1) = \frac{1+\gamma}{2}$.

NAPOMENA 14.3 *Ova procjena parametra varijance slučajne varijable $X \sim N(\mu, \sigma^2)$ nepoznatog očekivanja može koristiti za određivanje*

$$\delta = (n-1)\widehat{s^2}(\frac{1}{\lambda_1} - \frac{1}{\lambda_2})$$

širine inervala uz zadanu pozdanost γ za interval povjerenja

$$(\frac{(n-1)}{\lambda_2}\widehat{S^2}, \frac{(n-1)}{\lambda_1}\widehat{S^2}),$$

gdje su $\lambda_1 = z_{\frac{1-\gamma}{2}}$, $\lambda_2 = z_{\frac{1+\gamma}{2}}$ kvantili hikvadrat distribucije $\chi^2(n-1)$.

NAPOMENA 14.4 *Kvantili za hikvadrat distribuciju za $n = 4$, $\chi^2(4)$, $F(z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$:*

γ	0.95	0.99
$\frac{1+\gamma}{2}$	0.975	0.995
$z_{\frac{1+\gamma}{2}}$	11.14	14.86

γ	0.95	0.99
$\frac{1-\gamma}{2}$	0.025	0.01
$z_{\frac{1-\gamma}{2}}$	0.48	0.21

PRIMJER 14.2 *Neka slučajna varijabla ima normalnu distribuciju $X \sim N(\mu, \sigma^2)$ nepoznate varijance σ^2 i nepoznatog očekivanja μ. Uzet je uzorak veličine $n = 5$ i dobivena je vrijednost uzorka $(7, 8, 10, 9, 9)$. Odredite interval povjerenja za varijancu σ^2 slučajne varijable s pouzdanošću $\gamma = 0.95$.*

Rješenje: $P(\frac{(n-1)}{\lambda_2}\widehat{S^2} < \sigma^2 < \frac{(n-1)}{\lambda_1}\widehat{S^2}) = \gamma$.
Za varijancu σ^2 interval povjerenja pouzdanosti γ je $(\frac{(n-1)}{\lambda_2}\widehat{S^2}, \frac{(n-1)}{\lambda_1}\widehat{S^2})$, gdje su $\lambda_1 = z_{\frac{1-\gamma}{2}}$, $\lambda_2 = z_{\frac{1+\gamma}{2}}$ kvantili hikvadrat distribucije $\chi^2(n-1)$.
Za $n = 5$, iz tablice očitavamo za $\gamma = 0.95$, $\lambda_1 = z_{\frac{1-\gamma}{2}} = 0.48$,
$\lambda_2 = z_{\frac{1+\gamma}{2}} = 11.14$.

Koristeći date vrijednosti iz uzorka: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{5}(7+8+10+9+9) = 8.6$,
$\sum_{i=1}^{5}(x_i-\overline{x})^2 = 3^2+2^2+1^2+1^2 = 15$

$$\begin{aligned}\widehat{s^2} &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 = \frac{1}{5-1}\sum_{i=1}^{5}(x_i-8.6)^2 \\ &= \frac{1}{4}((7-8.6)^2+(8-8.6)^2+(10-8.6)^2+(9-8.6)^2+(9-8.6)^2) = 1.3\end{aligned}$$

dobivamo interval povjerenja za σ^2 :

$$\left(\frac{(n-1)}{\lambda_2}\widehat{s^2}, \frac{(n-1)}{\lambda_1}\widehat{s^2}\right) = \left(\frac{(5-1)}{11.14}\cdot 1.3, \frac{(5-1)}{0.48}\cdot 1.3\right) = (0.47, 10.83).$$

PRIMJER 14.3 *Neka slučajna varijabla ima normalnu distribuciju $X \sim N(\mu, \sigma^2)$ nepoznate varijance σ^2 i nepoznatog očekivanja μ. Uzet je uzorak veličine $n = 5$ i dobivena je vrijednost korigirane uzoračke varijance $\widehat{s^2} = 0.64$. Odredite interval povjerenja za varijancu σ^2 slučajne varijable s pouzdanošću $\gamma = 0.95$.*

Rješenje: $P(\frac{(n-1)}{\lambda_2}\widehat{S^2} < \sigma^2 < \frac{(n-1)}{\lambda_1}\widehat{S^2}) = \gamma$.
Za varijancu σ^2 interval povjerenja pouzdanosti γ je $(\frac{(n-1)}{\lambda_2}\widehat{S^2}, \frac{(n-1)}{\lambda_1}\widehat{S^2})$, gdje su $\lambda_1 = z_{\frac{1-\gamma}{2}}, \lambda_2 = z_{\frac{1+\gamma}{2}}$ kvantili hikvadrat distribucije $\chi^2(n-1)$.
Za $n = 5$, iz tablice očitavamo za $\gamma = 0.95$, $\lambda_1 = z_{\frac{1-\gamma}{2}} = 0.48$,
$\lambda_2 = z_{\frac{1+\gamma}{2}} = 11.14$.
Koristeći date vrijednosti iz uzorka $\widehat{s^2} = 0.64$ dobivamo interval povjerenja za σ^2 :

$$\left(\frac{(n-1)}{\lambda_2}\widehat{s^2}, \frac{(n-1)}{\lambda_1}\widehat{s^2}\right) = \left(\frac{(5-1)}{11.14}\cdot 0.64, \frac{(5-1)}{0.48}\cdot 0.64\right) = (0.229, 5.33).$$

Prisjetimo se motivacijskog primjera:

PRIMJER 14.4 *motiv*

U četiri mjerenja Rockwellove tvrdoće jedne ploče radnici du dobili sljedeće vrijednosti: 64.9, 64.1, 63.8, 64.0.
Odredite interval povjerenja za varijancu tvrdoće s pouzdanošću 99%.

Rješenje:

Za varijancu σ^2 interval povjerenja pouzdanosti γ je $(\frac{(n-1)}{\lambda_2}\widehat{S^2}, \frac{(n-1)}{\lambda_1}\widehat{S^2})$, gdje su $\lambda_1 = z_{\frac{1-\gamma}{2}}$, $\lambda_2 = z_{\frac{1+\gamma}{2}}$ kvantili hikvadrat distribucije $\chi^2(n-1)$.
Za $n = 4$, i $\gamma = 0.99$, iz tablice za $\chi^2(3)$ očitavamo za $\lambda_1 = z_{\frac{1-\gamma}{2}} = 0.07$
$\lambda_2 = z_{\frac{1+\gamma}{2}} = 12.84$.
Koristeći date vrijednosti iz uzorka $\widehat{s^2} = 0.233$ dobivamo interval povjerenja za σ^2 :

$$(\frac{(n-1)}{\lambda_2}\widehat{s^2}, \frac{(n-1)}{\lambda_1}\widehat{s^2}) = (\frac{3}{12.84} \cdot 0.233, \frac{(3)}{0.07} \cdot 0.233) = (0.054, 9.98).$$

Interval povjerenja za varijancu uz pouzdanost 99% je $[0.05, 9.98]$.

NAPOMENA 14.5 *Intervali povjerenja za varijancu i standardnu devijaciju ako je n veliki.*
(a) za $n > 30$, $P(\frac{2n\widehat{\sigma}^2}{(\sqrt{2n-3}+\lambda)^2} < \sigma^2 < \frac{2n\widehat{\sigma}^2}{(\sqrt{2n-3}-\lambda)^2}) = \gamma$,
$\lambda = z_{\frac{1+\gamma}{2}}$ *kvantili standardne normalne distribucije;*
(b) za $n \geq 100$, $P(\widehat{\sigma} - \lambda\frac{\widehat{\sigma}}{\sqrt{2n}} < \sigma < \widehat{\sigma} + \lambda\frac{\widehat{\sigma}}{\sqrt{2n}}) = \gamma$,
$\lambda = z_{\frac{1+\gamma}{2}}$ *kvantili standardne normalne distribucije.*

PRIMJER 14.5 *Uzet je uzorak veličine 100 za visinu 18-godišnjakinja i dobivene su vrijednosti za uzoračku aritmetičku sredinu* $\overline{x} = 165$ *cm i uzoračku varijancu* $\widehat{\sigma^2} = 5.8^2$ *cm.*
Pretpostavimo da visina djevojaka ima normalnu distribuciju.
(a) Odredite interval povjerenja za očekivnje μ *(srednju visinu) s pouzdanošću* $\gamma = 0.95$.
(b) Odredite interval povjerenja za standardnu devijaciju σ *s pouzdanošću* $\gamma = 0.95$.
(c) Koliki treba biti uzorak da s pouzdanošću $\gamma = 0.95$ *standardna devijacija ne odstupa od uzoračke standardne devijacije više od 5%?*

Rješenje:

(a) za $n > 30$, $P(\overline{X} - \lambda\frac{\widehat{\sigma}}{\sqrt{n}} < \mu < \overline{X} + \lambda\frac{\widehat{\sigma}}{\sqrt{n}}) = \gamma$, $\lambda = z_{\frac{1+\gamma}{2}}$ kvantili standardne normalne distribucije

$$\begin{aligned}
&(\overline{x} - \lambda\frac{\widehat{\sigma}}{\sqrt{n}} < \mu < \overline{x} + \lambda\frac{\widehat{\sigma}}{\sqrt{n}}) \\
&= \quad (165 - 1.96\frac{5.8}{\sqrt{100}} < \mu < 165. + 1.96\frac{5.8}{\sqrt{100}}) = (163.8, 166.2).
\end{aligned}$$

(b) za $n \geq 100$, $P(\widehat{\sigma} - \lambda \frac{\widehat{\sigma}}{\sqrt{2n}} < \sigma < \widehat{\sigma} + \lambda \frac{\widehat{\sigma}}{\sqrt{2n}}) = \gamma$, $\lambda = z_{\frac{1+\gamma}{2}}$ kvantili standardne normalne distribucije

$$
\begin{aligned}
&(\widehat{\sigma} - \lambda \frac{\widehat{\sigma}}{\sqrt{2n}} < \mu < \widehat{\sigma} + \lambda \frac{\widehat{\sigma}}{\sqrt{2n}}) \\
&\quad = \quad (5.8 - 1.96 \frac{5.8}{\sqrt{200}} < \mu < 5.8 + 1.96 \frac{5.8}{\sqrt{200}}) = (4.9, 6.6).
\end{aligned}
$$

(c) $P(\widehat{\sigma} - \lambda \frac{\widehat{\sigma}}{\sqrt{2n}} < \sigma < \widehat{\sigma} + \lambda \frac{\widehat{\sigma}}{\sqrt{2n}}) = \gamma$, $\lambda = z_{\frac{1+\gamma}{2}}$ kvantili standardne normalne distribucije.

Iz $P(5.8 - 1.96 \frac{5.8}{\sqrt{2n}} < \sigma < 5.8 + 1.96 \frac{5.8}{\sqrt{2n}}) = 0.95$ i uvjeta $1.96 \frac{5.8}{\sqrt{2n}} < 5.8 \cdot 5\%$ zaključujemo da je za $n \geq 769$.

γ	0.90	0.95	0.99
$\frac{1+\gamma}{2}$	0.95	0.975	0.995
$z_{\frac{1+\gamma}{2}}$	1.65	1.96	2.58

14.3 Ponovimo

INTERVAL POVJERENJA ZA VARIJANCU NORMALNE distribucije
(očekivanje poznato)

slučajni uzorak	$(X_1, X_2, ..., X_n)$, iz $N(\mu, \sigma^2)$
parametar	σ^2
pouzdanost	γ
interval povjerenja za σ^2	$P(G_1 \leq \sigma^2 \leq G_2) \geq \gamma$
	$G_1 = \frac{1}{\lambda_2} \sum_{i=1}^{n} (X_i - \mu)^2$
	$G_1 = \frac{1}{\lambda_1} \sum_{i=1}^{n} (X_i - \mu)^2$
λ_1	$z_{\frac{1-\gamma}{2}}$, kvantil $\chi^2(n)$
λ_2	$z_{\frac{1+\gamma}{2}}$, kvantil $\chi^2(n)$

INTERVAL POVJERENJA ZA VARIJANCU NORMALNE distribucije
(očekivanje nepoznato)

slučajni uzorak	$(X_1, X_2, ..., X_n)$, iz $N(\mu, \sigma^2)$
parametar	σ^2
pouzdanost	γ
interval povjerenja za σ^2	$P(G_1 \leq \sigma^2 \leq G_2) \geq \gamma$
	$G_1 = \frac{n-1}{\lambda_2} \cdot \widehat{S}^2$
	$G_2 = \frac{n-1}{\lambda_1} \cdot \widehat{S}^2$
λ_1	$z_{\frac{1-\gamma}{2}}$, kvantil $\chi^2(n-1)$
λ_2	$z_{\frac{1+\gamma}{2}}$, kvantil $\chi^2(n-1)$

Poglavlje 15

TESTIRANJE HIPOTEZA

Teorija testiranja statističkih hipoteza sastoji se u određivanju kriterija na osnovu kojeg pomoću eksperimentalnih vrijednosti slučajne varijable možemo odlučiti prihvaćamo li ili odbacujemo hipotezu. Parametarske hipoteze odnose se na parametre poznate funkcije distribucije slučajne varijable, a neparametarske se odnose na nepoznatu razdiobu.

Definicija 15.1 *(STATISTIČKA HIPOTEZA)*

Statistička hipoteza je bilo koja pretpostavka o distribuciji neke slučajne varijable (slučajnog vektora). Neparametarska hipoteza je pretpostavka o funkciji distribucije neke slučajne varijable. Parametarska hipoteza je pretpostavka o parametrima poznate distribucije neke slučajne varijable. Statističke hipoteze označavamo s $H_0, H_1, .., H_n$.

PRIMJER 15.1 $H_0(\lambda = \lambda_0)$ *je parametarska hipoteza za parametar* λ *u Poissonovoj distribuciji.*

Definicija 15.2 *(PROSTA HIPOTEZA)*

Parametarska hipoteza je prosta hipoteza ako sadrži samo jednu pretpostavku o parametru (npr. $H_0(\mu = \mu_0)$*). Parametarska hipoteza je složena ako se sastoji od konačno ili beskonačno mnogo prostih hipoteza*
(npr. $H_1(\mu > \mu_0)$, $H_1(\mu < \mu_0)$, $H_1(\mu \neq \mu_0)$*).*

Definicija 15.3 *(STATISTIČKI TEST)*

Testiranje statističkih hipoteza moguće je ako postaje barem dvije alternativne hipoteze: H_0 *nul-hipoteza i njoj alternativna* H_1 *hipoteza (u nekom smislu suprotna). Statistiku* T *na uzorku* $(X_1, X_2, ..., X_n)$ *pomoću koje se donosi odluka o prihvaćanju nul-hipoteze* *ili prihvaćanju neke od alternativno postavljenih hipoteza ako je dobivena vrijednost slučajnog uzorka $(x_1, x_2, ..., x_n)$ zovemo statističkom testom. Test statistika ima svoju poznatu distribuciju.*

PRIMJER 15.2 *Neka je nul-hipoteza $H_0(p = p_0)$ za slučajne varijable koje imaju $B(1, p)$ binomnu distribuciju.*
Test statistika $T = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ ima normalnu distribuciju $T \sim N(p, \frac{1}{n}p(1-p))$, za $n \to \infty$.

Definicija 15.4 *(GREŠKA PRVE VRSTE, NIVO ZNAČAJNOSTI)*

Testom je napravljena greška prve vrste ako se nije prihvatila ispravna nul-hipoteza. Vjerojatnost da se napravi greška prve vrste naziva se nivo (razina) značajnosti i označava s α.

Definicija 15.5 *GREŠKA DRUGE VRSTE, JAKOST TESTA*

Testom je napravljena greška druge vrste ako se prihvatila lažna nul-hipoteza.
Vjerojatnost da se napravi greška druge vrste označava s β.
Vjerojatnost da se ne napravi greška druge vrste zove se jakost testa označava s η, $\eta = 1-\beta$.
(Jakost testa je vjerojatnost da se odbaci lažna nul-hipoteza.)

Definicija 15.6 *Kritično područje*

Vrijednost t_{kr} statistike T na uzorku $(X_1, X_2, ..., X_n)$ na osnovu koje odlučujemo prihvacamo li ili odbacujemo nul-hipotezu H_0 zove se kritična točka.
Skup vrijednosti statistike T za koje se prihvaća H_0 zove se područje prihavćanja nul-hipoteze, a skup vrijednosti statistike T za koje se ne prihvaća H_0 zove se kritično područje.
Jednostrano kritično područje može biti određeno s uvjetom $T > t_{kr}$ za $t_{kr} > 0$ ili uvjetom $T < t_{kr}$ za $t_{kr} < 0$.
Dvostrano kritično područje određeno je s uvjetom $T < t_{kr1}$ i $T > t_{kr2}$ za $t_{kr1} < t_{kr2}$.

NAPOMENA 15.1 *Neka je zadan nivo značajnosti α i test statistika T za parametarsku hipotezu $H_0(parametar = parametar_0)$.*
(a) Za dvostrani test:

$$H_0 \; (parametar = parametar_0)$$

$$H_1 \; (parametar \neq parametar_0)$$

kritične točke t_{kr1}, t_{kr2}, određujemo iz uvjeta
$P(T < t_{kr1}) = \frac{\alpha}{2}$, $P(T > t_{kr2}) = \frac{\alpha}{2}$;
$t_{kr1} = z_{\frac{\alpha}{2}}$ i $t_{kr2} = z_{1-\frac{\alpha}{2}}$

su kvantili za F, funkciju distribucije test statistike T.
Ako je vrijednost test statistike $t \in (t_{kr1}, t_{kr2})$ *prihvaćamo nul-hipotezu* H_0*, inače prihvaćamo alternativnu hipotezu* H_1.
(b) Za jednostrani test:

$$H_0(parametar = parametar_0)$$

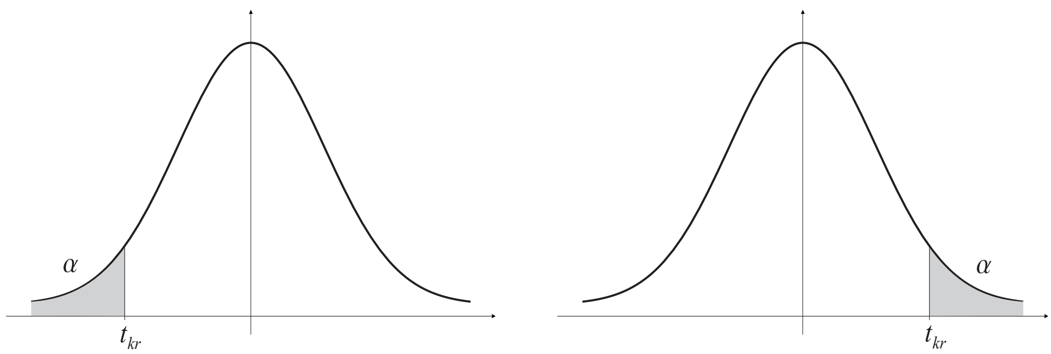
$$H_1(parametar < parametar_0)$$

lijevu kritičnu točku ($t_{kr} < 0$) *određujemo iz uvjeta*
$P(T < t_{kr}) = \alpha;$
$t_{kr} = z_\alpha$ *je kvantili za F, funkciju distribucije test statistike T.*
Ako je vrijednost test statistike $t \in (t_{kr}, \infty)$ *prihvaćamo nul-hipotezu* H_0*, inače prihvaćamo alternativnu hipotezu* H_1.
(c) Za jednostrani test:

$$H_0(parametar = parametar_0)$$

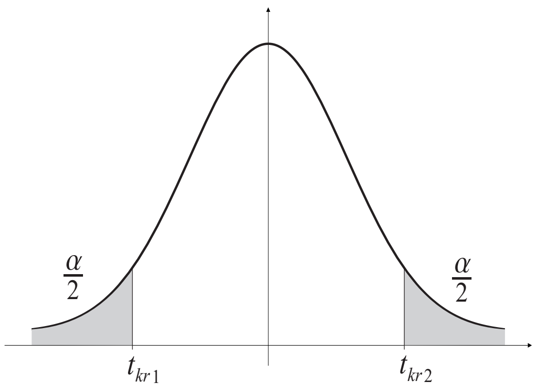
$$H_1(parametar > parametar_0)$$

desnu kritičnu točku ($t_{kr} > 0$) *određujemo iz uvjeta*
$P(T > t_{kr}) = \alpha;$
$t_{kr} = z_{1-\alpha}$ *je kvantili za F funkciju distribucije test statistike T.*
Ako je vrijednost test statistike $t \in (-\infty, t_{kr})$ *prihvaćamo nul-hipotezu* H_0*, inače prihvaćamo alternativnu hipotezu* H_1.



Slika 15.1: Kritično područje za jednostrani (lijevi i desni) test.

NAPOMENA 15.2 *Vrijednosti* α, β *ovise o* t_{kr}. *Želimo da obe greške budu male a to je kontradiktorno (ako* α *opada,* t_{kr} *se miče u desno i* β *raste.)*
U praksi se na početku izabere $\alpha = 0.05$ *ili* $\alpha = 0.01$*, zatim se odredi* t_{kr} *i na kraju se izračuna* β. *Ako je* β *veliko onda ponavljamo test pomoću većeg uzorka.*



Slika 15.2: Kritično područje za dvostrani test.

NAPOMENA 15.3 *Svi teoremi o testiranju hipoteza o parametrima (očekivanje, vjerojatnost i varijanca) su dani bez dokaza. Dokazi se temelje na rezultatima dobivenim u poglavlju Intervali povjerenja i u uvodu objašnjenog postupka.*

15.1 TEST HIPOTEZE ZA VJEROJATNOST BINOMNE DISTRIBUCIJE $n \to \infty$

MOTIV 15.1 *Proizvođač lijeka tvdri da je lijek za alergiju djelotvoran* 90% *u vremenskom razdoblju od 8 sati tako da se ukloni alergijska reakcija kože. Liječnik je vršio istraživanje. Na uzorku od 200 pacijenata s alergijskim problemima lijek je bio učinkovit kod njih 160. Koji je zaključak liječnik dobio- treba li vjerovati deklaraciji proizvođača? Pretpostavimo da je odluku donio uz razinu značajnosti* 1% *tj. vjerojatnost da je odbacio istinitu tvrdnju je* 1%.

TEOREM 15.1 *Ako je broj ponavljanja u Bernoullijevoj shemi veliki onda za parametarsku hipotezu $H_0(p = p_0)$ test statistika*

$$T = \sqrt{n}\frac{\overline{X} - p_0}{\sqrt{p_0(1-p_0)}}$$

ima standardnu normalnu distribuciju $T \sim N(0,1)$.
Neka je zadan nivo (razina) značajnosti α.
(a) Za dvostrani test:
nul-hipoteza $H_0(p = p_0)$ i alternativna $H_1(p \neq p_0)$
kritične točke t_{kr1}, t_{kr2}, određujemo iz uvjeta

$$P(T < t_{kr1}) = \frac{\alpha}{2},\ P(T > t_{kr2}) = \frac{\alpha}{2};$$

$t_{kr1} = z_{\frac{\alpha}{2}}$, $t_{kr2} = z_{1-\frac{\alpha}{2}}$ su kvantili za F^*, funkciju distribucije $T \sim N(0,1)$.
Ako je vrijednost test statistike
$t = \sqrt{n}\frac{\bar{x}-p_0}{\sqrt{p_0(1-p_0)}} \in (t_{kr1}, t_{kr2})$ prihvaćamo nulhipotezu H_0, inače prihvaćamo alternativnu H_1.
(b) Za jednostrani test:
nul-hipoteza $H_0(p = p_0)$ i alternativna $H_1(p < p_0)$
kritičnu točku t_{kr} određujemo iz uvjeta $P(T < t_{kr}) = \alpha$;
$t_{kr} = z_\alpha$ je kvantili za F^*, funkciju distribucije $T \sim N(0,1)$.
Ako je vrijednost test statistike
$t = \sqrt{n}\frac{\bar{x}-p_0}{\sqrt{p_0(1-p_0)}} \in (t_{kr1}, \infty)$ prihvaćamo nul-hipotezu H_0, inač prihvaćamo alternativnu H_1.
(c) Za jednostrani test:

nul-hipoteza $H_0(p = p_0)$ i alternativna $H_1(p > p_0)$
kritičnu točku t_{kr} određujemo iz uvjeta $P(T > t_{kr}) = \alpha$;
$t_{kr} = z_{1-\alpha}$ je kvantili za F^*, funkciju distribucije $T \sim N(0,1)$.
Ako je vrijednost test statistike
$t = \sqrt{n}\frac{\bar{x}-p_0}{\sqrt{p_0(1-p_0)}} \in (-\infty, t_{kr})$ *prihvaćamo nul-hipotezu H_0, inač prihvaćamo alternativnu H_1.*

PRIMJER 15.3 *Napravljen je slučajni pokus bacanje novčića. Pokus je ponovljen n=4040 puta i dobiveno je 2048 grbova. Znamo da je uspjeh u pokusu binomna slučajna varijabla s parametrom p vjerojatnost da padne grb. Testiramo nulhipotezu $H_0(p = 0.5)$ tj. testiramo hipotezu da je novčić ispravan. Alternativna hipozeza je $H_1(p \neq 0.5)$. Neka je nivo značajnosti $\alpha = 0.05$.*

Rješenje: Postavljamo hipoteze:
nul-hipoteza $H_0(p = 0.5)$
alternativna hipoteze $H_1(p \neq 0.5)$ $\quad H_1(p < 0.5)$ i $H_1(p > 0.5)$.
Za zadani $\alpha = 0.05$ provjeravamo dvostrani test.
$t_{kr1} = z_{0.025}$, $t_{kr2} = z_{0.975}$ su kvantili standardne normalne distribucije F^*.

$$-t_{kr1} = z_{1-0.025} = z_{0.975} = 1.96, \ \Rightarrow \ t_{kr1} = -1.96, \ t_{kr2} = z_{0.975} = 1.96.$$

Područje prihvaćanja za nul-hipotezu $H_0(p = 0.5)$ za nivo značajnosti $\alpha = 0.05$ je $(t_{kr1}, t_{kr2}) = (-1.96, 1.96)$.
Vrijednost test statistike $T(X_1, X_2, ..., X_n)$, je

$$t = \sqrt{n}\frac{\bar{x}-p_0}{\sqrt{p_0(1-p_0)}} = \sqrt{4040} \cdot \frac{\frac{2048}{4040} - 0.5}{\sqrt{0.5 \cdot (1-0.5)}} = 0.881\,04.$$

Kako je $t \in (t_{kr1}, t_{kr2})$ unutar područja prihvatljivosti nul-hipoteze, prihvaćamo nul-hipotezu $H_0(p = 0.5)$, novčić je ispravan.

Prisjetimo se motivacijskog primjera:

PRIMJER 15.4 *motiv*

Proizvođač lijeka tvdri da je lijek za alergiju djelotvoran 90% *u vremenskom razdoblju od 8 sati tako da se ukloni alergijska reakcija kože. Liječnik je vršio istraživanje. Na uzorku od 200 pacijenata s alergijskim problemima lijek je bio učinkovit kod njih 160. Koji je zaključak liječnik dobio- treba li vjerovati deklaraciji proizvođača? Pretpostavimo da je odluku donio uz razinu značajnosti* 1% *tj. vjerojatnost da je odbacio istinitu tvrdnju je* 1%.

Rješenje: Postavljamo hipoteze:
$H_0(p = 0.9)$, deklaracija o lijeku je isprvna
$H_1(p < 0.9)$ deklaracija je lažna.
U zadatku je zadano $\alpha = 0.01$, $p_0 = 0.9$, $\overline{x} = \frac{160}{200}$, $n = 200$.

Test statistika je $T = \sqrt{n}\dfrac{\overline{X} - p_0}{\sqrt{p_0(1-p_0)}}$ i ima standardnu normalnu distribuciju $T \sim N(0,1),\ n \to \infty$.
Za jednostrani test, kritičnu točku t_{kr} određujemo iz uvjeta $P(T < t_{kr}) = \alpha$. Tako je $t_{kr} = z_\alpha$ kvantili za F*, funkciju distribucije $T \sim N(0,1)$.
Kritična točka je $t_{kr} = z_\alpha = -z_{1-\alpha} = -2.33$.
Područje prihvaćanja za nul-hipotezu $H_0(p = 0.9)$ za nivo značajnosti $\alpha = 0.01$ je $(t_{kr}, \infty) = (-2.33, \infty)$.
Vrijednost test statistike $T(X_1, X_2, ..., X_n)$,

$$t = \sqrt{n}\frac{\overline{x} - p_0}{\sqrt{p_0(1-p_0)}} = \sqrt{200} \cdot \frac{\frac{160}{200} - 0.9}{\sqrt{0.9 \cdot (1 - 0.9)}} = -4.72.$$

Kako $t \notin (t_{kr}, \infty)$ nije unutar područja prihvatljivosti nul-hipoteze, prihvaćamo alternativnu $H_1(p < 0.9)$ tj. deklaracija nije vjerodostojna uz razinu značajnosti 1%.

15.2 TESTIRANJE HIPOTEZE ZA OČEKIVANJE NORMALNE DISTRIBUCIJE kad je varijanca poznata

(i za $n > 30$)

MOTIV 15.2 *Prekidna čvrstoća kablova proizvedenih u jednoj tvornici ima uzoračku aritmetičku sredinu 815 kg i standardnu devijaciju 45 kg. Novom tehnologijom u proizvodnom procesu tvrdi se da je prekidna čvrstoća povećana. Da testira tu tvrdnju kontrolor je testirao 50 kablova i dobio uzoračku aritmetičku sredinu 840 kg. Hoće li kontorlor prihvatiti tvrdnju proizvođača uz razinu značajnosti* 1%.

U 13. i 14. poglavlju (Intervali povjerenja) uočili smo da pomoću statistika $T(X_1, ..., X_n)$ donosimo zaključke o parametrima teorijske distribucije (očekivanje μ i varijanca σ^2). U slučaju velikog uzorka sve statsitike imaju normalnu distribuciju $N(\mu, \frac{\sigma^2}{n})$ bez obzira o teorijsku distribuciju populacije. Zato, za velike uzorke, ako nije poznata varijanca, možemo korigiranu uzoračku varijancu $\widehat{s^2}$ i uzoračku varijancu $\widehat{\sigma^2}$ uzeti za procjenu σ^2.

TEOREM 15.2 *Neka je $(X_1, X_2, ..X_n)$ slučajni uzorak slučajne varijable X koja ima poznatu distribuciju (normalna) s nepoznatim parametrom očekivanje μ i poznatom varijancom σ^2 (ili poznatom korigiranom uzoračkom varijancom $\widehat{s^2}$). Ako je veliki uzorak $(n \to \infty)$ za parametarsku hipotezu $H_0(\mu = \mu_0)$ test statistika $T = \sqrt{n}\frac{\overline{X}-\mu_0}{\sigma}$ ima standardnu normalnu distribuciju $T \sim N(0, 1)$.*
Neka je zadan nivo (razina) značajnosti α.
(a) Za dvostrani test:
nul-hipoteza $H_0(\mu = \mu_0)$ i alternativna $H_1(\mu \neq \mu_0)$,
kritične točke t_{kr1}, t_{kr2}, određujemo iz uvjeta

$$P(T < t_{kr1}) = \frac{\alpha}{2}, \; P(T > t_{kr2}) = \frac{\alpha}{2};$$

$t_{kr1} = z_{\frac{\alpha}{2}}$, $t_{kr2} = z_{1-\frac{\alpha}{2}}$ su kvantili za F, funkciju distribucije $T \sim N(0, 1)$.*
Ako je vrijednost test statistike $T(X_1, X_2, ..., X_n)$,
$t = \sqrt{n}\frac{\overline{x}-\mu_0}{\sigma} \in (t_{kr1}, t_{kr2})$ prihvaćamo nulhipotezu H_0, inače prihvaćamo alternativnu H_1.
(b) Za jednostarni test:
nul-hipoteza $H_0(\mu = \mu_0)$ i alternativna $H_1(\mu < \mu_0)$,

kritičnu točku t_{kr} određujemo iz uvjeta $P(T < t_{kr}) = \alpha$;
$t_{kr} = z_\alpha$ je kvantil za F^*, funkciju distribucije test statistike $T \sim N(0,1)$.
Ako je vrijednost test statistike $T(X_1, X_2, ..., X_n)$,
$t = \sqrt{n}\frac{\overline{x}-\mu_0}{\sigma} \in (t_{kr}, \infty)$ prihvaćamo nul-hipotezu H_0, inače prihvaćamo alternativnu H_1.
(c) Za jednostarni test:
nul-hipoteza $H_0(\mu = \mu_0)$ i alternativna $H_1(\mu > \mu_0)$,
kritičnu točku t_{kr} određujemo iz uvjeta $P(T > t_{kr}) = \alpha$;
$t_{kr} = z_{1-\alpha}$ je kvantil za F^*, funkciju distribucije test statistike $T \sim N(0,1)$.
Ako je vrijednost test statistike $T(X_1, X_2, ..., X_n)$,
$t = \sqrt{n}\frac{\overline{x}-\mu_0}{\sigma} \in (-\infty, t_{kr})$ prihvaćamo nul-hipotezu H_0, inače prihvaćamo alternativnu H_1.

PRIMJER 15.5 *Slučajna varijabla $X \sim N(\mu, \sigma^2 = 4)$. U uzorku veličine n=25 vrijednost uzoračke aritmetičke sredine je $\overline{x} = 14.7$. Za nivo značajnosti $\alpha = 0.01$ testirati*
(a) nul-hipotezu $H_0(\mu = 16)$ i alternativnu hipotezu $H_1(\mu \neq 16)$,
(b) nul-hipotezu $H_0(\mu = 16)$ i alternativnu hipotezu $H_1(\mu < 16)$.

Rješenje :
Zadan je uzorak iz normalne distribucije pa za parametarsku hipotezu $H_0(\mu = \mu_0)$ test statistika $T = \sqrt{n}\frac{\overline{X}-\mu_0}{\sigma}$ ima standardnu normalnu distribuciju $T \sim N(0,1)$.
Nivo značajnosti je $\alpha = 0.01$, i $n = 25$.
(a)Testiramo:
nul-hipotezu $H_0(\mu = 16)$ i alternativnu hipotezu $H_1(\mu \neq 16)$,
Kritične točke su kvantili za F^* :

$$t_{kr1} = z_{\frac{\alpha}{2}} = z_{0.005}, \ \ -t_{kr1} = z_{0.995}, \ \ -t_{kr1} = 2.58, \ \ t_{kr1} = -2.58.$$

$$t_{kr2} = z_{1-\frac{\alpha}{2}} = z_{0.995}, \ \ t_{kr2} = 2.58.$$

Područje prihvaćanja za nul-hipotezu $H_0(\mu = 16)$ za nivo značajnosti $\alpha = 0.01$ je $(t_{kr1}, t_{kr2}) = (-2.58, 2.58)$.
Vrijednost test statistike $T(X_1, X_2, ..., X_n)$,
$t = \sqrt{n}\frac{\overline{x}-\mu_0}{\sigma} = \sqrt{25}\frac{14.7-16}{2} = -3.25 \notin (t_{kr1}, t_{kr2})$ pa odbacujemo nul-hipotezu $H_0(\mu = 16)$ i prihvaćamo alternativnu hipotezu $H_1(\mu \neq 16)$.
(b) Testiramo:
nul-hipotezu $H_0(\mu = 16)$ i alternativnu hipotezu $H_1(\mu < 16)$.
Kritična točka je kvantili za F^* : $t_{kr} = z_\alpha = z_{0.01}, \ t_{kr} = -2.33$.
Područje prihvaćanja za nul-hipotezu $H_0(\mu = \mu_0)$ za nivo značajnosti $\alpha = 0.01$ je $(t_{kr}, \infty) = (-2.33, \infty)$.

Vrijednost test statistike $T(X_1, X_2, ..., X_n)$,
$t = \sqrt{n}\frac{\overline{x}-\mu_0}{\sigma} = \sqrt{25} \cdot \frac{14.7-16}{2} = -3.,25 \notin (-2.33, \infty)$, pa odbacujemo nul-hipotezu i prihvaćamo alternativnu hipotezu $H_1(\mu < 16)$.

Prisjetimo se motivacijskog primjera:

PRIMJER 15.6 *motiv*

Prekidna čvrstoća kablova proizvedenih u jednoj tvornici ima uzoračku aritmetičku sredinu 815 kg i standardnu devijaciju 45 kg. Novom tehnologijom u proizvodnom procesu tvrdi se da je prekidna čvrstoća povećana. Da testira tu tvrdnju kontrolor je testirao 50 kablova i dobio uzoračku aritmetičku sredinu 840 kg. Hoće li kontorlor prihvatiti tvrdnju proizvođača uz razinu značajnosti 1%.

Rješenje :
Uzorak je veliki $n > 30$, pa za parametarsku hipotezu $H_0(\mu = \mu_0)$ test statistika $T = \sqrt{n}\frac{\overline{X}-\mu_0}{\sigma}$ ima standardnu normalnu distribuciju $T \sim N(0, 1)$.
Testiramo:
nul-hipoteza $H_0(\mu = 815)$ i alternativna $H_1(\mu > 815)$
pomoću jednostranog testa.
U zadatku su poznati podatci $n = 50$, $\widehat{\sigma} = 45$, $\overline{x} = 840$, $\alpha = 0.01$.
Za veliki uzorak standardnu devijaciju σ procijenit ćemo sa uzoračkom devijacijom $\widehat{\sigma}$. Za jednostrani test, kritičnu točku t_{kr} određujemo iz uvjeta $P(T > t_{kr}) = \alpha$; $t_{kr} = z_{1-\alpha} = 2.33$ je kvantil standardne normalne distribucije.

Područje prihvaćanja za nul-hipotezu $H_0(\mu = 815)$ za nivo značajnosti 0.01 je $(-\infty, 2.33)$.
Vrijednost test statistike $T(X_1, X_2, ..., X_n)$,
$t = \sqrt{n}\frac{\overline{x}-\mu_0}{\sigma} = \sqrt{50}\frac{840-815}{45} = 3.93 \notin (-\infty, 2.33)$ pa odbacujemo nul-hipotezu i prihvaćamo alternativnu hipotezu $H_1(\mu > 815$. Kontrolor će prihvatiti tvrdnju proizvođača o poboljšanju prekidne čvrstoće.

15.3 TESTIRANJE HIPOTEZA ZA OČEKIVANJE NORMALNE DISTRIBUCIJE kad je varijanca nepoznata

MOTIV 15.3 *Da testira prekidnu čvrstoću užadi kontrolor je testirao 6 užadi i dobio uzoračku aritmetičku sredinu 3515 kg i korigiranu uzoračku varijancu 66 kg. Proizvođač tvrdi da je 3630 kg. Hoće li kontorlor prihvatiti tvrdnju proizvođača uz razinu značajnosti* 1%.

MOTIV 15.4 *Izvršena su mjerenja tlačne čvrstoće* $[\frac{kg}{cm^2}]$ *uzorka betona visoke kvalitete (A) i običnog betona (B). Dobiveni su sljedeće vrijednosti*

A	*357*	*356*	*413*
B	*346*	*358*	*302*

Pretpostavimo da su tlačne čvrstoće normalno distribuirane i da imaju jednake varijance. Testirajte hipotezu da je oba tipa betona imaju jednaka očekivanja $\mu_1 = \mu_2$ *uz alternativnu hipotezu da je* $\mu_1 > \mu_2$, *uz razinu značajnosti* 5%.

TEOREM 15.3 *Neka je* $(X_1, X_2, ..X_n)$ *slučajni uzorak slučajne varijable X koja ima normalnu distribuciju s nepoznatim parametrom očekivanje* μ *i nepoznatom varijancom* σ^2. *Za parametarsku hipotezu* $H_0(\mu = \mu_0)$ *test statistika*

$$T = \sqrt{n}\frac{\overline{X} - \mu_0}{\widehat{S}} = \sqrt{n-1}\frac{\overline{X} - \mu_0}{\widehat{\Sigma}}$$

ima Studentovu distribuciju $T \sim t(n-1)$.
Neka je zadan nivo (razina) značajnosti α.
(a) Za dvostrani test:
nul-hipoteza $H_0(\mu = \mu_0)$ *i alternativna* $H_1(\mu \neq \mu_0)$,
kritične točke t_{kr1}, t_{kr2}, *određujemo iz uvjeta*

$$P(T < t_{kr1}) = \frac{\alpha}{2},\ P(T > t_{kr2}) = \frac{\alpha}{2};$$

$t_{kr1} = z_{\frac{\alpha}{2}}$, $t_{kr2} = z_{1-\frac{\alpha}{2}}$ *su kvantili za F, funkciju distribucije* $T \sim t(n-1)$.
Ako je vrijednost test statistike $T(X_1, X_2, ..., X_n)$,
$t = \sqrt{n}\frac{\bar{x}-\mu_0}{\widehat{s}} \in (t_{kr1}, t_{kr2})$ *prihvaćamo nulhipotezu* H_0, *inače prihvaćamo alternativnu* H_1.

(b) Za jednostarni test:
nul-hipoteza: $H_0(\mu = \mu_0)$ *i alternativna* $H_1(\mu < \mu_0)$*, kritičnu točku* t_{kr} *određujemo iz uvjeta* $P(T < t_{kr}) = \alpha$.
$t_{kr} = z_\alpha$ *je kvantil za F, funkciju distribucije test statistike* $T \sim t(n-1)$.
Ako je vrijednost test statistike $T(X_1, X_2, ..., X_n)$,
$t = \sqrt{n}\frac{\overline{x}-\mu_0}{\widehat{s}} \in (t_{kr}, \infty)$ *prihvaćamo* H_0*, inače prihvaćamo alternativnu* H_1.
(c) Za jednostarni test:
nul-hipoteza $H_0(\mu = \mu_0)$ *i alternativna* $H_1(\mu > \mu_0)$ *kritičnu točku* t_{kr} *određujemo iz uvjeta* $P(T > t_{kr}) = \alpha$;
$t_{kr} = z_{1-\alpha}$*, je kvantil za F, funkciju distribucije test statistike* $T \sim t(n-1)$.
Ako je vrijednost test statistike $T(X_1, X_2, ..., X_n)$,
$t = \sqrt{n}\frac{\overline{x}-\mu_0}{\widehat{s}} \in (-\infty, t_{kr})$ *prihvaćamo nulhipotezu* H_0*, inače prihvaćamo alternativnu* H_1.

NAPOMENA 15.4 *Kvantili za Studentovu distribuciju za* $n = 5,\ t(4)$,
$F(z_\alpha) = \alpha$:

α	0.05	0.01
z_α	-2.13	-1.53
$z_{1-\alpha}$	2.13	1.53

α	0.05	0.01
$\frac{\alpha}{2}$	0.025	0.005
$z_{\frac{\alpha}{2}}$	-2.78	-4.60
$z_{1-\frac{\alpha}{2}}$	2.78	4.60

PRIMJER 15.7 *Neka slučajna varijabla ima normalnu distribuciju* $X \sim N(\mu, \sigma^2)$ *nepoznate varijance* σ^2*. Uzet je utorak veličine* $n = 5$ *i dobivena je vrijednost uzoračke aritmetičke sredine* $\overline{x} = 10.2$*, i vrijednost korigirane uzoračke varijance* $\widehat{s^2} = 0.64$.
Za nivo značajnosti $\alpha = 0.05$ *testirati*
(a) nul-hipotezu $H_0(\mu = 10)$ *i alternativnu hipotezu* $H_1(\mu \neq 10)$,
(b) nul-hipotezu $H_0(\mu = 10)$ *i alternativnu hipotezu* $H_1(\mu < 10)$.

Rješenje:
Za parametarsku hipotezu $H_0(\mu = \mu_0)$ test statistika

$$T = \sqrt{n}\frac{\overline{X}-\mu_0}{\widehat{S}} = \sqrt{n-1}\frac{\overline{X}-\mu_0}{\widehat{\Sigma}}$$

ima Studentovu distribuciju $T \sim t(n-1)$.
Za nivo značajnosti $\alpha = 0.05$ i $n = 5$ (a) testiramo:

nul-hipotezu $H_0(\mu = 10)$ i alternativnu hipotezu $H_1(\mu \neq 10)$.
Za $\alpha = 0.05,\ n-1 = 4$,
kritične točke su kvantili Studentove distribucije $t(5-1)$:

$$t_{kr1} = z_{\frac{\alpha}{2}} = z_{0.025},\ \ -t_{kr1} = z_{0.975},\ \ -t_{kr1} = 2.78,\ \ t_{kr1} = -2.78,$$

$$t_{kr2} = z_{1-\frac{\alpha}{2}} = z_{0.995},\ \ t_{kr2} = 2.78.$$

Područje prihvaćanja za nulhipotezu $H_0(\mu = \mu_0)$ za nivo značajnosti $\alpha = 0.05$ je $(t_{kr1}, t_{kr2}) = (-2.78, 2.78)$.
Vrijednost test statistike $T(X_1, X_2, ..., X_n)$,
$t = \sqrt{n}\frac{\overline{x}-\mu_0}{\widehat{s}} = \sqrt{5}\cdot\frac{10.2-10}{0.8} = 0.559\,02 \in (t_{kr1}, t_{kr2})$ je upala u područje prihvaćanja, pa prihvaćamo nulhipotezu $H_0(\mu = 10)$.
(b) Testiramo:
nul-hipotezu $H_0(\mu = 10)$ i alternativnu hipotezu $H_1(\mu < 10)$.
Za $\alpha = 0.05$ kritična točka je kvantil Studentove distribucije $t(5-1)$:

$$t_{kr} = z_{\alpha} = z_{0.05},\ \ -t_{kr} = z_{0.95},\ \ -t_{kr} = 2.13,\ \ t_{kr} = -2.13.$$

Područje prihvaćanja za nul-hipotezu $H_0(\mu = 10)$ za nivo značajnosti $\alpha = 0.05$ je $(t_{kr}, \infty) = (-2.13, \infty)$.
Vrijednost test statistike $T(X_1, X_2, ..., X_n)$,
$t = \sqrt{n}\frac{\overline{x}-\mu_0}{\widehat{s}} = \sqrt{5}\cdot\frac{10.2-10}{0.8} = 0.559 \in (-2.13, \infty)$ pa prihvaćamo nul-hipotezu $H_0(\mu = 10)$.

Prisjetimo se motivacijskog primjera:

PRIMJER 15.8 *motiv*

Da testira prekidnu čvrstoću užadi kontrolor je testirao 6 užadi i dobio uzoračku aritmetičku sredinu 3515 kg i uzoračku standardnu devijaciju 66 kg. Proizvođač tvrdi da je 3630 kg. Hoće li kontorlor prihvatiti tvrdnju proizvođača uz razinu značajnosti 1%.

Rješenje:
Testira se:
nul-hipoteza $H_0(\mu = 3630)$ i alternativna $H_1(\mu < 3630)$
pomoću jednostranog testa.

Test statistika $T = \sqrt{n}\frac{\overline{X}-\mu}{\widehat{S}} = \sqrt{n-1}\frac{\overline{X}-\mu}{\widehat{\Sigma}}$ ima studentovu distribuciju $T \sim t(n-1)$.
Za jednostrani test kritičnu točku t_{kr} određujemo iz uvjeta $P(T < t_{kr}) = \alpha$.

$t_{kr} = z_\alpha$, je kvantil za F, funkciju distribucije test statistike $T \sim t(n-1)$.
Za $\alpha = 0.01$ i $n = 6$, $t_{kr} = z_\alpha = z_{0.01}$, $-t_{kr} = z_{0.99}$, $-t_{kr} = 3.37$, $t_{kr} = -3.37$.
Područje prihvaćanja za nul-hipotezu $H_0(\mu = 3630)$ za nivo značajnosti $\alpha = 0.01$ je $(-3.37, \infty)$.
Vrijednost test statistike $T(X_1, X_2, ..., X_n)$,
$t = \sqrt{n-1}\frac{\bar{x}-\mu_0}{\widehat{\sigma}} = \sqrt{5} \cdot \frac{3515-3630}{66} = -3.89 \notin (-3.37, \infty)$ pa ne prihvaćamo nul-hipotezu $H_0(\mu = 3630)$, nego alternativnu $H_1(\mu < 3630)$. Kontrolor će odbaciti tvrdnju proizvođača s razinom značajnosti 1%.

NAPOMENA 15.5 *Neka su $(X_1, X_2, ...X_n)$ i $(Y_1, Y_2, ..., Y_n)$ zadana dva slučajna uzorka iz normalne distribucije nepoznatih očekivanja μ_1 i μ_2 ali jedankih varijanci $\sigma_1^2 = \sigma_2^2$. Ako testiramo $H_0(\mu_1 = \mu_2)$*
uz alternativne hipoteze $H_1(\mu_1 > \mu_2)$ ili $H_1(\mu_1 < mu_2)$ ili $H_1(\mu_1 \neq \mu_2)$
definiramo $\mu = \mu_1 - \mu_2$ i testiramo $H_0 : \mu = 0$
uz alternativne hipoteze $H_1(\mu > 0)$ ili $H_1(\mu < 0)$ ili $H_1(\mu \neq 0)$.

Prisjetimo se motivacijskog primjera:

PRIMJER 15.9 *motiv*
Izvršena su mjerenja tlačne čvrstoće $[\frac{kg}{cm^2}]$ uzorka betona visoke kvalitete (A) i običnog betona (B). Dobiveni su sljedeće vrijednosti

A	*357*	*359*	*413*
B	*346*	*358*	*302*

Pretpostavimo da su tlačne čvrstoće normalno distribuirane i da imaju jednake varijance. Testirajte hipotezu da je oba tipa betona imaju jednaka očekivanja $\mu_1 = \mu_2$ uz alternativnu hipotezu da je $\mu_1 > \mu_2$, uz razinu značajnosti 5%.

Rješenje:
Definiramo novu varijablu razlike čvrstoća D i uzorak razlike čvrstoća

A-B	11	1	111

Uzoračaka aritmetička sredina za D je $\overline{d} = 41$, a korigiranqauzoračka varijanca $\widehat{s_D^2} = 3700$
Za ovaj uzorak testiramo nul-hipotezu $H_0(\mu = 0)$
uz alternativnu hipotezu $H_1(\mu > 0)$. Koristimo test statistiku $T = \sqrt{n}\frac{\overline{X}-\mu_0}{\widehat{S}}$ koja ima

studentovu distribuciju $T \sim t(n-1)$.
Za jednostrani test, kritičnu točku t_{kr} određujemo iz uvjeta $P(T > t_{kr}) = \alpha$.
$t_{kr} = z_{1-\alpha}$, je kvantil za F, funkciju distribucije test statistike $T \sim t(n-1)$.
U zadatku je zadana razina značajnosti $\alpha = 0.05$ i $n = 3$ pa iz tablice $t(2)$ očitavamo $t_{kr} = z_{1-\alpha} = 2.92$.

Područje prihvaćanja za nul-hipotezu $H_0(\mu = 0)$ za nivo značajnosti $\alpha = 0.05$ je $(-\infty, 2.92)$.
Ako je vrijednost test statistike $t = \sqrt{n}\frac{\bar{d}-\mu_0}{\widehat{s_D}} \in (t_{kr}, \infty)$ prihvaćamo nul-hipotezu, inače prihvaćamo alternativnu.
$t = \sqrt{3}\frac{41-0}{60.82} = 1.167 \in (-\infty, 2.92)$, pa prihvaćamo nul-hiptezu $H_0 : \mu = 0$, tj. nul-hipotezu $H_0(\mu_1 = \mu_2)$.

S razinom značajnosti 5% prihvaćamo hipotezu da oba betona imaju jednaka očekivanja čvrstoće.

15.4 TESTIRANJE HIPOTEZA za varijancu normalne razdiobe

MOTIV 15.5 *Očekivana masa palete opeka je 1166.4 kg. Standardna devijacija mase palete opeka je 5.5 kg. Kontrolor je testirao uzorak od 20 paleta i dobio uzoračku standardnu devijaciju 6.5 kg. Može li kontrolor zaključiti da je standardna devijacija u porastu uz nivo značajnosti* 5%*?*

TEOREM 15.4 *Neka je $(X_1, X_2, ..X_n)$ slučajni uzorak slučajne varijable X koja ima normalnu distribuciju s nepoznatim parametrom varijancom σ^2. Za parametarsku hipotezu $H_0(\sigma^2 = \sigma_0^2)$ test statistika*

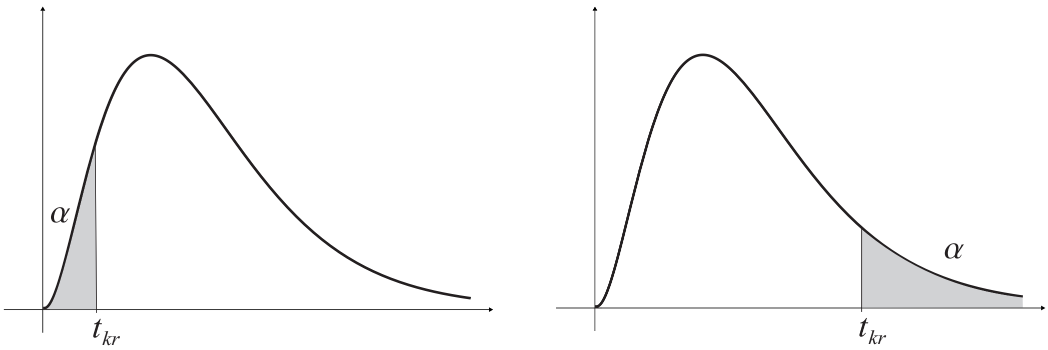
$$T = \frac{n-1}{\sigma_0^2}\widehat{S}^2 = \frac{n}{\sigma_0^2} \cdot \widehat{\Sigma}^2$$

ima hikvadrat distribuciju s $(n-1)$ stupnjeva slobode $T \sim \chi^2(n-1)$.
Neka je zadan za nivo (razina) značajnosti α.
(a) Za dvostarni test:
nul-hipoteza $H_0(\sigma^2 = \sigma_0^2)$ i alternativna $H_1(\sigma^2 \neq \sigma_0^2)$,
kritične točke t_{kr1}, t_{kr2}, određujemo iz uvjeta

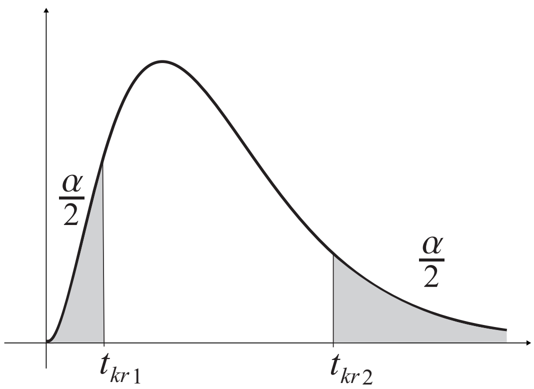
$$P(T < t_{kr1}) = \frac{\alpha}{2},\ P(T > t_{kr2}) = \frac{\alpha}{2}.$$

$t_{kr1} = z_{\frac{\alpha}{2}}$, $t_{kr2} = z_{1-\frac{\alpha}{2}}$ su kvantili za funkciju distribucije $\chi^2(n-1)$.
Ako je vrijednost test statistike $T(X_1, X_2, ..., X_n)$,
$t = \frac{(n-1)}{\sigma_0^2}\widehat{s}^2 \in (t_{kr1}, t_{kr2})$
prihvaćamo nul-hipotezu H_0, inače prihvaćamo alternativnu H_1.
(b) Za jednostrani test:
nul-hipoteza $H_0(\sigma^2 = \sigma_0^2)$ i alternativna $H_1(\sigma^2 < \sigma_0^2)$,
kritičnu točku t_{kr} određujemo iz uvjeta $P(T < t_{kr}) = \alpha$.
$t_{kr} = z_\alpha$ je kvantil za funkciju distribucije $\chi^2(n-1)$.
Ako je vrijednost test statistike $T(X_1, X_2, ..., X_n)$,
$t = \frac{(n-1)}{\sigma_0^2}\widehat{s}^2 \in (t_{kr}, \infty)$
prihvaćamo nul-hipotezu H_0, inače prihvaćamo alternativnu H_1.
(c) Za jednostarni test:
nul-hipoteza $H_0(\sigma^2 = \sigma_0^2)$ i alternativna $H_1(\sigma^2 > \sigma_0^2)$,
kritičnu točku t_{kr} određujemo iz uvjeta $P(T > t_{kr}) = \alpha$.
$t_{kr} = z_{1-\alpha}$, je kvantil za funkciju distribucije $\chi^2(n-1)$.
Ako je vrijednost test statistike $T(X_1, X_2, ..., X_n)$,

$t = \frac{(n-1)}{\sigma_0^2}\widehat{s}^2 \in (-\infty, t_{kr})$
prihvaćamo nul-hipotezu H_0, inače prihvaćamo alternativnu H_1.



Slika 15.3: Kritično područje za jednostrani (lijevi i desni) test.



Slika 15.4: Kritično područje za dvostrani test.

NAPOMENA 15.6 *Kvantili za hikvadrat distribuciju za $n = 4$, $\chi^2(4)$,
$F(z_\alpha) = \alpha$:*

α	0.05	0.01
z_α	0.71	0.29
$z_{1-\alpha}$	9.48	11.27

α	0.05	0.01
$\frac{\alpha}{2}$	0.025	0.005
$z_{\frac{\alpha}{2}}$	0.48	0.21
$z_{1-\frac{\alpha}{2}}$	11.14	14.86

PRIMJER 15.10 *Neka slučajna varijabla ima normalnu distribuciju
$X \sim N(\mu, \sigma^2)$. U uzorku veličine $n = 5$ vrijednost korigirane uzoračke varijance je*

$\widehat{s^2} = 0.64$ *Za nivo značajnosti* $\alpha = 0.05$ *testirati*
(a) nul-hipotezu $H_0(\sigma^2 = 0.8^2)$ *i alternativnu hipotezu* $H_1(\sigma^2 \neq 0.8)$,
(b) nul-hipotezu $H_0(\sigma^2 = 0.8^2)$ *i alternativnu hipotezu* $H_1(\sigma^2 < 0.8)$.

Rješenje :
Za parametarsku hipotezu $H_0(\sigma^2 = \sigma_0^2)$ test statistika

$$T = \frac{n-1}{\sigma_0^2}\widehat{S^2} = \frac{n}{\sigma_0^2} \cdot \widehat{\Sigma^2}$$

ima hikvadrat distribuciju s $(n-1)$ stupnjeva slobode $T \sim \chi^2(n-1)$.

(a) Testitamo:
nul-hipotezu $H_0(\sigma^2 = 0.8^2)$ i alternativnu hipotezu $H_1(\sigma^2 \neq 0.8^2)$,

Za $\alpha = 0.05$, $n = 5$ kritične točke su kvantili distribucije $\chi^2(4)$:

$$t_{kr1} = z_{\frac{\alpha}{2}} = z_{0.025}, \quad t_{kr1} = 0.48,$$

$$t_{kr2} = z_{1-\frac{\alpha}{2}} = z_{0.995}, \quad t_{kr2} = 11.14.$$

Područje prihvaćanja za nul-hipotezu $H_0(\sigma^2 = 0.8^2)$ za nivo značajnosti
$\alpha = 0.05$ je $(t_{kr1}, t_{kr2}) = (0.48, 11.14)$.
Vrijednost test statistike $T(X_1, X_2, ..., X_n)$,
$t = \frac{(n-1)}{\sigma_0^2}\widehat{s^2} = \frac{4}{0.64}0.64 = 4 \in (0.48, 11.14)$ pa prihvaćamo nul-hipotezu H_0.
(b) Testiramo:
nul-hipotezu $H_0(\sigma^2 = 0.8^2)$ i alternativnu hipotezu $H_1(\sigma^2 < 0.8^2)$.
Za $\alpha = 0.05$, $n = 5$, kritična točka je kvantil distribucije $\chi^2(4)$:
$t_{kr} = z_\alpha = z_{0.05}, \; t_{kr} = 0.71$.
Područje prihvaćanja za nul-hipotezu $H_0(\sigma^2 = 0.64)$ za nivo značajnosti $\alpha = 0.05$ je $(t_{kr}, \infty) = (0.71, \infty)$.
Vrijednost test statistike $T(X_1, X_2, ..., X_n)$,
$t = \frac{(n-1)}{\sigma_0^2}\widehat{s^2} = \frac{4}{0.64}0.64 = 4 \in (0.71, \infty)$ pa prihvaćamo nul-hipotezu $H_0(\sigma^2 = 0.64)$.

Prisjetimo se motivacijskog primjera:

PRIMJER 15.11 *motiv*

Očekivana masa palete opeka je 1166.4 kg. Standardna devijacija mase palete opeka je 5.5 kg. Kontrolor je testirao uzorak od 20 paleta i dobio uzoračku standardnu devijaciju 6.5 kg. Može li kontrolor zaključiti da je standardna devijacija u porastu uz nivo značajnosti 1%?

Rješenje :
U zadatku su zadani $n = 20$, $\widehat{\sigma} = 6.5$, i nivo značajnosti $\alpha = 0.01$.
Ako se testira:
nul-hipoteza $H_0(\sigma = 5.5)$ i alternativna $H_1(\sigma > 5.5)$,
pomoću jednostranog testa, izabiremo test statistiku

$$T = \frac{n}{\sigma_0^2}\widehat{\Sigma}^2$$

koja ima hikvadrat distribuciju s $(n-1)$ stupnjeva slobode $T \sim \chi^2(19)$.
Kritičnu točku t_{kr} određujemo iz uvjeta $P(T > t_{kr}) = \alpha$.
$t_{kr} = z_{1-\alpha}$, je kvantil za funkciju distribucije $\chi^2(n-1)$.
Za nivo značajnosti $\alpha = 0.01$ očitamo iz tablice $\chi^2(19)$ vrijednost $t_{kr} = z_{0.99} = 36.19$.
Područje prihvaćanja za nul-hipotezu $H_0(\sigma = 5.5)$ je $(-\infty, 36.19)$.
Vrijednost test statistike $T(X_1, X_2, ..., X_n)$,
$t = \frac{n}{\sigma_0^2}\widehat{\sigma}^2 = \frac{20}{5.5^2} \cdot 6.5^2 = 34.35 \in (-\infty, 36.19)$, pa prihvaćamo o nul-hipotezu $H_0(\sigma = 5.5)$.
Kontrolor će zaključiti da se standardna devijacija nije povećala uz nivo značajnosti 1%.

15.5 Ponovimo

TEST HIPOTEZA za parametar t

slučajni uzorak	$(X_1, X_2, ..., X_n)$
parametar	t
nivo (razina) značajnosti	$\alpha = 0.01$ ili 0.05
nul-hipoteza	$H_0(t = t_0)$
alternativna hipoteze	$H_0(t \neq t_0)$, $H_0(t < t_0)$,$H_0(t > t_0)$
test statistika	T uz H_0
dvostrani test za t	$P(T < t_{kr1}) + P(T > t_{kr2}) = \alpha$
	$t_{kr1} = z_{\frac{\alpha}{2}}$ $t_{kr2} = z_{1-\frac{\alpha}{2}}$
prihvat za H_0	$t \in (t_{kr1}, t_{kr2})$
jednostrani test za t	$P(T < t_{kr}) = \alpha$
	$t_{kr} = z_{\alpha}$
prihvat za H_0	$t \in (t_{kr}, \infty)$
jednostrani test za t	$P(T > t_{kr}) = \alpha$
	$t_{kr} = z_{1-\alpha}$
prihvat za H_0	$t \in (-\infty, t_{kr})$

TEST HIPOTEZA za p $(n > 30)$

slučajni uzorak	$(X_1, X_2, ..., X_n)$ $n > 30$ $B(n, p)$
parametar p,	
nivo (razina) značajnosti	$\alpha = 0.01$ ili 0.05
nul-hipoteza	$H_0(p = p_0)$
alternativna hipoteze	$H_0(p \neq p_0)$, $H_0(p < p_0)$,$H_0(p > p_0)$
test statistika	$T = \sqrt{n}\frac{\overline{X}-p_0}{\sqrt{p_0\cdot(1-p_0)}} \sim N(0,1)$

TEST HIPOTEZA za μ $(n > 30)$

slučajni uzorak	$(X_1, X_2, ..., X_n)$ $n > 30$ ili$N(\mu, \sigma^2)$
parametar μ,	σ^2 poznat
nivo (razina) značajnosti	$\alpha = 0.01$ ili 0.05
nul-hipoteza	$H_0(\mu = \mu_0)$
alternativna hipoteze	$H_0(\mu \neq \mu_0)$, $H_0(\mu < \mu_0)$,$H_0(\mu > \mu_0)$
test statistika	$T = \sqrt{n}\frac{\overline{X}-\mu_0}{\sigma} \sim N(0,1)$

TEST HIPOTEZA za μ (normalne distribucije)

slučajni uzorak	$(X_1, X_2, ..., X_n)$ $n < 30$ ili $N(\mu, \sigma^2)$
parametar μ,	σ^2 nepoznat
nivo (razina) značajnosti	$\alpha = 0.01$ ili 0.05
nul-hipoteza	$H_0(\mu = \mu_0)$
alternativna hipoteze	$H_0(\mu \neq \mu_0)$, $H_0(\mu < \mu_0)$,$H_0(\mu > \mu_0)$
test statistika	$T = \sqrt{n}\frac{\overline{X}-\mu_0}{\widehat{S}} \sim t(n-1)$

TEST HIPOTEZA za σ^2

slučajni uzorak	$(X_1, X_2, ..., X_n)$ $n < 30$ ili $N(\mu, \sigma^2)$
parametar σ^2,	
nivo (razina) značajnosti	$\alpha = 0.01$ ili 0.05
nul-hipoteza	$H_0(\sigma^2 = \sigma_0^2)$
alternativna hipoteze	$H_0(\sigma^2 \neq \sigma_0^2)$, $H_0(\sigma^2 < \_0)$,$H_0(\sigma^2 > \sigma_0^2)$
test statistika	$T = \sqrt{n-1}\frac{1}{\sigma_0}\widehat{S}^2 \sim \chi^2(n-1)$

Bibliografija

[1] V. Devidé, Riješeni zadaci iz više matematike, Školska knjiga, Zagreb, 1972.

[2] M. Ilijašević, Ž. Pauše, Riješeni primjeri i zadaci iz vjerojatnosti i statistike, Zagreb-poduzeće za grafičku djelatnost, Samobor, 1990.

[3] E. Kreyszig, Advanced engineering mathematics, John Wiley & sons, Inc., New York Chichester Brisbane Toronto Singapore, 1993.

[4] V. M. Milošević, Teorijska statistika, Naučna knjiga, Beograd, 1990.

[5] B. Ostle, Statistics in research, The Iowa tate University Press, 1963.

[6] Ž. Pauše, Vjerojatnost-informacija-stohastički procesi, Školska knjiga, Zagreb, 1988.

[7] Ž. Pauše, Uvod u matematičku statistiku, Školska knjiga, Zagreb, 1993.

[8] N. Sarapa, Vjerojatnost, Školska knjiga, Zagreb, 1987.

[9] M. R. Spiegel, Shaum's outline of Theory and problems of Probability and statistics, McGRAW-Hill Book Company, New York St. Louis San Francisco Toronto Sydney, 1990.

[10] D. Veljan, Kombinatorika s teorijom grafova, Školska knjiga, Zagreb, 1989.

[11] S. V. Vukadinović, Elementi teorije verovatnoće i matematičke statistike, Privredni pregled, Beograd, 1988.