1 Slučajni opit

Eksperiment-opit je moćan metod za spoznaju stvarnosti. Opitom se ispituje odnos između uslova-uzroka i posljedice. Deterministički opit-isti uslovi uvijek dovode do istog rezultata. Kod slučajnog opita ostvarivanje određenih uslova ne dovodi do determinističkog rezultata. Teorija vjerovatnoće izučava matematički model slučajnog opita.

Primjer 1.1 *Novčić se baca jednom.* **Ishodi-elementarni događaji** *koji se mogu desiti u ovom opitu su: palo je pismo P; pao je grb G.* **Prostor ishoda**, *koristi se oznaka Ω je skup skup čiji su elementi ishodi koji se mogu desiti u opitu. U našem slučaju $\Omega = \{P, G\}$.*

Primjer 1.2 *Novčić se baca četiri puta.* $\Omega = \{PPPP, PPPG, ..., GGGG\}$.

Primjer 1.3 *Kocka se baca jednom.* $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Primjer 1.4 *Kocka se baca dva puta.* $\Omega = \{(1, 1), (1, 2), ..., (6, 6)\}$.

Primjer 1.5 *Novčić se baca do prvog padanja pisma.* $\Omega = \{P, GP, GGP, ...\}$.

Primjer 1.6 *Strijelac gađa u kružnu metu.*

Primjer 1.7 *Braunovo kretanje.*

Rezultati slučajnih opita se nazivaju **slučajnim događajima**, ubuduće ćemo govoriti o događajima. Realizacija događaja, što je prirodno, je ekvivalentna sa realizacijom nekog ishoda koji događaju odgovara te se događaj identifikuje sa skupom ishoda koji tom događaju odgovaraju. Po završenom opitu, ako se ostvari ishod koji odgovara događju, konstatujemo da se događaj realizovao, a u suprotnom da se nije realizovao. U opitu u kome dva puta bacamo kocku možemo govoriti o događaju A da je zbir palih brojeva 5. Skup ishoda koji odgovaraju događaju A ćemo takođe označiti sa A i $A = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$.

Neka su A i B skupovi-događaji iz Ω.

$A \bigcup B = \{\omega : \omega \in A \text{ ili } \omega \in B\}$. Iz ovog zapisa zaključujemo: $A \bigcup B$ je događaj koji se realizuje ako i samo ako se realizuje ili događaj A ili događaj B (ili ekvivalentno rečenom, $A \bigcup B$ je događaj koji se realizuje ako i samo ako se realizuje bar jedan od događaja A ili B). Dakle, $A \bigcup B$ je događaj da se realizuje ili događaj A ili događaj B ($A \bigcup B$ je događaj da se realizuje bar jedan od događaja A ili B).

$A \bigcap B = \{\omega : \omega \in A \text{ i } \omega \in B\}$. Iz ovog zapisa zaključujemo: $A \bigcap B$ je događaj koji se realizuje ako i samo ako se realizuje i događaj A i događaj B (ili ekvivalentno rečenom, $A \bigcap B$ je događaj koji se realizuje ako i samo ako se realizuju oba događaja A i B). Dakle, $A \bigcap B$ je događaj da se realizuje i događaj A i događaj B ($A \bigcap B$ je događaj da se realizuju oba događaja A i B). Ako je $AB = \emptyset$, tada kažemo da se događaji A i B uzajamno isključuju.

$A \backslash B = \{\omega : \omega \in A, \omega \notin B\}$ je događaj da se realizuje događaj A i ne realizuje događaj B.

$A^c = \{\omega : \omega \notin A\}$ je događaj (događaj suprotan događaju A) da se u opitu ne realizuje događaj A.

Ako je $A \subset B$ tada kažemo da događaj A implicira (povlači) događaj B. Naime, iz realizacije događaja A slijedi da je realizovan neki ishod iz A, a zbog $A \subset B$, taj ishod je iz B. Zaključujemo, realizovan je događaj B. Relacija $A \subset B$ se može interpretirati i na sljedeći način: događaj A se ne može ostvariti ako se ne ostvari događaj B. Osoba ne može biti majka ako nije žena.

Ω je izvjestan događaj. $\emptyset$ je nemoguć događaj.

Primjer 1.8 *U opitu u kome se kocka baca dva puta A je događaj da je zbir palih brojeva $\leqslant 3$, B je događaj da u drugom bacanju padne paran broj, C je događaj da u drugom bacanju padne šestica. Naći $A \bigcup B, A \bigcap B, A^c, B \backslash A$ i dokazati da je $C \subseteq B$.*

Neke od podskupova prostora ishoda Ω ćemo nazivati događajima. Prirodno je zahtijevati da kolekciji događaja pripada izvjestan događaj Ω, nemoguć događaj $\emptyset$ i da je ta kolekcija zatvorena u odnosu na skupovne operacije. Kolekciju događaja ćemo označavati sa $\mathfrak{F}$.

Definicija 1.1 *Kolekcija događaja $\mathfrak{F}$ je polje ako važi:*

(A1) $\Omega \in \mathfrak{F}$.

(A2) Ako $A \in \mathfrak{F}$ *tada* $A^c \in \mathfrak{F}$.

(A3) Ako $A \in \mathfrak{F}$ *i* $B \in \mathfrak{F}$ *tada* $A \bigcup B \in \mathfrak{F}$.

Definicija 1.2 *Kolekcija događaja $\mathfrak{F}$ je σ polje ako važe (A1), (A2), (A3) i*

(A4) Ako $A_1 \in \mathfrak{F}$, $A_2 \in \mathfrak{F}, \ldots$ *tada* $\displaystyle\bigcup_{i=1}^{\infty} A_i \in \mathfrak{F}$.

σ polje je zatvoreno u odnosu na prebrojivo presjecanje. Zaista, ako $A_1 \in \mathfrak{F}$, $A_2 \in \mathfrak{F}, \ldots$ tada je $\displaystyle\bigcap_{i=1}^{\infty} A_i = \Big(\bigcup_{i=1}^{\infty} A_i^c\Big)^c \in \mathfrak{F}$. Lako se dokazuje da je σ polje zatvoreno u odnosu na operaciju razlike skupova. Dokazuje se da je presjek σ polja takođe σ polje.

Primjer 1.9 $\Omega = \{1, 2, 3, ...\}$ *i kolekciju* $\mathfrak{F}$ *čine skupovi iz* Ω *koji su konačni ili su njihovi komplementi konačni. Kolekcija* $\mathfrak{F}$ *je polje koje nije* σ *polje. Ako kolekciju* $\mathfrak{W}$ *čine skupovi koji su prebrojivi ili su njihovi komplementi prebrojivi, tada je kolekcija* $\mathfrak{F}$ σ *polje.*

2 Vjerovatnosni prostor

Definicija 2.1 *Neka je* $\mathfrak{F}$ σ*-polje na prostoru ishoda* Ω*. Uređeni par* $(\Omega, \mathfrak{F})$ *se naziva* **mjerljivi prostor**.

Definicija 2.2 *Neka je* $(\Omega, \mathfrak{F})$ *mjerljivi prostor. Funkcija* $P : \mathfrak{F} \longrightarrow R$ *je* **vjerovatnoća** na $\mathfrak{F}$ *ako važi:*

P1. $P(A) \geqslant 0$ *za svaki* $A \in \mathfrak{F}$; $P(\Omega) = 1$.

P2. Ako su $A_i \in \mathfrak{F}, i \in N$ *i* $A_i \cap A_j = \emptyset$, $i \neq j$ *tada je* $P\left(\sum\limits_{i=1}^{\infty} A_i\right) = \sum\limits_{i=1}^{\infty} P(A_i)$.

Definicija 2.3 *Uređena trojka* $(\Omega, \mathfrak{F}, P)$*, gdje je* $\mathfrak{F}$ σ*-polje na* Ω *i* P *vjerovatnoća na* $\mathfrak{F}$*, naziva se* **vjerovatnosni prostor**.

Vjerovatnosni prostor je osnovni objekat u Teoriji vjerovatnoća. Svojstvo $P(A) \geqslant 0, A \in \mathfrak{F}$, je svojstvo nenegativnosti vjerovatnoće, a svojstvo $P(\Omega) = 1$ je svojstvo normiranosti vjerovatnoće. Svojstvo iz aksiome P2 je svojstvo σ–aditivnosti vjerovatnoće.

Neka je $(\Omega, \mathfrak{F}, P)$ vjerovatnosni prostor. Elementi σ–polja $\mathfrak{F}$ su **događaji**, a broj $P(A)$, $A \in \mathfrak{F}$, se naziva **vjerovatnoća događaja** A.

U sljedećoj teoremi će biti evidentirana neka značajna svojstva vjerovatnoće.

Teorema 2.1 *Neka je* $(\Omega, \mathfrak{F}, P)$ *vjerovatnosni prostor. Tada vrijedi:*
(a) $P(\emptyset) = 0$.
(b) Ako su $A_1, \ldots, A_n$ *događaji tada je*
$P\left(\sum\limits_{i=1}^{n} A_i\right) = \sum\limits_{i=1}^{n} P(A_i)$
(svojstvo **konačne aditivnosti vjerovatnoće**).
(c) Ako su A *i* B *događaji,* $A \subseteq B$ *tada je* $P(A) \leqslant P(B)$
(svojstvo **monotonosti vjerovatnoće**).
(d) Ako je A *događaj tada je* $0 \leqslant P(A) \leqslant 1$.
(e) Ako je A *događaj tada je* $P(A^c) = 1 - P(A)$
(**vjerovatnoća suprotnog događaja** A^c).
(f) Ako su A *i* B *događaji tada je* $P(A \cup B) = P(A) + P(B) - P(AB)$.

♦ (a) U aksiomi P2 stavimo $A_1 = \Omega$ i $A_i = \emptyset$, $i \geqslant 2$. Dobijamo

$$1 = 1 + P(\emptyset) + P(\emptyset) + \ldots \Longrightarrow P(\emptyset) = 0$$

(b) U P2 stavimo $A_i = \emptyset$, $i > n$ i iskoristimo (a).
(c) $A \subseteq B \Longrightarrow B = A \cup (B \setminus A)$. Zbog (b) je

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

(d) Za svaki događaj A važi: $\emptyset \subseteq A \subseteq \Omega$ odakle iz monotonosti vjerovatnoće slijedi tvrđenje.
(e) $A + A^c = \Omega \Longrightarrow P(A) + P(A^c) = 1$.
(f) Imamo $A \cup B = A + (B \setminus A)$, $(A \cap B) + (B \setminus A) = B$. Odavde, zbog (b), dobijamo

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

Nakon sabiranja jednakosti i sređivanja dobijamo tvrđenje.♦

Teorema 2.2 Silvesterova formula. *Neka je $(\Omega, \mathfrak{F}, P)$ vjerovatnosni prostor i $A_i \in \mathfrak{F}$, $i = 1, 2, \ldots, n$. Tada važi*

$$\begin{aligned} P(\bigcup_{i=1}^{n} A_i) &= \sum_{1 \leqslant i \leqslant n} P(A_i) - \sum_{1 \leqslant i < j \leqslant n} P(A_i A_j) \\ &+ \sum_{1 \leqslant i < j < k \leqslant n} P(A_i A_j A_k) - \cdots + (-1)^{n+1} P(\bigcap_{i=1}^{n} A_i). \end{aligned} \tag{2.1}$$

♦ Dokaz ćemo sprovesti indukcijom po n. Za $n = 1$ tvrđenje je očigledno, a za $n = 2$ to je tvrđenje (f) iz Teoreme 1. Pretpostavimo da tvrđenje vrijedi za sve familije od najviše $n-1$ događaja. Stavimo

$$B = \bigcup_{i=1}^{n-1} A_i,\ C_i = A_i A_n,\ i < n.$$

Tada vrijedi

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P(B \cup A_n) = P(B) + P(A_n) - P(BA_n). \tag{2.2}$$

Osim toga je $BA_n = \bigcup\limits_{i=1}^{n-1} C_i$ pa zbog induktivne pretpostavke imamo

$$P(B) = \sum_{1 \leqslant i < n} P(A_i) - \sum_{1 \leqslant i < j < n} P(A_i A_j) + \ldots + (-1)^n P(\bigcap_{i=1}^{n-1} A_i) \tag{2.3}$$

i

$$
\begin{aligned}
P(BA_n) &= \sum_{1\leqslant i<n} P(C_i) - \sum_{1\leqslant i<j<n} P(C_iC_j) + ... + (-1)^n P\left(\bigcap_{i=1}^{n-1} C_i\right) \\
&= \sum_{1\leqslant i<n} P(A_iA_n) - \sum_{1\leqslant i<j<n} P(A_iA_jA_n) + ... + (-1)^n P(\bigcap_{i=1}^{n} A_i). \qquad (2.4)
\end{aligned}
$$

Iz (2.2),(2.3),(2.4) slijedi (2.1).

Primijetimo da u desnoj strani jednakosti (2.1) prva suma ima $\binom{n}{1}$ članova, druga $\binom{n}{2}$ članova,..., a n–ta $\binom{n}{n}$ član. ♦

Razmotrićemo model u kojem je skup $\Omega = \{\omega_1, \omega_2, ...\}$ konačan ili prebrojiv, a σ polje događaja $\mathfrak{F}$ je $\mathbb{P}(\Omega)$. Svakom ishodu ω_i, $i \in I$, pridružuje se broj $p(\omega_i), i \in I$ tako da važi:

a) $0 \leqslant p(\omega_i) \leqslant 1, i \in I$,

b) $\sum_{i\in I} p(\omega_i) = 1$.

Neka je

$$P : \mathfrak{F} \to [0,1], P(A) = \sum_{i:\omega_i\in A} p(\omega_i), A \in \mathfrak{F}.$$

Lako se dokazuje da je trojka $(\Omega, \mathfrak{F}, P)$ vjerovatnosni prostor i naziva se **diskretni vjerovatnosni prostor**.

U praksi se često susreću opiti sa konačno mnogo ravnopravnih-simetričnih ishoda. Zbog ravnopravnosti ishoda u slučaju kada je $\Omega = \{\omega_1, \omega_2, ..., \omega_N\}$ razumno je svakom ishodu pridružiti vjerovatnoću $\frac{1}{N}$ tj. $p(\omega_1) = ... = p(\omega_1) = \frac{1}{N}$ odakle zaključujemo da je za svako $A \in \mathfrak{F}$

$$P(A) = \frac{\mid A \mid}{\mid N \mid} \qquad (2.5)$$

Dakle vjerovatnoća događaja A je količnik broja ishoda koji odgovaraju događaju A i broja svih mogućih ishoda. Formulom (2.5) je data tzv. **klasična (ili Laplasova) definicija vjerovatnoće**.

U sljedeća dva primjera ishodi nisu ravnopravni.

Primjer 2.1 *Novčić se baca do prvog padanja grba ali najviše dva puta.*

$\Omega = \{P, GP, GG\}; P(P) = 0,5; P(GP) = P(GG) = 0,25.$

Primjer 2.2 *Novčić se baca do prvog padanja pisma.* $\Omega = \{P, GP, GGP, ...\}, P(P) = 0,5; P(GP) = 0,25; P(GGP) = 0,125, ...$

Zapaža se da se u dugim serijama opita relativna učestalost događaja ponaša stabilno. Ta stabilnost učestalosti ukazuje na to da se objektivno prisutna slučajnost realizacije događaja može kvantitativno mjeriti. Matematička formalizacija gore pomenutog zapažanja se ostvaruje kroz zakon velikih brojeva. Teorija vjerovatnoće se ne bavi zadatkom pravilnog zadavanja vjerovatnoća $p_i, i = 1, 2, ..., N$. Pri zadavanju ovih vjerovatnoća vodi se računa o intuitivnoj predstavi broja p_i kao relativnoj učestalosti ishoda ω_i u dugoj seriji opita. Jedan od glavnih vjerovatnosnih zadataka je da na osnovu vjerovatnoća ishoda tražimo vjerovatnoće složenih događaja.

Primjer 2.3 *Komplet u kojem je 36 karata se na slučajan način dijeli na dva jednakobrojna dijela. Kolika je vjerovatnoća da se u oba dijela nalazi po 9 crnih i crvenih karata?*

►Broj podjela je određen izborom 18 karata iz kompleta od 36 karata, a tih izbora ima $\binom{36}{18}$. Iz skupa od 18 crnih karata, 9 karata možemo izabrati na $\binom{18}{9}$ načina. Isti rezon primjenjujemo i za crvene karte. Koristeći princip proizvoda, dobijamo

$$p = \frac{\binom{18}{9}\binom{18}{9}}{\binom{36}{18}} \approx 0,26.$$

Prilikom računanja je korišćena Stirlingova formula

$$n! \sim \sqrt{2\pi n}\, n^n\, e^{-n}.$$

Navešćemo neke relativne frekvencije događaja čiju smo vjerovatnoću izračunali. Podaci potiču iz jedne serije od 100 ponavljanja opita – podjela karata. Nakon 5 podjela relativna frekvencija je bila 0, 4, nakon 10 podjela je bila 0, 3, nakon 26 podjela je bila 0, 23, nakon 50 podjela je bila 0, 24, a nakon svih 100 podjela relativna frekvencija je 0, 24. Znači, na početku serije relativna frekvencija ima veliku fluktuaciju, a zatim se sa rastom serije relativna frekvencija stabilizuje.◄

Primjer 2.4 *U kutiji se nalaze cedulje na kojima su brojevi od* 1 *do* 10. *Iz kutije se po modelu bez vraćanja vadi* 5 *cedulja. Kolika je vjerovatnoća da brojevi na izvađenim ceduljama obrazuju rastući niz? R:*$\frac{1}{5!}$.

Primjer 2.5 *U kutiji se nalazi b bijelih i c crnih kuglica. Dva igrača jedan za drugim vade kuglicu i pobjeđuje onaj koji prvi izvuče bijelu kuglicu. Kolika je vjerovatnoća pobjede igrača koi započinje igru? Uraditi oba modela.*

a) Model sa vraćanjem. $\frac{b+c}{b+2c}$. b) Model bez vraćanja.

$$c \text{ parno}, \quad p = b\sum_{l=0}^{c} \frac{c^{[l]}}{(b+c)^{[l+1]}}, n^{[k]} = n(n-1)...(n-k+1), \text{sumira se po parnim indeksima,}$$
$$c \text{ neparno}, \quad p = b\sum_{l=0}^{c-1} \frac{c^{[l]}}{(b+c)^{[l+1]}}, \text{sumira se po parnim indeksima.}$$

Primjer 2.6 *Izračunati vjerovatnoću da za 30 osoba između 12 mjeseci u godini , 6 mjeseci sadrži po 2, a ostalih 6 mjeseci po 3 njihova rođendana. R:* $p = \binom{12}{6}\frac{30!}{2^6 6^6 12^{30}} \approx 0,00035$.

Primjer 2.7 *r kuglica se razmješta u n kutija. Izračunati vjerovatnoću da*

a) u svakoj od prvih r kutija bude tačno jedna kuglica ($n \geqslant r$), R: $p = \frac{r!}{n^r}$.

b) u prvoj kutiji bude tačno r_1 kuglica, u drugoj kutiji bude tačno r_2 kuglica,..., u n-toj kutiji bude tačno r_n kuglica, $\sum\limits_{i=1}^{n} r_i = r$, R: $p = \frac{r!}{n^r \prod\limits_{i=1}^{n} r_i!}$.

c) u jednoj od kutija bude r_1, u nekoj drugoj r_2 itd pri čemu su brojevi $r_1, r_2, ..., r_n$ među sobom različiti, $\sum\limits_{i=1}^{n} r_i = r$. R: $p = \frac{r!n!}{n^r \prod\limits_{i=1}^{n} r_i!}$.

Primjer 2.8 *U kutiji se nalazi $n-1$-na bijela i jedna crvena kuglica. Iz kutije se po modelu bez vraćanja vadi $k, 1 \leqslant k \leqslant n$ kuglica. Kolika je vjerovatnoća da među njima bude crvena?*

$$p = \frac{\binom{1}{1}\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{k}{n}.$$

Primjer 2.9 *m muškaraca i n žena sjedaju u red. Kolika je vjerovatnoća da sve žene sjede jedna pored druge?*

$$p = \frac{m+1}{\binom{m+n}{n}}.$$

Primjer 2.10 *Na četiri strane kocke A utisnute su po 2 tačke, a na dvije strane utisnuto je po 5 tačaka. Na svim stranama kocke B utisnute su po 3 tačke. Na četiri strane kocke C utisnute su po 4 tačke, a na dvije po jedna tačka. Biraju se dvije kocke, istovremeno se bacaju i "pobjeđuje" ona kocka na kojoj padne više tačaka (padne veći broj).*

► Analizirajmo varijante. Ako igraju kocke A i B, B pobjeđuje ako na A padne 2. Dakle, pobjedu kocke B povlače parovi 2 na A i 6 na B, a tih parova ima $4 \cdot 6 = 24$. Kako mogućih ishoda ima

$6 \cdot 6 = 36$ (6 je broj strana i na A i na B), to je vjerovatnoća pobjede kocke B, $p_B = \frac{24}{36} = \frac{2}{3}$, a vjerovatnoća pobjede kocke A, $p_A = \frac{1}{3}$. Ako igraju kocke B i C, vjerovatnoća pobjede kocke C je $p_C = \frac{2}{3}$ (kocka C pobjeđuje ako na njoj padne 4), dok je $p_B = \frac{1}{3}$. Ako igraju A i C, imamo $p_A = \frac{4\cdot2+2\cdot6}{36} = \frac{5}{9}$ (A pobjeđuje u varijantama $1^0 : 2$ na A, 1 na C, $2^0 :\ 5$ na A) i $p_C = \frac{4}{9}$.

Primijetimo, u igri kocaka A i B, bolja je kocka B – ima veću šansu (vjerovatnoću) na pobjedu; u igri B i C, bolja je kocka C; u igri A i C, bolja je kocka A. Dakle, relacija "biti bolji (bolja)" u našem slučaju nije tranzitivna.

Ako svaka od dvije osobe odabere po jednu kocku sa namjerom da otpočnu opisanu igru, u boljoj je poziciji osoba koja kocku bira nakon što je svoju kocku već odabrala prva osoba. Naime, ako je prva osoba odabrala A, druga će odabrati B; ako je prva odabrala B, druga će odabrati C; ako je prva odabrala C, druga će odabrati A. Prirodno, druga osoba uvijek bira kocku za koju je vjerovatnoća pobjede veća.◄

Primjer 2.11 *Iz skupa od n osoba koje sjede za okruglim stolom, na slučcajan način bira se k osoba, $k \leqslant \frac{n}{2}$. Kolika je vjerovatnoća da nikoje dvije izabrane osobe ne sjede jedna do druge?*

►Numerišimo stolice sa $1, 2, ..., n$, koristimo numeraciju koja prati kretanje kazaljke na satu, a zatim sto "presjecanjem" između stolica n i 1 pretvorimo u pravougaoni (stolice $1, 2, .., n$ se nalaze na jednoj strani pravougaonog stola). Tokom analize, stolice na kojima sjede odabrane osobe ćemo obilježiti sa a, a neodabrane sa b. U procesu računanja broja povoljnih ishoda razmotrimo dvije varijante.

Prva. Na stolici sa rednim brojem 1 sjedi neka od k odabranih osoba. Na stolicama 2 i n moraju sjedjeti neodabrane osobe. Isključimo stolicu n i stolice na kojima sjede odabrane osobe. Ostala je $n - k - 1$ stolica na kojima sjede neodabrane osobe. Da bi rekonstruisali raspored koji nam je generisao stolice koje su ostale, treba odabrati $k - 1$ od preostalih stolica i na prvoj desnoj poziciji smjestiti odabranu. Recimo, $n = 8, k = 3$. Ostale su 4 stolice, dakle rekonstrukcija kreće od *bbbb*. Imajući u vidu da je u ovoj varijanti poznato da je stolica 1 obilježena sa a, a stolice 2 i 8 sa b, jedan od rekonstrukcija je recimo *abbabbab*. Mi smo iz četvorke *bbbb* odabrali prvi i četvrti član.

Druga. Razmatra se slučaj kada na stolici 1 sjedi neodabrana osoba.

$$p = \frac{\binom{n-k-1}{k-1} + \binom{n-k}{k}}{\binom{n}{k}} \quad ◄$$

Primjer 2.12 *Za okrugli sto sijedaju osobe iz grupe od n bračnih parova. Kolika je vjerovatnoća da svaka od $k, 1 \leqslant k \leqslant n$, fiksiranih žena sjedi pored svog supruga?*

▶Izdvojićemo jednu od k fiksiranih žena i od njene stolice tretirati permutaciju. Suprug izdvojene žene može zauzeti jednu od dvije susjedne pozicije. Svaki par stolica na kojima sjede supružnici iz preostalih $k-1$ parova (parova koje formiraju preostalih $k-1$ fiksiranih žena) tretiraćemo kao jedan element tj. blok. $k-1$ blok i $2n-2k$ osoba koje nisu članovi k izdvojenih parova se mogu poređati na $(k-1+2n-2k)! = (2n-k-1)!$ načina. U svakom bloku svoje stolice supružnici mogu odabrati na 2 načina. Dakle,

$$p_k = \frac{2^k(2n-k-1)!}{(2n-1)!}. \blacktriangleleft$$

Primjer 2.13 *n nula i n jedinica formira slučajan niz. Kolika je vjerovatnoća da su parovi u nizu napravljeni od 0 i 1?*

▶Primjeri nizova: U nizu 00101101 parovi su $00; 10; 11; 01$. U nizu 01101010 parovi su $01; 10; 10; 10$ te niz odgovara uslovu.

Niz od n nula i n jedinica je određen pozicijama koje su zauzele recimo nule (ostale pozicije zauzimaju jedinice). Broj načina da se od $2n$ pozicija odabere n za nule je $\binom{2n}{n}$. Niz koji nama odgovara je onaj kod koga u svakom paru imamo nulu. Nula se u svakom paru može naći na jednoj od dvije pozicije u paru. Dakle $p = \frac{2^n}{\binom{2n}{n}}$.

Ekvivalentan je sljedeći zadatak.

n osoba je izulo svoje cipele, a zatim je sa police na kojima su cipele bile ostavljene, svaka osoba nasumce uzela po dvije cipele. Kolika je vjerovatnoća da je svaka osoba uzela po jednu lijevu i jednu desnu cipelu?◀

Primjer 2.14 *Na dva polja šahovske table su postavljene dame. Kolika je vjerovatnoća da se dame napadaju?*

▶Dame će se napadati ako su na istoj horizontali ili istoj vertikali ili istoj dijagonali. Ima po 8 horizontala i vertikala, po 4 dijagonale dužine $2, 3, 4, 5, 6, 7$ i dvije dijagonale dužine 8. Dakle

$$p = \frac{8\binom{8}{2} + 8\binom{8}{2} + 2\binom{8}{2} + 4\binom{2}{2} + 4\binom{3}{2} + 4\binom{4}{2} + 4\binom{5}{2} + 4\binom{6}{2} + 4\binom{7}{2}}{\binom{64}{2}} = \frac{81}{224}. \blacktriangleleft$$

Primjer 2.15 (Zadatak o rasijanom dekanu.) *Na svečanoj dodjeli diploma bilo je n studenata i dekan im je nasumce podijelio diplome. Kolika je vjerovatnoća da je bar jedan student dobio svoju diplomu?*

▶Numerišimo studente brojevima od 1 do n. Označimo sa A_1 događaj da prvi student dobije svoju diplomu, sa A_2 događaj da drugi student dobije svoju diplomu,..., sa A_n događaj da n–ti student dobije svoju diplomu. Ako sa W označimo događaj čiju vjerovatnoću tražimo, tada je $W = A_1 \cup A_2 \cup \dots A_n$. Da bismo našli $P(W)$ primijenićemo Silvesterovu formulu. Nakon primjene ove formule, treba izračunati vjerovatnoće presjeka raznih događaja iz kolekcije $A_1, \dots, A_n$. Kako je ideologija u računanju ovih vjerovatnoća ista, demonstriraćemo račun, recimo na primjeru $P(A_1A_2)$. n diploma se n–torici studenata može podijeliti na $n!$ načina. Kada prvi i drugi student dobiju svoje diplome, preostalih $n-2$ diplome preostalim studentima (ima ih $n-2$) može biti podijeljeno na $(n-2)!$ načina. Dakle, $P(A_1, A_2) = \frac{(n-2)!}{n!}$. Imamo

$$P(W) \;=\; 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots + \frac{(-1)^{n-1}}{n!}.$$

Ako $n \to \infty$, dobijeni zbir teži ka $1 - e^{-1} \approx 0,6322$. Zbog brze konvergencije niza – zbira, u slučaju kada je $n \geqslant 20$, dobijena vjerovatnoća se odlično aproksimira sa $0,6322$. Iz istog razloga, rezultati se zanemarljivo razlikuju ako je, recimo, $n = 100$ ili $n = 200$.

Ovaj zadatak se može formulisati u mnogo varijanti. Navedimo jednu. Složili smo 100 listova, zaduvao je vjetar i razbacao listove. Zatim smo listove nasumce ponovo složili. Kolika je vjerovatnoća da će se bar jedan list naći na rednom mjestu na kojem je bio i ranije?◄

Primjer 2.16 *Kolika je vjerovatnoća da u seriji od 20 bacanja kocke ne budu zabilježeni svi brojevi?*

► A_1 u seriji nije zabilježena jedinica,..., A_6 u seriji nije zabilježena šestica.

$$P(\bigcup_{i=1}^{6} A_i) = 6\frac{5^{20}}{6^{20}} - \binom{6}{2}\frac{4^{20}}{6^{20}} + ... + \binom{6}{5}\frac{1}{6^{20}}. \blacktriangleleft$$

Primjer 2.17 *Iz grupe od* 8 *bračnih parova bira se* 6 *osoba. Kolika je vjerovatnoća da među izabranima ne postoji bračni par?*

►Numerišimo bračne parove sa $1, 2, ..., 8$. Neka je A_1 događaj da je među izabranim osobama bračni par 1 (tj. oba supružnika), A_2 bračni par 2,... Tražena vjerovatnoća je

$$p = 1 - P\Big(\bigcup_{i=1}^{8} A_i\Big) = 1 - \binom{8}{1}\frac{\binom{14}{4}}{\binom{16}{6}} + \binom{8}{2}\frac{\binom{14}{2}}{\binom{16}{6}} - \binom{8}{3}\frac{1}{\binom{16}{6}}. \blacktriangleleft$$

Primjer 2.18 *Neka je* $(\Omega, \mathfrak{F}, P)$ *vjerovatnosni prostor i* $A_i \in \mathfrak{F}$, $i = 1, 2, \dots, n$. *Tada važi*

$$\begin{aligned} P(\bigcap_{i=1}^{n} A_i) \;&=\; \sum_{1\leqslant i\leqslant n} P(A_i) - \sum_{1\leqslant i<j\leqslant n} P(A_i \cup A_j) \\ &\quad + \sum_{1\leqslant i<j<k\leqslant n} P(A_i \cup A_j \cup A_k) - \cdots + (-1)^{n+1} P(\bigcup_{i=1}^{n} A_i\;) \end{aligned}$$

U rješavanju nekih važnih vjerovatnosnih zadataka koristi se sljedeća dobro poznata teorema iz Kombinatorike (Teorema 1, zbirka str 37).

Teorema 2.3 *Neka je prostor ishoda Ω konačan skup, $\mathfrak{F} = \mathbb{P}(\Omega)$ i $A_1, A_2, ..., A_n$ su događaji. Tada je*

$$|A_1A_2..A_n| = |\Omega| - \sum_{1\leqslant i\leqslant n} |A_i^c| + \sum_{1\leqslant i<j\leqslant n} |A_i^cA_j^c| - ... + (-1)^{n-1}|A_1^c...A_n^c|.$$

Primjer 2.19 *Kocka se baca do pojave svih šest strana. Izračunati vjerovatnoću događaja da će biti izvedeno tačno $n, n \geqslant 6$, bacanja.*

► Ω je skup n torki čiji su elementi iz $S = \{1,2,3,4,5,6\}, |\Omega| = 6^n$. Odgovaraju n torke koje se završavaju sa nekim brojem iz S (6 mogućnosti), a na prvih $n-1$ mjesta se nalaze **svi** brojevi iz S sa izuzetkom broja koji je na n tom mjestu. $n-1$ torki napravljenih od 5 brojeva (svi učestvuju) ima

$$5^{n-1} - \binom{5}{1}4^{n-1} + \binom{5}{2}3^{n-1} - \binom{5}{3}2^{n-1} + \binom{5}{4}.$$

(vidi Teoremu (prilagođavamo je slučaju kada se kocka baca $n-1$ puta), ako je na n tom mjestu recimo 6, tada $n-1$ torke pravimo sa brojevima $1,2,3,4,5$; skup - događaj A_1 - učestvuje 1 u $n-1$ torki, skup - događaj A_2 - učestvuje 2 u $n-1$ torki itd). Dakle,

$$\begin{aligned} p &= \frac{6\left[5^{n-1} - \binom{5}{1}4^{n-1} + \binom{5}{2}3^{n-1} - \binom{5}{3}2^{n-1} + \binom{5}{4}\right]}{6^n} \\ &= \left(\frac{5}{6}\right)^{n-1} - 5\left(\frac{4}{6}\right)^{n-1} + 10\left(\frac{3}{6}\right)^{n-1} - 10\left(\frac{2}{6}\right)^{n-1} + 5\left(\frac{1}{6}\right)^{n-1}. \blacktriangleleft \end{aligned}$$

Primjer 2.20 *r kuglica se razmješta u n kutija, $r \geqslant n$. Kolika je vjerovatnoća da tačno $m, 0 \leqslant m \leqslant n-1$ kutija bude prazno?*

►Zapisujemo redni broj kutije u koju ide kuglica $1,2,...,r$ i dobijamo r torku čiji su članovi iz $S = \{1,2,...,n\}, |\Omega| = n^r$. Odgovaraju sve r torke napravljene od $n-m$ brojeva iz S (svi ti brojevi učestvuju). $n-m$ brojeva iz S možemo izabrati na $\binom{n}{m}$ načina. Izabrane brojeve označimo sa $s_1, s_2, ..., s_{n-m}$. Izračunajmo koliko ima r torki koje nam odgovaraju. Neka je A_1 događaj-skup s_1 učetvuje; ..., A_{n-m} događaj-skup s_{n-m} učetvuje. Sada je

$$\begin{aligned} |A_1...A_{n-m}| &= (n-m)^r - \binom{n-m}{1}(n-m-1)^r + \binom{n-m}{2}(n-m-2)^r... \\ &\quad +(-1)^{n-m-1}\left[n-m-(n-m-1)\right]^r \end{aligned}$$

te je

$$p = \frac{\binom{n}{m}|A_1...A_{n-m}|}{n^r} = \binom{n}{m}\sum_{\nu=o}^{n-m-1}(-1)^{\nu}\binom{n-m}{\nu}\left(1-\frac{m+\nu}{n}\right)^r. \blacktriangleleft$$

Primjer 2.21 *Kocka se baca 7 puta. Kolika je vjerovatnoća da je suma palih brojeva 27?*

► $x_1 + ... + x_7 = 27, 1 \leqslant x_i \leqslant 6$. $y_i = x_i - 1; y_1 + ... + y_7 = 20, 0 \leqslant y_i \leqslant 5$.

$y_1 + ... + y_7 = 20, 0 \leqslant y_i$ ima $\binom{26}{6}$ rješenja. Koristićemomo formulu uključenja isključenja. Tražimo koliko ima rješenja takvih da je jedna nepoznata $\geqslant 6$, $z_7 = y_7 - 6; y_1 + ... + y_6 + z_7 = 14$ nenegativnih rješenja ima $\binom{20}{6}$. Dvije nepoznate $\geqslant 6$, $y_1 + ... + y_5 + z_6 + z_7 = 8$ ima ih $\binom{14}{6}$ i na kraju tri nepoznate $\geqslant 6$, $y_1 + ...y_4 + z_5 + z_6 + z_7 = 2$ ima ih $\binom{8}{6}$. Dakle,

$$p = \frac{1}{6^7}\left[\binom{26}{6} - \binom{20}{6}\binom{7}{1} + \binom{14}{6}\binom{7}{2} - \binom{8}{6}\binom{7}{3}\right]. ◄$$

3 Uslovna vjerovatnoća i nezavisnost događaja

Neka je A događaj iz slučajnog opita koji je matematički modeliran trojkom $(\Omega, \mathcal{F}, P)$. Ako osim kompleksa uslova koji generišu opit nema drugih ograničenja koja mogu uticati na računanje vjerovatnoće događaja A, tada kažemo da je $P(A)$ bezuslovna vjerovatnoća događaja A.

Međutim, praksa često nameće potrebu za računanjem vjerovatnoće događaja A pri dopunskom uslovu da je ostvaren neki događaj B. Takve vjerovatnoće se zovu **uslovne**. Koristićemo zapis $P(A|B)$; njime je označena vjerovatnoća događaja A pri uslovu B (ili preciznije, pri uslovu da je ostvaren događaj B).

Da bismo približili motive za definiciju kroz koju se ukazuje na postupak računanja uslovne vjerovatnoće, navešćemo dva primjera.

Primjer 3.1 *Kocka za igru se baca dva puta. Neka je A događaj da je u oba bacanja pao paran broj, a B događaj da je zbir brojeva u oba bacanja $\leqslant 6$.*

►Konstatujmo

$$\begin{aligned}
&A = \{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\},\\
&B = \{(1,1),(1,2),(1,3),(1,4),(1,5),(2,1),(2,2),(2,3),(2,4),(3,1),\\
&(3,2),(3,3),(4,1),(4,2),(5,1)\},
\end{aligned}$$

odakle slijedi $P(A) = \frac{9}{36}$, $P(B) = \frac{15}{36}$. Kako računati $P(A|B)$? Izvjesno je, B je ostvareno pa se skup ishoda sužava i redukuje na skup koji ima 15 elemenata. Među tim elementima – ishodima nađimo one koji odgovaraju događaju A. Ti elementi su $\{(2,2),(2,4),(4,2)\}$, a ovim pretraživanjem je u stvari nađen skup – događaj AB. Slijedeći ideju iz klasične definicije vjerovatnoće imamo

$$P(A|B) = \frac{3}{15} = \frac{\frac{3}{36}}{\frac{15}{36}} = \frac{P(AB)}{P(B)}. \blacktriangleleft$$

Primjer 3.2 *Slučajno se bira tačka iz kvadrata K dužine stranice 1. Neka su A i B dva podskupa sa K. Neka je A događaj da je slučajno izabrana tačka iz A, a B događaj da je slučajno izabrana tačka iz B. Nađimo $P(A|B)$.*

$\blacktriangleright$Budući da je B izvjesno ostvareno, tj. slučajno izabrana tačka pripada skupu B, skup ishoda se svodi na B. Tražeći među elementima skupa B one koji pripadaju skupu A, mi dobijamo skup AB. Slijedeći ideju iz geometrijske vjerovatnoće, sada imamo

$$P(A|B) = \frac{\text{mes } AB}{\text{mes } B} = \frac{P(AB)}{P(B)}.$$

Posljednja jednakost slijedi iz činjenice mes $K = 1$. $\blacktriangleleft$

Vratimo se izlaganju teorije. Neka je $(\Omega, \mathcal{F}, P)$ prostor vjerovatnoće i neka je $B \in \mathcal{F}$, $P(B) > 0$. Definišimo preslikavanje $P_B : \mathcal{F} \longrightarrow [0,1]$ sa:

$$P_B(A) = P(A|B) = \frac{P(AB)}{P(B)},\ A \in \mathcal{F}. \tag{3.1}$$

Lako se provjerava da je P_B vjerovatnoća na $\mathcal{F}$. Nazivamo je uslovna vjerovatnoća uz uslov B. Druga jednakost u formuli (3.1) služi kao obrazac za računanje uslovne vjerovatnoće. Budući da su primjeri (3.1) i (3.2) u svježem sjećanju, jasno je odakle potiče motiv za drugu jednakost u (3.1).

Znači, svaki događaj B za koji je $P(B) > 0$ generiše vjerovatnosni prostor $(\Omega, \mathcal{F}, P_B)$.

Iz jednakosti (3.1) slijedi da uz uslove $P(A) > 0$, $P(B) > 0$ važi **formula množenja vjerovatnoća**

$$P(AB) = P(B)P(A|B) = P(A)P(B|A). \tag{3.2}$$

Postoji prirodno uopštenje formule (3.2). Naime

$$P(A_1A_2...A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)...P(A_n|A_1A_2...A_{n-1}),$$

uz pretpostavku da su vjerovatnoće događaja iz uslova pozitivne.

Definicija 3.1 *Konačna ili prebrojiva familija $H_i, i \in I$ događaja u prostoru vjerovatnoće $(\Omega, \mathcal{F}, P)$ je* **potpuni sistem događaja** *ako je $P(H_i) > 0$ za svako $i \in I$ i $\sum_{i \in I} H_i = \Omega$.*

Sljedeće dvije teoreme se često primjenjuju prilikom rješavanja zadataka.

Teorema 3.1 (Formula potpune vjerovatnoće.) *Neka je $H_i, i \in I$ potpuni sistem događaja u vjerovatnosnom prostoru $(\Omega, \mathcal{F}, P)$. Tada za svako $A \in \mathcal{F}$ važi*

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

♦ Za proizvoljno $A \in \mathcal{F}$ važi

$$\begin{aligned} P(A) = P(A\Omega) = P\left(A \sum_{i \in I} H_i\right) = P\left(\sum_{i \in I}(AH_i)\right) \\ = \sum_{i \in I} P(AH_i) = \sum_{i \in I} P(A|H_i)P(H_i). \blacklozenge \end{aligned}$$

Teorema 3.2 (Bajesova formula.) *Neka je $H_i, i \in I$ potpuni sistem događaja u vjerovatnosnom prostoru $(\Omega, \mathcal{F}, P)$ i neka je $A \in \mathcal{F}$, $P(A) > 0$. Tada za svako $i_0 \in I$ važi*

$$P(H_{i_0}|A) = \frac{P(A|H_{i_0})P(H_{i_0})}{\sum\limits_{i \in I} P(A|H_i)P(H_i)}.$$

♦ Iz formule proizvoda vjerovatnoća i formule potpune vjerovatnoće slijedi

$$P(H_{i_0}|A) = \frac{P(AH_{i_0})}{P(A)} = \frac{P(A|H_{i_0})P(H_{i_0})}{\sum\limits_{i \in I} P(A|H_i)P(H_i)}. \blacklozenge$$

Budući da događaji iz sistema $H_i, i \in I$ čine jednu potpunu kolekciju mogućih događaja, govorićemo da su H_i **hipoteze**.

Definicija 3.2 *Događaji A i B su* **nezavisni** *ako je*

$$P(AB) = P(A)P(B) \tag{3.3}$$

. U slučaju nejednakosti, događaji su **zavisni.**

Primijetimo da je u jednakosti iz definicije učešće događaja A i B simetrično. Uz uslov $P(A)P(B) > 0 \Leftrightarrow P(A) > 0 \wedge P(B) > 0$ važi

$$P(AB) = P(A)P(B) \Leftrightarrow P(A|B) = P(A) \Leftrightarrow P(B|A) = P(B).$$

Naime, prvu jednakost smo dijelili prvo sa $P(B)$, a zatim sa $P(A)$. Dakle, uz formulisani uslov, nezavisnost je ekvivalentan sa $P(A|B) = P(A)$, tj. uslov da je ostvaren B ne utiče na vjerovatnoću realizacije A. Isto važi kada A i B zamijene mjesta. Jednakost iz definicije obuhvata slučaj kada je vjerovatnoća jednog od događaja jednaka 0. Ako je $P(A) = 0$ i B proizvoljan događaj, tada je $P(AB) = 0$ te je $P(AB) = P(A)P(B)$, tj. događaji A i B su nezavisni. Ako je $P(A) = 1$ i B proizvoljan događaj, tada iz $P(A \cup B) = 1$ i $P(A \cup B) = P(A) + P(B) - P(AB)$ slijedi $P(AB) = P(A)P(B)$. Dakle A i B su nezavisni događaji.

Teorema 3.3 *Neka je $P(B) > 0$. Događaji A i B su nezavisni ako i samo ako je $P(A|B) = P(A)$.*

♦Neka je $P(AB) = P(A)P(B)$. Uz ovaj uslov dobijamo $P(A|B) = \frac{P(AB)}{P(B)} = P(A)$. Obrnuto, uz uslov $P(A|B) = P(A)$ imamo $P(A|B) = P(A) = \frac{P(AB)}{P(B)} \Rightarrow P(AB) = P(A)P(B)$.♦

Dakle, uz uslov $P(A)P(B) > 0$ važi

$$P(AB) = P(A)P(B) \Leftrightarrow P(A|B) = P(A) \Leftrightarrow P(B|A) = P(B).$$

Jednakost $P(A|B) = P(A)$ se odnosi na slučaj kada je vjerovatnoća realizacije događaja A uz uslov da je ostvaren događaj B jednaka bezuslovnoj. Dakle, realizacija događaja B nije "uticala" na vjerovatnoću realizacije događaja A. Ako je $P(A) = 0$, tada je za $\forall B \in \mathcal{F}, P(AB) = P(A)P(B)$ tj. događaji A i B su nezavisni. Takođe, za $\forall B \in \mathcal{F}$, događaji B i Ω su nezavisni. Primijetimo, definicija 3.2 nije opterećena uslovom $P(A) > 0, P(B) > 0$.

Kod rješavanja praktičnih vjerovatnosnih zadataka, nezavisnost se obično prihvata na osnovu naših intuitivnih predstava o opitu. Prirodno je smatrati da u opitu u kome bacamo dva novčića, pojava, recimo grba kod jednog novčića nema uticaja na pojavu, recimo, grba na drugom novčiću (osim ako novčići nisu fizički povezani npr. tankom žicom). Takođe, to što je neka žena rodila sina nema uticaja na pol djeteta koje će roditi neka druga žena te je vjerovatnoća da su dva slučajno odabrana novorođenčeta dječaci $\frac{1}{2} \cdot \frac{1}{2}$.

Postoje situacije kada nezavisnost događaja nije očigledna i ustanovljava se tek nakon provjere važenja jednakosti (3.3) (pogledati primjere (3.13) i (3.14)).

U vjerovatnoći se pojavljuje potreba za razmatranjem nezavisnosti familije događaja. Formalizujmo.

Definicija 3.3 *Neka je $(\Omega, \mathcal{F}, P)$, vjerovatnosni prostor i $A_i \in \mathcal{F}, i \in I$ proizvoljna familija događaja. Kažemo da je to* **familija nezavisnih događaja** *ako za svaki konačni podskup različitih indeksa $i_1, i_2, ..., i_k \in I$ važi*

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

Do za sada izloženu teoriju ćemo ilustrovati primjerima.

Primjer 3.3 *Profesor je za ispit pripremio n dobrih i m loših cedulja. Kolika je vjerovatnoća da student koji cedulju izvlači kao $l+1$ po redu, $0 \leqslant l \leqslant m+n-1$, izvuče dobru cedulju?*

►Neka je A događaj čiju vjerovatnoću tražimo, a B_k, $0 \vee (l-m) \leqslant k \leqslant n \wedge l$, događaj da prvih l studenata izvuku k dobrih cedulja. Na osnovu formule potpune vjerovatnoće dobijamo

$$P(A) = \sum_{k=0\vee(l-m)}^{(n-1)\wedge l} P(A|B_K)P(B_k) = \sum_{k=0\vee(l-m)}^{(n-1)\wedge l} \frac{n-k}{n+m-l} \frac{\binom{n}{k}\binom{m}{l-k}}{\binom{m+n}{l}}.$$

Kad smo zadavali B_k, maksimalna vrijednost za k je bila $n \wedge l$. Primijetimo, ako je $n \leqslant l$ tada za $k = n \wedge l = n$ važi $P(A|B_k) = P(A|B_n) = 0$ te nema potrebe da u sumi računamo član sa indeksom $n \wedge l$. Dakle sumiranje završavamo sa $k = (n-1) \wedge l$. Ako je $l < n$ tada je najveći indeks u sumi l i on se dobija bez obzira da li je najveći indeks zadat kao $n \wedge l$ ili $(n-1) \wedge l$. Ovim smo opravdali korektnost sumiranja do indeksa $k = (n-1) \wedge l$.

Nakon sređivanja opšteg člana u sumi, dobijamo

$$P(A) = \frac{n}{n+m} \sum_{k=0\vee(l-m)}^{(n-1)\wedge l} \frac{\binom{n-1}{k}\binom{m}{l-k}}{\binom{n+m-1}{l}} = \frac{n}{n+m}.$$

Naime, suma je jednaka 1. Objasnimo zbog čega. U slučaju $l = 0$ jednakost je očigledna. Skoncetrišimo se na model sa $n-1$ dobrih i m loših cedulja, i cedulje izvlači l, $1 \leqslant l \leqslant n+m-1$ studenata. Ako sa k označimo broj izvučenih dobrih cedulja, tada je $0 \vee (l-m) \leqslant k \leqslant (n-1) \wedge l$, a vjerovatnoće da se izvuče k dobrih cedulja se nalaze u sumi. Sumirajući te vjerovatnoće dobijamo 1. Vidimo da u rezultatu ne figuriše l. Dakle, bez obzira na poziciju studenta u "redu", vjerovatnoća izvlačenja dobre cedulje je $\frac{n}{m+n}$.

Naš model sa ceduljama je ekvivalentan sa sljedećim. U kutiji se nalazi m bijelih i $n-m$ crnih kuglica. Po modelu bez vraćanja se vade kuglice. Izračunajmo vjerovatnoće da

a) je j-ta izvađena kuglica bijela?

b) su i-ta i j-ta izvađena kuglica bijele, $i \neq j$?

c) je i-ta bijela a j-ta crna, $i \neq j$?

a) Kuglice se mogu izvući na $n!$ načina. Svakom od tih izvlačenja-nizova pridružujemo istu vjerovatnoću (mjeru) $\frac{1}{n!}$. Interesuje nas koliko ima nizova sa bijelom kuglicom na j-tom mjestu. j-to mjesto u nizu popunjava neka od m bijelih, a ostalih $n-1$ mjesta popunjavaju ostalih $n-1$ kuglica.

Dakle,

$$p = \frac{m(n-1)(n-2)...1}{n!} = \frac{m}{n}.$$

Možemo rezonovati i ovako. j ta izvađena kuglica može biti bilo koja od n kuglica iz kutije (kuglice su ravnopravne), a nama odgovara bilo koja od m bijelih. Dakle, $p = \frac{m}{n}$.

b) j-to mjesto u nizu popunjava neka od m bijelih, i-to mjesto u nizu popunjava neka od $m-1$ bijelih, a za preostalih $n-2$ mjesta ostaje $n-2$ kuglica. Dakle,

$$p = \frac{m(m-1)(n-2)...1}{n!} = \frac{m(m-1)}{n(n-1)}.$$

Možemo rezonovati i ovako. Dvije bijele kuglice koje će biti izvađene u i tom i j tom vađenju možemo izabrati na $\binom{m}{2}$ načina, dvije kuglice koje će biti izvađene u i tom i j tom vađenju možemo izabrati na $\binom{n}{2}$ načina. Dakle,

$$p = \frac{\binom{m}{2}}{\binom{n}{2}} = \frac{m(m-1)}{n(n-1)}.$$

c) $p = \frac{m(n-m)}{n(n-1)}$.◀

Primjer 3.4 *Iz kutije u kojoj se se nalazi 10 bijlih, 11 crvenih i 12 crnih kuglica, po modelu a) sa vraćanjem; b) bez vraćanja, vade se kuglica. Kolika je vjerovatnoća da će bijela kugica biti izvađena prije crne?*

Primjer 3.5 *U svakoj od dvije kutije se nalazi po* $10B$ *i* $5C$ *kuglica. Iz prve kutije se vadi kuglica, prebacuje u drugu, zatim se iz druge vadi kuglica i prebacuje u prvu. Na kraju se iz prve kutije vadi kuglica. Kolika je vjerovatnoća da je ta kuglica bijela?*

▶Uvedimo hipoteze $H_1: \ BB$ (u prvom vađenju je iz prve kutije izvučena bijela kuglica, a iz druge kutije je izvučena bijela kuglica), $H_2: \ BC$, $H_3: \ CB$, $H_4: \ CC$. Ako sa W označimo događaj čiju vjerovatnoću tražimo, dobijamo

$$\begin{aligned} P(W) &= P(W|H_1)P(H_1) + P(W|H_2)P(H_2) + P(W|H_3)P(H_3) + \\ &+ P(W|H_4)P(H_4) = \frac{2}{3}\cdot\frac{2}{3}\cdot\frac{11}{16} + \frac{6}{15}\cdot\frac{2}{3}\cdot\frac{5}{16} + \frac{11}{15}\cdot\frac{1}{3}\cdot\frac{10}{16} + \\ &+ \frac{2}{3}\cdot\frac{1}{3}\cdot\frac{6}{16} = \frac{2}{3}. \end{aligned}$$

Dakle, vjerovatnoće prije i poslije prebacivanja su iste.◀

Primjer 3.6 *Igrači* A *i* B *igraju do bankrota jednog od njih. Na početku igrač* A *ima* a *eura, a igrč* B *ima* b *eura. U svakoj partiji vjerovatnoća pobjede igrača* A *je* p*, a igrača* B *je* q*,* $p + q = 1$*. Kad* *partiju dobije igrač A tada mu igrač B daje euro, a kad partiju dobije igrač B tada mu igrač A daje euro. Naći vjerovatnoću bankrota za svakog igrača.*

▶Neka je $p_n, n = 0, 1, ..., a + b$ vjerovatnoća bankrota igrača A kad ima n eura. Jasno, $p_{a+b} = 0, p_0 = 1$. Neka je H_1-u prvoj partiji pobjeđuje A, H_2-u prvoj partiji pobjeđuje B. Sada je

$$\begin{aligned}
p_n &= P(A|H_1)P(H_1) + P(A|H_2)P(H_2) = pp_{n+1} + qp_{n-1} \Rightarrow \\
&(p+q)p_n = pp_{n+1} + qp_{n-1} \Rightarrow q(p_n - p_{n-1}) = p(p_{n+1} - p_n), n = 1, 2, ..., a + b - 1. \\
1^0 \quad & p = q = \frac{1}{2}, \ p_{n+1} - p_n = p_n - p_{n-1} = ... = p_1 - p_0 = c \Rightarrow p_n = p_0 + nc, \\
n = \quad & a + b \Rightarrow c = -\frac{1}{a+b} \Rightarrow p_n = 1 - \frac{n}{a+b}, p_a = \frac{b}{a+b}, q_b = \frac{a}{a+b}. \\
2^0 \quad & p \neq q. \ q^n \prod_{k=1}^{n} (p_k - p_{k-1}) = p^n \prod_{k=1}^{n} (p_{k+1} - p_k) \Rightarrow p_{n+1} - p_n = \left(\frac{q}{p}\right)^n (p_1 - 1). \\
& p_{a+b} - p_n = \sum_{k=n}^{a+b-1} (p_{k+1} - p_k) = \sum_{k=n}^{a+b-1} \left(\frac{q}{p}\right)^k (p_1 - 1) = (p_1 - 1)\frac{\left(\frac{q}{p}\right)^n - \left(\frac{q}{p}\right)^{a+b}}{1 - \frac{q}{p}}. \\
& p_{a+b} = 0 \Rightarrow p_n = (1 - p_1)\frac{\left(\frac{q}{p}\right)^n - \left(\frac{q}{p}\right)^{a+b}}{1 - \frac{q}{p}}.
\end{aligned}$$

Stavljajući u posljednjem izrazu $n = 0$ dobijamo $1 - p_1$ i na kraju je

$$p_n = \frac{\left(\frac{q}{p}\right)^{a+b} - \left(\frac{q}{p}\right)^n}{\left(\frac{q}{p}\right)^{a+b} - 1} \Rightarrow p_a = \frac{1 - \left(\frac{p}{q}\right)^b}{1 - \left(\frac{p}{q}\right)^{a+b}}, q_b = \frac{1 - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^{a+b}}. ◀$$

Primjer 3.7 *U kutiji se nalaze tri novčića. Dva su "normalna", a treći je iskovan tako da na obje strane ima pismo. Slučajno se bira jedan novčić i baca četiri puta. Naći vjerovatnoću da je uzet "normalni" novčić ako je u sva četiri bacanja palo pismo?*

▶Označimo sa A događaj da je u sva četiri bacanja palo pismo, sa H_1 hipotezu da je uzet "normalni" novčić, a sa H_2 hipotezu da je uzet "felerični" novčić. Primjenom Bajesove formule dobijamo

$$P(H_1|A) = \frac{P(A|H_1)P(H_1)}{P(A|H_1)P(H_1) + P(A|H_2)P(H_2)} = \frac{\frac{1}{16} \cdot \frac{2}{3}}{\frac{2}{3} \cdot \frac{1}{16} + 1 \cdot \frac{1}{3}} = \frac{1}{9}. ◀$$

Primjer 3.8 *U svakoj od dvije kutije se nalazi po N kuglica, u prvoj kutiji je M_1 bijelih, u drugoj je M_2 bijelih. Slučajno se bira kutija, a zatim se iz nje po modelu sa vraćanjem vadi n kuglica. Sa H_1 označimo hipotezu da je izabrana prva, a sa H_2 hipotezu da je izabrana druga kutija. Neka je W događaj da su sve izvađene kuglice bijele. Izračunati $P(H_k|W), k = 1, 2$.*

$$\blacktriangleright\ P(H_k|W) = \frac{0,5\left(\frac{M_k}{N}\right)^n}{0,5\left(\frac{M_1}{N}\right)^n + 0,5\left(\frac{M_2}{N}\right)^n} = \frac{M_1^n}{M_1^n + M_2^n}, k = 1, 2.$$

Ako je $M_2 < M_1$, tada je $\lim_{n\to\infty} P(H_1|W) = 1$. U ovom slučaju, znanje o realizaciji događaja W suštinski mijenja našu apriornu informaciju o hipotezama H_1 i H_2. ◀

Primjer 3.9 *U kutiji se nalazi* 91 *kuglica i svaka može biti bijela ili crna. Iz kutije je po modelu bez vraćanja izvučeno* 19 *kuglica i među njima je registrovano* 7 *bijelih i* 12 *crnih kuglica. Naći najvjerovatniji prvobitni sastav kutije.*

▶Označimo sa H_k hipotezu da je u kutiji k, $k = 0, 1, \dots, 91$, bijelih kuglica. Naravno, $P(H_k) = \frac{1}{92}$ za svako k. Označimo sa A događaj da je izvučeno 7B i 12C kuglica. Nakon primjene Bajesove formule dobijamo

$$P(H_k|A) = \frac{P(H_k)P(A|H_k)}{\sum\limits_{i=0}^{91} P(H_i)P(A|H_i)} = \frac{P(A|H_k)}{\sum\limits_{i=0}^{91} P(A|H_i)}.$$

Zadatak traženja maksimuma za $P(H_k|A)$ ekvivalentan je zadatku traženja maksimuma za $P(A|H_k)$. Imamo

$$P(A|H_k) = \frac{\binom{k}{7}\binom{91-k}{12}}{\binom{91}{19}}.$$

Nakon kraće analize se dobija $P(A|H_k) < P(A|H_{k+1})$ za $k \leqslant 32$ i $P(A|H_k) > P(A|H_{k+1})$ za $k \geqslant 33$. Odavde zaključujemo da je najvjerovatniji sastav kutije 33B i 58C. Primijetimo da je $\frac{7}{19} \cdot 91 \approx 33$ te možemo konstatovati da dobijeni rezultat korespondira sa našom intuicijom.◀

Primjer 3.10 *U svakoj od n kutija se nalazi b bijelih i c crnih kuglica. Iz prve se kutije u drugu prebacuje jedna, zatim se iz druge u treću prebacuje jedna,.... Kolika je vjerovatnoća da se iz n-te izvadi bijela kuglica?*

▶ $H_k, k = 1, 2, .., n$ je događaj da se iz $k$te kutije izvadi bijela kuglica.

$$P(H_{k+1}) = P(H_{k+1} \mid H_k)P(H_k) + P(H_{k+1} \mid H_k^c)P(H_k^c) = \frac{b + P(H_k)}{b + c + 1}, k = 1, 2, ..., n - 1.$$

Iz $P(H_1) = \frac{b}{b+c}$ slijedi $P(H_2) = ... = P(H_n) = \frac{b}{b+c}$. ◀

Primjer 3.11 *U kutiji je k kuglica, svaka od njih sa vjerovatnoćom $\frac{1}{2}$ (nezavisno od ostalih) može biti bijela ili crna. Iz kutije n puta vadimo kuglicu, model sa vraćanjem. Između izvađenih kuglica m su bijele, $0 < m < n$. Kolika je vjerovatnoća da u kutiji ima tačno s bijelih kuglica $(0 < s < k)$?*

$$\blacktriangleright\ P(H_s \mid W) = \frac{P(H_s)P(W \mid H_s)}{\sum\limits_{i=1}^{k-1} P(W \mid H_i)P(H_i)} = \frac{\binom{k}{s}\frac{1}{2^k}\binom{n}{m}\left(\frac{s}{k}\right)^m\left(1-\frac{s}{k}\right)^{n-m}}{\sum\limits_{i=1}^{k-1}\binom{n}{m}\left(\frac{i}{k}\right)^m\left(1-\frac{i}{k}\right)^{n-m}\binom{k}{i}\frac{1}{2^k}} =$$
$$= \frac{\binom{k}{s}s^m(k-s)^{n-m}}{\sum\limits_{i=1}^{k-1}\binom{k}{i}(k-i)^{n-m}}. \blacktriangleleft$$

Primjer 3.12 *Iz skupa $\{1,2,...,n\}$ jedan za drugim se biraju dva broja. Kolika je vjerovatnoća da je razlika između prvog i drugog izvađenog broja $\geqslant m, 0 < m < n$?*

$$\blacktriangleright\ P(W) = \sum_{i=m+1}^{n} P(W \mid H_i)P(H_i) = \frac{1}{n}\Big(\frac{1}{n-1} + \frac{2}{n-1} + ... + \frac{n-m}{n-1}\Big) = \frac{(n-m)(n-m+1)}{2(n-1)n}. \blacktriangleleft$$

Primjer 3.13 *Iz kompleta od 32 karte, slučajno se vadi karta. Neka je A događaj da je izvučena karta as, a B događaj da je izvučena karta krsta. Da li su događaji A i B nezavisni?*

$\blacktriangleright$ $P(A) = \frac{4}{32} = \frac{1}{8}$, $P(B) = \frac{8}{32} = \frac{1}{4}$, $P(AB) = \frac{1}{32}$ (postoji samo jedan as krsta). Dakle $P(AB) = P(A)P(B)$ te su događaji A i B nezavisni. Da li se nezavisnost narušava kada se u komplet stavi "prazna" karta?$\blacktriangleleft$

Primjer 3.14 *U kvadratu $K = (0,1)\times(0,1)$ slučajnio se bira tačka. Neka je A događaj da je izabrana tačka iz oblasti $A = \{(x,y) : (x,y) \in K, x > a\}$, $0 < a < 1$ i neka je B događaj da je izabrana tačka iz oblasti $B = \{(x,y) : (x,y) \in K, y > b\}$, $0 < b < 1$. Da li su događaji A i B nezavisni?*

Primjer 3.15 *U kutiji se nalaze četiri kuglice na kojima su zapisani redom brojevi $2,3,5,30$. Iz kutije se vadi kuglica. Označimo sa A_k, $k \in \{2,3,5,30\}$ događaj da je broj na izvučenoj kuglici djeljiv sa k.*

$\blacktriangleright$Imamo

$$P(A_2) = P(A_3) = P(A_5) = \frac{1}{2},\ \frac{1}{4} = P(A_2A_3) = P(A_2)P(A_3)$$
$$= P(A_2A_5) = P(A_2)P(A_5) = P(A_3A_5) = P(A_3)P(A_5),$$
$$\frac{1}{4} = P(A_2A_3A_5) \neq P(A_2)P(A_3)P(A_5) = \frac{1}{8}.$$

Događaji A_2, A_3, A_5 su nezavisni u parovima, ali ne i u ukupnosti.$\blacktriangleleft$

Primjer 3.16 *Kocka se baca dva puta. Izdvojimo događaje*

$$A = \{(i,j) : j \in \{1,2,5\}\},\ B = \{(i,j) : j \in \{4,5,6\}\},$$
$$C = \{(i,j) : i + j = 9\}.$$

Imamo,

$$P(A) = P(B) = \frac{1}{2}, P(C) = \frac{1}{9}, P(AB) = \frac{1}{6}, P(ABC) = \frac{1}{36},$$

pa možemo konstatovati

$$P(AB) \neq P(A)P(B) \text{ i } P(ABC) = P(A)P(B)P(C). \blacktriangleleft$$

Primjer 3.17 *U svakom susretu igrači A i B imaju jednake šanse da osvoje bod. Pobjeđuje igrač koji prvi osvoji 6 bodova. Naći vjerovatnoću događaja da pobijedi igrač A ako trenutno vodi rezultatom 4 : 2.*

►Sa A ćemo simbolički označavati pobjedu igrača A u pojedinačnoj igri, a sa B pobjedu igrača B u pojedinačnoj igri. B će biti ukupni pobjednik ako se ostvari neka od shema BBBB, ABBBB, BABBB, BBABB, BBBAB. Zbog podrazumijevane nezavisnosti između igara, imamo da je vjerovatnoća pobjede igrača B

$$p_B = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + 4\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{16},$$

a vjerovatnoća pobjede igrača A je $p_A = \frac{13}{16}$. ◄

Primjer 3.18 *U populaciji je* 2% *oboljelih. Kada se testira oboljela osoba test je pozitivan sa vjerovatnoćom 0,99, a kad se testira zdrava test je pozitivan u* $0,5\%$ *slučajeva.*

a) Kolika je vjerovatnoća da je test pozitivan?

b) Ako je test pozitivan, kolika je vjerovatnoća da je testirana oboljela osoba?

►A- test je pozitivan, H_1 - testira se oboljela osoba, H_2 - testira se zdrava osoba. $P(H_1) = 0,02$, $P(H_2) = 0,98$, $P(A|H_1) = 0,99$, $P(A|H_2) = 0,005$. a) Nakon primjene formule potpune vjerovatnoće dobijamo $P(A) = 0,0247$. b) Nakon primjene Bajesove formule dobijamo $P(H_1|A) = 0,8016$. ◄

Primjer 3.19 *Kolika je vjerovatnoća da se u seriji bacanja kocke tri uzastopne šestice pojave prije dvije uzastopne jedinice?*

►Neka je p_1 vjerovatnoća ostvarivanja našeg događaja ako serija počinje sa 6, p_2 ako serija počinje sa 1, p_3 ako serija počinje sa $2, 3, 4$ ili 5 i p vjerovatnoća ostvarivanja našeg događaja. Nakon primjene formule potpune vjerovatnoće, dobijamo

$$p = \frac{1}{6}p_1 + \frac{1}{6}p_2 + \frac{4}{6}p_3.$$

Potražimo p_1. Serija je počela sa 6 i može biti nastavljena sa 6, 1 i nekim od "neutralnih" brojeva $2,3,4,5$. Ako je "nastavak" 6 mogući su sljedeći nastavci do konačne realizacije našeg događaja: Prvi sljedeći broj je 6 i time se traženi događaj realizuje, prvi sljedeći broj je 1, dvije prethodne šestice gube na značaju i ulazimo u model koji počinje sa 1, prvi sljedeći broj je neki od "neutralnih" i ulazimo u model koji počinje sa "neutralnim" brojem. Ako je "nastavak" 1 ulazimo u model koji počinje sa 1. Ako je "nastavak" neki "neutralni" broj ulazimo u model koji počinje sa "neutralnim" brojem. Sada dobijamo

$$p_1 = \frac{1}{6}(\frac{1}{6} + \frac{1}{6}p_2 + \frac{4}{6}p_3) + \frac{1}{6}p_2 + \frac{4}{6}p_3 \Longleftrightarrow 36p_1 - 7p_2 - 28p_3 = 1.$$

Slično potražimo p_2 i p_3 i dobijemo još dvije jednačine: $p_1 - 6p_2 + 4p_3 = 0$, $p_1 + p_2 - 2p_3 = 0$. Nakon rješavanja sistema jednačina dobija se $p = 7/50$. ◀

Primjer 3.20 *Kocka se baca n puta. Kolika je vjerovatnoća da se šestica pojavi paran boj puta (za potrebe zadatka nulu računati kao paran broj)*

▶ Neka je H_1 hipoteza da serija počinje sa šesticom, H_2 hipoteza da počinje nekim drugim brojem i neka je p_n tražena vjerovatnoća. Imamo,

$$p_n = P(H_1)(1 - p_{n-1}) + P(H_2)p_{n-1} = \frac{1}{6} + \frac{2}{3}p_{n-1},\, p_1 = \frac{5}{6}.$$

Nakon rješavanja formirane rekurzivne relacije dobijamo $p_n = \dfrac{1}{2}(1 + (2/3)^n)$. ◀

Primjer 3.21 *Kocka se baca n puta. Kolika je vjerovatnoća da se u seriji bacanja ne pojavi par PP (pismo, pismo)?*

▶ Neka je H_1 hipoteza da serija počinje pismom, a H_2 hipoteza da počinje grbom. U slučaju ostvarivanja H_1, traženi događaj će se realizovati ako se u sljedećem bacanju dobije grb i u preostalih $n-2$ bacanja ne ostvari par PP. U slučaju ostvarivanja H_2, traženi događaj će se realizovati ako se u sljedećih $n-1$ bacanja ne pojavi par PP. Dakle,

$$p_n = \frac{1}{2}\frac{1}{2}p_{n-2} + \frac{1}{2}p_{n-1},\, p_0 = p_1 = 1.$$

Nakon rješavanja formirane rekurzivne relacije, dobija se traženo p_n. ◀

Primjer 3.22 *Izvodi se serija bacanja se novčića, u slučaju padanja grba igrač dobija jedan poen, u slučaju padanja pisma igrač dobija dva poena. Poeni se sabiraju. Kolika je vjerovatnoća da će igrač u jednom trenutku imati n poena?*

▶ $p_1 = \frac{1}{2}, p_2 = \frac{3}{4}, p_n = \frac{1}{2}p_{n-1} + \frac{1}{2}p_{n-2} \Rightarrow p_n = \frac{2}{3} + (-1)^n \frac{1}{3 \cdot 2^n}$. ◀

4 Slučajne promjenljive (veličine)

DEFINICIJA 4.1 *Neka je* $(\Omega, \mathcal{F}, P)$ *vjerovatnosni prostor. Funkcija* $X : \Omega \to R$ *je* **slučajna promjenljiva** *ako je za svaki polusegment* $[a, b)$ *skup* $S = \{\omega : X(\omega) \in [a, b)\} \in \mathcal{F}$ *tj. događaj.*

Izostavljanjem uslova, pojavio bi se neprihvatljiv slučaj da za neki polusegment $[a, b)$ "upadanje" X u $[a, b)$ nije događaj.

Svaka slučajna promjenljiva X generiše vjerovatnoću-vjerovatnosnu mjeru na jednoj familiji skupova sa R. Mjera nije zadata na $\mathcal{P}(R)$. Ta vjerovatnoća se naziva raspodjela slučajne promjenljive X, koristi se oznaka P_X. Vjerovatnoća-mjera skupa iz pomenute familijie jednaka je vjerovatnoći sa kojom X "upada" u skup. $P_X(a) = P\{X = a\}$, $P_X[a, b) = P\{\omega : X(\omega) \in [a, b)\}$.

Slučajnom promjenljivom X svakom mogućem ishodu $\omega \in \Omega$ pridružujemo njegovu numeričku karakteristiku-obilježje $X(\omega)$. Ako se u opitu realizovao ishod ω_0 tada kažemo da je slučajna promjenljiva "uzela" vrijednost $X(\omega_0)$. Vrijednost-broj $X(\omega_0)$ je realizacija slučajne promjenljive X. U terminologiji koja je uobičajena u fizici može se reći da slučajna promjenljiva mjeri ishode.

Primjer 4.1 *Neka je* $A \in \mathfrak{F}$ *i neka je*

$$I_A(\omega) := \begin{cases} 0, & \omega \in A^c \\ 1, & \omega \in A \end{cases}.$$

Slučajna promjenljiva I_A se naziva **indikator događaja** A. Iz definicije slijedi, $I_A = 1$ ako i samo ako se ostvari događaj A. Informacija da je I_A uzeo vrijednost 1 implicira da je u opitu ostvaren događaj A. ◄

DEFINICIJA 4.2 *Za slučajnu promjenljivu čiji je skup realizacija konačan ili prebrojiv kažemo da je* **diskretnog tipa**. *Ako je kodomen konačan, tada govorimo o* **prostoj slučajnoj promjenljivoj**.

Standardni način bilježenja raspodjele vjerovatnoće diskretne slučajne promjenljive X je dat tabelom

$$X : \begin{array}{cccccc} x_1 & x_2 & x_3 & . & . & . \\ p_1 & p_2 & p_3 & . & . & . \end{array}.$$

Primjer 4.2 *Tetraedar na čijim su stranama zapisani brojevi* $1, 2, 3, 4$ *baca se dva puta. Neka je* X *zbir palih brojeva.*

Imamo, $\Omega = \{(1,1),(1,2),...,(4,4)\}$, $\mathfrak{F} = \mathbb{P}(\Omega)$ i P je indukovano činjenicom da je $P(\omega) = \frac{1}{16}, \forall\omega \in \Omega$. Formirajmo skupove

$$A_1 = \{X(\omega) = 2\} = \{(1,1)\}, A_2 = \{X(\omega) = 3\} = \{(1,2),(2,1)\},$$
$$A_3 = \{X(\omega) = 4\} = \{(1,3),(2,2),(3,1)\}, A_4 = \{X(\omega) = 5\} = \{(1,4),(4,1),(2,3),(3,2)\}$$
$$A_5 = \{X(\omega) = 6\} = \{(2,4),(4,2),(3,3)\}, A_6 = \{X(\omega) = 7\} = \{(3,4),(4,3)\}, A_7 = \{X(\omega) = 8\} = \{(4,4)\}$$

i neka je $x_1 = 2, x_2 = 3, x_3 = 4, x_4 = 5, x_5 = 6, x_6 = 7, x_7 = 8$. Sada je

$$X(\omega) = \sum_{k=1}^{7} x_k I_{A_k}(\omega),$$

a raspodjela slučajne promjenljive X je data shemom

$$X : \begin{array}{ccccccc} 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \frac{1}{16} & \frac{1}{8} & \frac{3}{16} & \frac{1}{4} & \frac{3}{16} & \frac{1}{8} & \frac{1}{16} \end{array}.$$

Naime,

$$p_k = P(A_k), k = 1,2,...,7. \blacktriangleleft$$

Primjer 4.3 *Slučajna promjenljiva X je broj bacanja kocke do prvog uzastopnog padanja dva ista broja.*

$$P\{X = k\} = \frac{6\cdot5\cdot5\cdot5\cdots5\cdot5\cdot5\cdot1}{6\cdot6\cdot6\cdots6} = \frac{1}{6}\left(\frac{5}{6}\right)^{k-2}, k = 2,3,4....$$

Primjer 4.4 *Kocka se baca deset puta. Naći raspodjelu slučajne promjenljive X koja je jednaka broju pojavljivanja brojeva djeljivih sa tri.*

Opit tretiramo kao Bernuilijevu shemu sa 10 ponavljanja (bacanja) opita u kome su ishodi padanje broja koji je djeljiv sa 3-uspjeh i padanje broja koji nije djeljiv sa 3-neuspjeh. Vjerovatnoća uspjeha je $\frac{1}{3}$. Slučajna promjenljiva X prebrojava uspjehe te možemo konstatovati $X : \mathcal{B}(10, \frac{1}{3})$, iza čega stoji

$$P\{X = k\} = \binom{10}{k}\left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{10-k}, \; k = 0,1,\ldots,10. \blacktriangleleft$$

Primjer 4.5 *r čestica se raspoređuje na n orbita oko atomskog jezgra. Naći raspodjelu slučajne promjenljive X koja je jednaka broju čestica raspoređenih na orbiti koja je najbliža jezgru.*

Pozicioniranje svih čestica na orbitama tretiramo kao Bernulijevu shemu, uspjehom se smatra pozicioniranje čestice na orbiti najbližoj jezgru. Vjerovatnoća uspjeha je $\frac{1}{n}$, r čestica bira orbitu te je $X : \mathcal{B}(r, \frac{1}{n})$.

$$P\{X = k\} = \binom{r}{k}\left(\frac{1}{n}\right)^k\left(1 - \frac{1}{n}\right)^{r-k}, k = 0,1,2,...,r.$$

Primijetimo, ako je $r = 2n$ tada

$$P\{X = k\} \to \frac{2^k}{k!}e^{-2}, n \to \infty, k = 0, 1, 2, ... \blacktriangleleft$$

Granični slučajevi binomne raspodjele dovode do dvije značajne raspodjele: Puasonove i normalne.

Teorema 4.1 *Neka je* $S_n : \mathcal{B}(n, p_n), (n \in \mathbb{N})$ *i neka je* $\lim_{n\to\infty} np_n = \lambda, 0 < \lambda < \infty$. *Tada za svako* $k = 0, 1, 2, ...$ *važi*

$$\lim_{n\to\infty} P\{S_n = k\} = \frac{\lambda^k}{k!}e^{-\lambda}.$$

Uvedimo oznaku $np_n = \lambda_n$. Na osnovu uslova imamo $\lim_{n\to\infty} \lambda_n = \lambda$. Za proizvoljno $k = 0, 1, 2, ..$ imamo

$$\begin{aligned} P\{S_n = k\} &= \frac{n(n-1)...(n-k+1)}{k!}\Big(\frac{\lambda_n}{n}\Big)^k\Big(1 - \frac{\lambda_n}{n}\Big)^{n-k} = \\ &= \frac{\lambda_n^k}{k!}\Big(1 - \frac{\lambda_n}{n}\Big)^n\Big(1 - \frac{1}{n}\Big)...\Big(1 - \frac{k-1}{n}\Big)\Big(1 - \frac{\lambda_n}{n}\Big)^{-k} \to \frac{\lambda^k}{k!}e^{-\lambda}, n \to \infty. \blacklozenge \end{aligned}$$

Granične vjerovatnoće iz prethodne teoreme su vezane za realizacije $k, k = 0, 1, 2, ...$, a kako je $\sum_{k=1}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda} = 1$ to teorema sugeriše slučajnu promjenljivu $S_\infty = X$ sa raspodjelom

$$P\{X = k\} = \frac{\lambda^k}{k!}e^{-\lambda},\ k = 0, 1, 2, \ldots, \lambda > 0.$$

Za ovako raspodijeljenu slučajnu promjenljivu kažemo da ima **Puasonovu** raspodjelu sa parametrom λ i koristimo zapis $X : \mathcal{P}(\lambda)$. Puasonova raspodjela je rasprostranjena u praksi. Puasonovu raspodjelu ima: broj uključivanja na web sajt u toku sata, broj odšteta koje u toku mjeseca svojim klijentima isplati osiguravajuća kompanija, broj zemljotresa u nekom području u toku deset godina. Navešćemo primjer u kojem su n i p u relaciji iz teoreme. Radioaktivni elementi emituju α čestice u savršeno slučajnim momentima. α čestica je jezgro helijuma i ona nastaje u trenutku kada se raspadne atom radioaktivnog elementa. U djeliću radijuma mase 1 pikogram (pikogram, oznaka pg=$10^{-12}g$) ima 10^{12} atoma. Za proizvoljni atom radijuma vjerovatnoća da se raspadne u toku jedne sekunde je $p = 10^{-12}$ i atomi se nezavisno raspadaju. Broj α čestica koje "grumen" radijuma mase 5pg emituje u toku sekunde ima $\mathcal{P}(5)$ raspodjelu.

Ako je $S_n : \mathcal{B}(n, p_n)$ tada na osnovu Teoreme (4.1) za velike n i male p_n važi približna formula

$$P\{S_n = k\} \approx \frac{\lambda_n^k}{k!}e^{-\lambda_n},\ k = 0, 1, 2, \ldots, n, \lambda_n = np_n.$$

Ova aproksimacija se obično upotrebljava ako je $n \geqslant 20$, $np_n < 10$.

Primjer 4.6 *Kolika je vjerovatnoća da među 1000 slučajno odabranih osoba njih četvoro slavi rođendan prvog januara?*

Slučajna promjenljiva X koja među odabranim "prebrojava" osobe rođene prvog januara ima $\mathcal{B}(1000, \frac{1}{365})$ raspodjelu koja se aproksimira sa $P(2,73)$ raspodjelom. Dakle, tražena vjerovatnoća je

$$\frac{(2,73)^4}{4!}e^{-2,73} = 0,151.$$

Primjer 4.7 *U kutiji je n bijelih i m crnih kuglica. Slučajno se a) po modelu sa vraćanjem, b) po modelu bez vraćanja, izvlači $k, 1 \leqslant k \leqslant n+m$ kuglica. Naći raspodjelu slučajne promjenljive X koja je jednaka broju izvučenih bijelih kuglica. Preciznije je, broju izvlačenja u kojima je izvučena bijela kuglica, ali prva formulacija je opšteprihvaćena.*

a) $X : \mathcal{B}(k, \frac{n}{n+m})$ jer je vjerovatnoća izvlačenja bijele kuglice u svakom izvlačenju $\frac{n}{n+m}$.

b)
$$P\{X = l\} = \frac{\binom{n}{l}\binom{m}{k-l}}{\binom{n+m}{k}}, \ \max\{0, k-m\} \leqslant l \leqslant \min\{n, k\}. \tag{4.1}$$

Raspodjela zadata sa (4.1) se naziva **hipergeometrijska** i koristi se notacija $X : H(n, m, k)$.

U slučaju kada $\frac{n}{m+n} \to p$, $0 < p < 1$, $n \to \infty$, tada

$$P\{X = l\} \to \binom{k}{l}p^l q^{k-l}, l = 0, 1, ..., k, \ p = 1 - q, \ n \to \infty.$$

Dakle, granična raspodjela je binomna. ◄

Primjer 4.8 *Naći raspodjelu slučajne promjenljive koja je*

a) jednaka broju ponavljanja neuspjeha (ili prosto broju neuspjeha) do ostvarivanja prvog uspjeha,

b) jednaka broju ponavljanja opita do ostvarivanja prvog uspjeha,

c) jednaka broju neuspjeha do ostvarivanja r tog uspjeha, $r \geqslant 1$.

Vjerovatnoća uspjeha je p.

a) $P\{X = k\} = q^k p, k = 0, 1, 2, ...; \sum\limits_{k=0}^{\infty} pq^k = \frac{p}{1-q} = 1.$

b) $P\{Y=k\} = q^{k-1}p, k = 1,2,....; \sum\limits_{k=1}^{\infty} pq^{k-1} = 1.$

c) $P\{Z=k\} = \binom{k+r-1}{r-1} q^k p^r = \binom{k+r-1}{k} p^r q^k = \binom{-r}{k} p^r \big(-q\big)^k, k = 0,1,2,...$

Raspodjela iz a) se naziva **geometrijska** sa parametrom p i koristi se notacija $X : \mathcal{G}(p)$, a iz c) **negativna binomna** sa parametrima r i p i koristi se notacija $Z : \mathcal{B}i(r,p)$. U slučaju $r = 1$ dobija se geometrijska raspodjela. Primijetima da je $Y = X + 1$.

Funkcija raspodjele slučajne promjenljive

Definicija 4.3 **Funkcija raspodjele vjerovatnoće slučajne promjenljive** X (ili kratko **funkcija raspodjele**) je funkcija $F_X = F : R \longrightarrow [0,1]$ zadata sa

$$F(x) = P\{X < x\} = P\{\omega :\ X(\omega) < x\},\ x \in R.$$

Primjer 4.9 *Novčić se baca dva puta. Neka je X broj palih pisama. Imamo:*

$$\Omega = \{(GG), (GP), (PG), (PP)\}, \mathfrak{F} = \mathbb{P}(\Omega)$$

i vjerovatnoća se zadaje tako što se svakom elemntarnom ishodu pridruži vjerovatnoća $\frac{1}{4}$.

$$X : \begin{array}{ccc} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{array}.$$

Za $x \leqslant 0$ imamo $F(x) = P\{\omega :\ X(\omega) < x\} = P\{\emptyset\} = 0$.

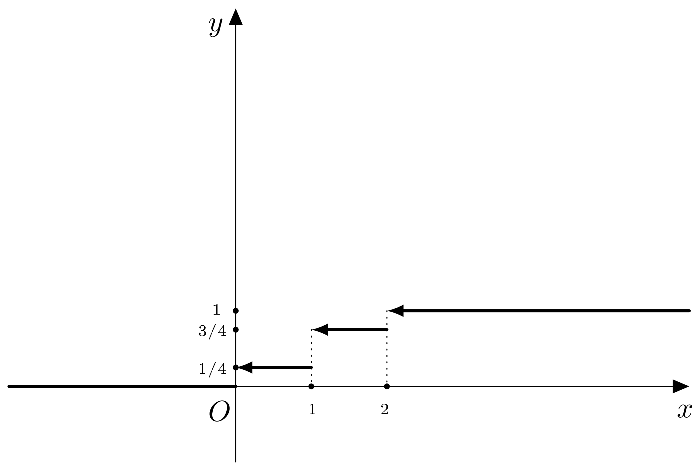
Za $0 < x \leqslant 1$ imamo $F(x) = P\{\omega :\ X(\omega) < x\} = P\{(GG)\} = 1/4$.

Za $1 < x \leqslant 2$ imamo $F(x) = P\{\omega :\ X(\omega) < x\} = P\{(GG), (GP), (PG)\} = 3/4$.

Za $x > 2$ imamo $F(x) = P\{\omega :\ X(\omega) < x\} = P\{\Omega\} = 1$.

Funkcija raspodjele slučajne promjenljive X je

$$F(x) = \begin{cases} 0, & x \leqslant 0 \\ 1/4, & 0 < x \leqslant 1 \\ 3/4, & 1 < x \leqslant 2 \\ 1, & x > 2 \end{cases}.$$



Grafik funkcije raspodjele za Primjer 1.9.

Primijetimo da se raspodjela slučajne promjenljiva X može pročitati sa grafika funkcije $F(x)$. Naime, tačke u kojima funkcija $F(x)$ ima skok su vrijednosti slučajne promjenljive, a veličina skoka jednaka je vjerovatnoći sa kojom se uzima odgovarajuća vrijednost. ◀

Navešćemo svojstva funkcije raspodjele.

a) $F(x)$ **je monotono rastuća funkcija.**

b) $F(-\infty) = \lim\limits_{x\to-\infty} F(x) = 0,\ F(\infty) = \lim\limits_{x\to\infty} F(x) = 1.$

c) **Za** $\forall x \in R$ **važi** $F(x-0) = F(x)$ **tj. funkcija** $F(x)$ **je neprekidna sa lijeve strane u svim tačkama.**

d) Neka je $[a,b)$ proizvoljan segment na realnoj pravoj. **Tada je**

$P\{X \in [a,b)\} = F(b) - F(a)$. Dokažimo.

$$\{\omega :\ X(\omega) \in [a,b)\} = \{a \leqslant X < b\} = \{X < b\} \setminus \{X < a\},$$

dobijamo

$$P\{X \in [a,b)\} = P\{\{X < b\} \setminus \{X < a\}\} = P\{X < b\} - P\{X < a\} = F(b) - F(a).$$

e) Neka je x_0 proizvoljan realan broj. **Važi:** $P\{X = x_0\} = F(x_0+0) - F(x_0)$.

g) Neka je F_X funkcije raspodjele vjerovatnoće slučajne promjenljive X. Neposredno iz f) slijedi: funkcija F_X je neprekidna u tački x_0 ako i samo ako je $P\{X = x_0\} = 0$.

Definicija 4.4 *Funkcija* $F : \mathbb{R} \to [0,1]$*: koja je*

1^0 *monotono rastuća,*

2^0 *neprekidna sa lijeve strane i*

3^0 $F(-\infty) = 0, \;\; F(\infty) = 1$

se naziva **funkcija raspodjele.**

Teorema 4.2 *Neka je* $F(x)$, $x \in \mathbb{R}$ *funkcija raspodjele. Tada postoje vjerovatnosni prostor* $(\Omega, \mathfrak{F}, P)$ *i slučajna promjenljiva* X *na njemu takvi da je* $F_X = F$.

Definicija 4.5 *Za slučajnu promjenljivu* X *kažemo da je* **apsolutno neprekidnog tipa** *ako postoji nenegativna funkcija* g *takva da je*

$$F_X(x) = \int\limits_{-\infty}^{x} g(t)dt \text{ za } \forall x \in \mathbb{R}. \tag{4.2}$$

Za raspodjelu slučajne promjenljive apsolutno neprekidnog tipa kažemo da je apsolutno neprekidnog tipa.

Funkcija $F_X(x)$ je neprekidna te je za svako $x_0 \in R$, $P\{X = x_0\} = 0$. Funkcija g se naziva **gustina slučajne promjenljive** X ili prosto **gustina od** X. Jasno, $\int\limits_{-\infty}^{\infty} g(t)dt = F_X(\infty) = 1$. Iz Analize znamo da u svim tačkama neprekidnosti funkcije g važi $F'(x) = g(x)$.

Neka je $[a,b)$ proizvoljni polusegment na realnoj pravoj i neka je X slučajna promjenljiva apsolutno neprekidnog tipa sa funkcijom raspodjele F i funkcijom gustine g. Imamo

$$P\{X \in [a,b)\} = F(b) - F(a) = \int\limits_{-\infty}^{b} g(t)dt - \int\limits_{-\infty}^{a} g(t)dt = \int\limits_{a}^{b} g(t)dt.$$

Teorema o srednjoj vrijednosti integrala implicira

$$P\{x_0 \leqslant X < x_0 + \Delta x\} = g(x_0)\Delta x + o(\Delta x), \, \Delta x \to 0,$$

gdje je g gustina slučajne promjenljive X.

Teorema 4.3 *Nenegativna funkcija* g *je funkcija gustine neke slučajne promjenljive* X *ako i samo ako važi* $\int\limits_{-\infty}^{\infty} g(t)dt = 1$.

Teorijski dio izlaganja o raspodjelama apsolutno neprekidnog tipa ćemo završiti konstatacijom da se u literaturi pojavljuje definicija koja je ekvivalentna sa definicijom (4.5).

DEFINICIJA 4.6 *Za slučajnu promjenljivu X kažemo da je* **apsolutno neprekidnog tipa** *ako postoji nenegativna funkcija g takva da je za svaki $[a,b)$*

$$P\{X \in [a,b)\} = \int_a^b g(t)dt.$$

Primjer 4.10 *Neka je slučajna promjenljiva X jednaka izabranom broju u opitu slučajnog izbora broja sa intervalu (a,b). Za proizvoljni interval $(c,d) \subset (a,b)$ imamo*

$$P\{c < X < d\} = \frac{d-c}{b-a} = \int_c^d \frac{1}{b-a}dt$$

te zaključujemo da je gustina

$$g(t) = \begin{cases} \frac{1}{b-a}, & t \in (a,b), \\ 0, & t \notin (a,b). \end{cases}$$

Sada je

$$F(x) = \int_{-\infty}^{x} g(t)dt = \begin{cases} 0, & x \leqslant a, \\ \frac{x-a}{b-a}, & a < x \leqslant b, \\ 1, & x \geqslant b. \end{cases}$$

Za slučajnu promjenljivu X kažemo da ima ravnomjernu (uniformnu) raspodjelu na intervalu (a,b) i koristimo zapis $X : \mathcal{U}(a,b)$. Budući da je realizacija slučajne promjenljive X jednaka slučajno izabranoj tački-broju sa (a,b), koristi se formulacija: X je slučajna tačka sa intervala (a,b). ◀

Primjer 4.11 *Slučajna promjenljiva X ima* **eksponencijalnu** *raspodjelu sa parametrom λ, ako je njena gustina $g(x) = \lambda e^{-\lambda x}$, $x > 0$, $\lambda > 0$. Koristimo notaciju $X : \mathcal{E}(\lambda)$.*

Primjer 4.12 *Slučajna promjenljiva X ima* **normalnu** *raspodjelu sa parametrima m i σ^2, ako je njena gustina*

$$\varphi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{(x-m)^2}{2\sigma^2}}, \ x \in R.$$

Koristimo notaciju $X : \mathcal{N}(m, \sigma^2)$. ◀

Postoje algoritmi koji modeliraju konkretne raspodjele. U rezultatu, algoritmi generišu brojeve -podatke koji su po realnoj pravoj "razbacani" u skladu sa modeliranom raspodjelom.

5 Slučajni vektori

DEFINICIJA 5.1 *Funkcija* $\mathbb{X} = (X_1, X_2) : \Omega \longrightarrow \mathbb{R}^2$ *se naziva* **slučajni vektor** *ako je za svaki pravougaonik* $[a_1, b_1) \times [a_2, b_2)$ *skup* $\{\omega : (X_1, X_2)(\omega) \in [a_1, b_1) \times [a_2, b_2)\}$ *događaj.*

Teorema 5.1 *Funkcija* $\mathbb{X} = (X_1, X_2) : \Omega \longrightarrow \mathbb{R}^2$ *je slučajni vektor ako i samo ako je svaka funkcija* $X_i, i = 1, 2$ *slučajna promjenljiva.*

DEFINICIJA 5.2 *Za slučajni vektor čiji je skup realizacija konačan ili prebrojiv kažemo da je* **diskretnog tipa.**

Primjer 5.1 *Inicijalni opit u kome posmatramo događaje*

$$A_i, \sum_{i=1}^{k} A_i = \Omega, 1 \leqslant i \leqslant k, k \geqslant 3, P(A_i) = p_i, \sum_{i=1}^{k} p_i = 1$$

ponavljamo n *puta, opiti su nezavisni. Neka je* S_1 *broj opita u kojima se realizuje događaj* A_1*,* S_2 *broj opita u kojima se realizuje događaj* A_2*, ...,* S_k *broj opita u kojima se realizuje događaj* A_k*. Primijetimo,* $S_1 + S_2 + ... + S_k = n$*. Imamo*

$$\begin{aligned} &P\{(S_1, S_2, ..., S_k) = (n_1, n_2, ..., n_k)\} = P\{S_1 = n_1, S_2 = n_2, ..., S_k = n_k\} \\ &= \frac{n!}{n_1! n_2! ... n_{k-1}! n_k!} p_1^{n_1} p_2^{n_2} ... p_k^{n_k}, n_1, n_2, ..., n_k \geqslant 0, n_1 + n_2 + ... + n_k = n. \end{aligned}$$

Znamo da je

$$(p_1 + p_2 + ... + p_k)^n = \sum_{\substack{n_1+n_2+...+n_k=n \\ n_1, n_2, ..., n_k \geqslant 0}} \frac{n!}{n_1! n_2! ... n_{k-1}! n_k!} p_1^{n_1} p_2^{n_2} ... p_k^{n_k}.$$

Slučajni vektor $(S_1, S_2, ..., S_k)$ *ima* **polinomnu raspodjelu** *i koristimo zapis*

$$(S_1, S_2, ..., S_k) : \mathbb{M}(n, p_1, p_2, ..., p_k).$$

Primjer 5.2 *Na intervalu* $(0, 1)$ *slučajno biramo n tačaka. Naći raspodjelu slučajnog vekotora* (S_1, S_2, S_3) *gdje je* S_1 *broj izabranih tačaka sa* $(0, 1/6)$*,* S_2 *broj izabranih tačaka sa* $(1/6, 1/2)$*, a* S_3 *broj izabranih tačaka sa* $(1/2, 1)$*.*

$$P\{(S_1, S_2, S_3) = (n_1, n_2, n_3)\} = \frac{n!}{n_1! n_2! n_3!} (1/6)^{n_1} (1/3)^{n_2} (1/2)^{n_3}, n_1, n_2, n_3 \geqslant 0, n_1 + n_2 + n_3 = n,$$

$(S_1, S_2, S_3) : \mathbb{M}(n, 1/6, 1/3, 1/2)$. ◀

Definicija 5.3 *Slučajni vektor* $\mathbb{X} = (X_1, X_2)$ *je* **apsolutno neprekidnog tipa** *ako postoji nenegativna funkcija* $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$, *takva da je za svaki pravougaonik* $[a_1, b_1) \times [a_2, b_2)$

$$P\{(X_1, X_2) \in [a_1, b_1) \times [a_2, b_2)\} = \iint\limits_{[a_1,b_1)\times[a_2,b_2)} g(u, v)dudv.$$

Primjer 5.3 *Neka je slučajni vektor* $W = (X, Y)$ *jednak koordinatama slučajno izabrane tačke iz oblasti* $G \in R^2$. *Za proizvoljnu oblast* $g \subset G$

$$P\{W \in g\} = \frac{mes(g)}{mes(G)} = \iint\limits_{g} \frac{1}{mes(G)} dudv$$

te zaključujema da je gustina slučajnog vektora W

$$g(u, v) = \begin{cases} 0, & (u, v) \notin G, \\ \frac{1}{\operatorname{mes} G}, & (u, v) \in G. \end{cases}$$

Za slučajni vektor W *kažemo da ima ravnomjernu (uniformnu) raspodjelu na oblasti* G *i koristimo zapis* $W : \mathcal{U}(G)$. *Slučajni vektor* W *se naziva slučajna tačka sa* G.◀

Neka je (X, Y) diskretni slučajni vektor sa raspodjelom

$$P\{(X, Y) = (x_i, y_j)\} = p_{i,j}, i = 1, 2, .., j = 1, 2, ..., \sum_i \sum_j p_{i,j} = 1.$$

Potražimo marginalne raspodjele.

$$q_i = P\{X = x_i\} = P\{X = x_i, \sum_j \{Y = y_j\}\} = \sum_j P\{X = x_i, Y = y_j\} = \sum_j p_{i,j}.$$

Slično, $r_j = P\{Y = y_j\} = \sum\limits_i p_{i,j}$.

Primjer 5.4 *Tri kuglice se na slučajan način razmještaju u tri kutije. Neka je* X *broj kuglica u prvoj kutiji, a* Y *broj zauzetih kutija. Naći raspodjelu slučajnog vektora* (X, Y). *Da li su* X *i* Y *nezavisne slučajne promjenljive?*

▶Kutije ćemo označavati brojevima $1, 2$ i 3, a za ishode ćemo proglasiti trojke kod kojih su "koordinate" određene rednim brojem kutije u koju se smješta redom, prva, druga, treća kuglica. Znači,

$$\Omega = \{(1, 1, 1), (1, 1, 2), (1, 2, 1), \ldots, (3, 3, 3)\}, \overline{\Omega} = 27.$$

Sada je, recimo, $(X,Y)(1,2,1)=(2,2)$. Pokažimo na jednom primjeru kako se dobijaju elementi tabele kojom je zadata raspodjela vektora (X,Y). Npr.

$$\begin{aligned}&P\{\omega:(X,Y)(\omega)=(1,2)\}=\\&=P\{(1,2,2),(2,1,2),(2,2,1),(1,3,3),(3,1,3),(3,3,1)\}=\frac{6}{27}.\end{aligned}$$

Raspodjela slučajnog vektora (X,Y) se može predstaviti tabelom

$Y \diagdown X$	0	1	2	3
1	$\frac{2}{27}$	0	0	$\frac{1}{27}$
2	$\frac{6}{27}$	$\frac{6}{27}$	$\frac{6}{27}$	0
3	0	$\frac{6}{27}$	0	0

Sada imamo,

$$X:\begin{array}{cccc}0&1&2&3\\\frac{8}{27}&\frac{12}{27}&\frac{6}{27}&\frac{1}{27}\end{array}\ \text{ i }\ Y:\begin{array}{ccc}1&2&3\\\frac{3}{27}&\frac{18}{27}&\frac{6}{27}\end{array}.$$

Vjerovatnoće sa kojima slučajna promjenljiva X uzima odgovarajuće vrijednosti se dobijaju sumiranjem elemenata iz kolone u kojoj se nalazi odgovarajuća vrijednost. Kad se radi sa slučajnom promjenljivom Y, ulogu kolone preuzima vrsta.

Pokažimo da su komponente X i Y zavisne slučajne promjenljive. Dovoljno je primijetiti da je

$$P\{(X,Y)=(2,2)\}=\frac{6}{27}\neq P\{X=2\}P\{Y=2\}=\frac{6}{27}\frac{18}{27}=\frac{4}{27}. \blacktriangleleft$$

Primjer 5.5 *Slučajne promjenljive X i Y su nezavisne, $X:\mathcal{P}(\lambda)$, $Y:\mathcal{P}(\mu)$. Naći raspodjelu slučajne promjenljive $Z=X+Y$.*

$\blacktriangleright$ $H_l=\{Y=l\}, l=0,1,2...$ Zbog nezavisnosti komponenti X i Y, nakon primjene formule potpune vjerovatnoće dobijamo

$$\begin{aligned}P\{Z=k\}&=\sum_{l=0}^{k}P\{X+Y=k|Y=l\}P\{Y=l\}=\sum_{l=0}^{k}P\{X=l-k\}P\{Y=l\}\\&=\frac{e^{-(\lambda+\mu)}}{k!}\sum_{l=0}^{k}\binom{k}{l}\mu^l\lambda^{k-l}=\frac{(\lambda+\mu)^k}{k!}e^{-(\lambda+\mu)},\ k=0,1\ldots\end{aligned}$$

Dakle, $Z:\mathcal{P}(\lambda+\mu)$. $\blacktriangleleft$

Primjer 5.6 *Kocka se baca do prvog padanja šestice. Neka je X broj registrovanih parnih a Y broj registrovanih neparnih brojeva. Naći raspodjelu slučajnog vektora (X,Y).*

$$P\{(X,Y)=(k,l)\}=\binom{k+l-1}{l}\Big(\frac{1}{2}\Big)^l\Big(\frac{1}{3}\Big)^{k-1}\frac{1}{6}, k=1,2,...,l=0,1,2,...$$

Primjer 5.7 *Iz kutije u kojoj su nalaze kuglice sa brojevima* $1,2,...,100$ *po modelu bez vraćanja se izvlači* 10 *kuglica. Naći raspodjelu slučajne promjenljive* Y_6*-šesti po veličini među izvučenim brojevima.*

$$P\{Y_6=k\}=\frac{\binom{k-1}{5}\binom{100-k}{4}}{\binom{100}{10}}, 6\leqslant k\leqslant 96.$$

Primjer 5.8 *Iz kutije u kojoj se nalaze kuglice numerisane brojevima* $1,2,...,60$*, po modelu bez vraćanja se vadi* 10 *kuglica. Naći raspodjelu slučajnog vektora* (Y_1,Y_{10}) *gdje je* Y_1 *najmanji, a* Y_{10} *najveći od izvađenih brojeva.*

$$P\{(Y_1,Y_{10})=(k,l)\}=\frac{\binom{l-k-1}{8}}{\binom{60}{10}}.$$

6 Matematičko očekivanje, disperzija i koeficijent korelacije

Neka je ξ slučajna promjenljiva sa raspodjelom $P\{\xi=x_i\}=p_i$, $i=1,2,...,k$. Opit u kome se pojavljuje slučajna promjenljiva ξ ponavljamo n puta u uslovima koji garantuju nezavisnost opita. Neka su u tih n ponavljanja opita registrovane sljedeće realizacije slučajne promjenljive ξ: $a_1,a_2,...,a_n$. Nađimo aritmetičku sredinu registrovanih realizacija tj.

$$\frac{a_1+...+a_n}{n}=\sum_{i=1}^{k}x_i\frac{\nu_i}{n},$$

gdje je ν_i broj registrovanih realizacija x_i. Ako $\frac{\nu_i}{n}$ tj. relativnu frekvenciju registrovanja realizacije x_i u našoj seriji zamijenimo sa p_i dobijamo veličinu $\sum\limits_{i=1}^{k}x_ip_i=E\xi$ koja se naziva matematičko očekivanje slučajne promjenljive ξ. U slučaju kada je n veliko, $\frac{\nu_i}{n}\approx p_i$, zakon velikih brojeva, te u tom slučaju konstatujemo da je

$$E\xi\approx\frac{a_1+...+a_n}{n}.$$

Dakle, matematičko očekivanje slučajne promjenljive je približno jednako aritmetičkoj sredini realizacija slučajne promjenljive u dugoj seriji. Aritmetičke sredine realizacija iz nekoliko dugih serija će se grupisati oko matematičkog očekivanja.

Definicija 6.1 *Neka je ξ slučajna promjenljiva sa raspodjelom $P\{\xi = x_i\} = p_i$, $i = 1, 2, ..., k$.* **Matematičko očekivanje slučajne promjenljive** *ξ u oznaci*

$$E\xi := \sum_{i=1}^{k} x_i p_i.$$

Definicija 6.2 *Neka je ξ slučajna promjenljiva sa raspodjelom $P\{\xi = x_i\} = p_i$, $i = 1, 2, ...$.* **Matematičko očekivanje slučajne promjenljive** *ξ u oznaci*

$$E\xi := \sum_{i=1}^{\infty} x_i p_i,$$

ako ovaj red apsolutno konvergira. Ako red ne konvergira apsolutno, tada kažemo da slučajna promjenljiva ξ nema matematičko očekivanje.

Objasnimo zbog čega se zahtijeva apsolutna konvergencija reda. Prisjetimo se jednog detalja iz matematičke analize.

Primjer 6.1 *Ako u razvoju $\ln(1+x) = -\sum_{n=1}^{\infty} \frac{(-x)^n}{n}$, $-1 < x \leqslant 1$, stavimo $x = 1$, dobijamo*

$$\ln 2 = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = (1 - \frac{1}{2}) + (\frac{1}{3} - \frac{1}{4}) + \cdots = \frac{1}{2} + \frac{1}{12} + \cdots$$

Potražimo

$$(1 - \frac{1}{2} - \frac{1}{4}) + (\frac{1}{3} - \frac{1}{6} - \frac{1}{8}) + \cdots = \frac{1}{4} + \frac{1}{24} + \cdots = \frac{\ln 2}{2}.$$

Znači, u redu $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$ smo ispremještali članove (drugačije ih poređali, drugačije indeksirali) i na taj način smo dobili red koji konvergira ka drugoj vrijednosti. Ovu mogućnost pregrupisavanja članova u red čija je suma različita od sume polaznog reda, daju uslovno konvergentni redovi. Kod apsolutno konvergentnih redova, bez obzira na pregrupisavanje – preindeksiranje članova reda, suma reda je uvijek ista.◀

Ako red $\sum_{i=1}^{\infty} x_i p_i$ uslovno konvergira tada njegova suma zavisi od poretka članova. Zbog toga, za različite poretke članova, aritmetičke sredine realizacija se grupišu oko različitih brojeva što je u suprotnosti sa konceptom matematičkog očekivanja kao jedinstvenog broja oko kog se gupišu aritmetičke sredine realizacija. Apsolutno konvergencija obezbjeđuje nezavisnost sume od poretka članova te nema anomalije o kojoj smo govorili.

Iz definicije (6.1) je jasno da svaka prosta slučajna promjenljivaa ima očekivanje (često se prefiks matematičko izostavlja).

Jednostavno se dokazuju sljedeća svojstva matematičkog očekivanja diskretnih slučajnih promjenljivih. Sva navedena svojstva važe i u opštem slučaju.

S1) $|EX| \leqslant E|X|$.

S2) $Ec = c$, c je konstanta.

S3) Ako je $X \leqslant Y$, tada je $EX \leqslant EY$.

S4) $E(a_1X_1 + \cdots + a_nX_n) = a_1EX_1 + \cdots + a_nEX_n$.

S5) Neka su X i Y nezavisne diskretne slučajne promjenljive. Tada je $EXY = EXEY$.

Primjer 6.2 *Neka je $S_n : \mathcal{B}(n,p)$. Tada je*

$$ES_n = \sum_{k=0}^{n} k\binom{n}{k}p^k q^{n-k} = np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-1-(k-1))!}p^{k-1}q^{n-1-(k-1)}[l=k-1]$$
$$= np\sum_{l=0}^{n-1}\frac{(n-1)!}{l!(n-1-l)!}p^l q^{n-1-l} = np\sum_{l=0}^{n-1}\binom{n-1}{l}p^l q^{n-1-l} = np(p+q)^{n-1} = np. \blacktriangleleft$$

Primjer 6.3 *Naći očekivanu sumu brojeva koji se izvlače u igri LOTO (izvlači se 7 brojeva iz skupa $\{1,2,\ldots,39\}$).*

►Neka je X_1 prvoizvučeni broj, neka je X_2 drugooizvučeni broj itd. Imamo

$$X_i : \begin{array}{cccc} 1 & 2 & \ldots & 39 \\ \frac{1}{39} & \frac{1}{39} & \ldots & \frac{1}{39} \end{array},\ i = 1,2,\ldots 7.$$

Jednaka raspodijeljenost slučajnih promjenljivih X_i, $i = 1,2,\ldots,7$, je posljedica činjenice da je vjerovatnoća da recimo, broj 13 bude izvučen u prvom izvlačenju ista kao i vjerovatnoća da bude izvučen u drugom, trećem itd., dakle uvijek je $\frac{1}{39}$. Situacija je analogna onoj iz primjera 3.3. Tražena suma je $W = X_1 + \cdots + X_7$, pa je

$$EW = EX_1 + \cdots + EX_7 = 7 \cdot \frac{1}{39}(1 + 2 + \cdots + 39) = 140.$$

Zahvaljujući aditivnosti očekivanja, očekivanje slučajne promjenljive W smo našli bez prethodnog traženja raspodjele ove slučajne promjenljive.◄

Definicija 6.3 *Neka je X apsolutno neprekidna slučajna promjenljiva sa gustinom $g(x)$. Matematičko očekivanje slučajne promjenljive X u oznaci*

$$EX := \int_{-\infty}^{\infty} xg(x)\mathrm{d}x,$$

ako ovaj integral apsolutno konvergira. Ako integral ne konvergira apsolutno, tada kažemo da slučajna promjenljiva X nema matematičko očekivanje.

Zbog tijesne veze na relaciji sumiranje – integraljenje (sjetimo se da je integral uopštena suma) i veze na relaciji vjerovatnoća – funkcija gustine, ovakva definicija je bila očekivana.

Sljedeća teorema ima veliki teorijski i praktični značaj.

Teorema 6.1 Osnovna teorema za matematičko očekivanje. *Neka je X apsolutno neprekidna slučajna promjenljiva sa gustinom $g(x)$ i neka je $f : R \longrightarrow R$. Tada je*

$$Ef(X) = \int\limits_{-\infty}^{\infty} f(x)g(x)\mathrm{d}x.$$

Dakle, očekivanje slučajne promjenljive $Y = f(X)$ možemo dobiti bez prethodnog traženja njene funkcije raspodjele.

Primjer 6.4 $X : \mathcal{U}(a, b)$, $EX = \int\limits_a^b \frac{x}{a+b}\,\mathrm{d}x = \frac{a+b}{2}$. ◀

Primjer 6.5 $X : \mathcal{E}(\lambda)$, $EX = \int\limits_0^\infty \lambda x e^{-\lambda x}\mathrm{d}x = \frac{1}{\lambda}$. ◀

Primjer 6.6 $X : \mathcal{N}(m, \sigma^2)$, $EX = \frac{1}{\sqrt{2\pi\sigma^2}} \int\limits_{-\infty}^{\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}}\mathrm{d}x =$
$= \frac{m}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{-\frac{v^2}{2}}\mathrm{d}v = m$, *gdje je* $v = \frac{x-m}{\sigma}$. *Koristili smo poznati rezultat* $\int\limits_{-\infty}^{\infty} e^{-x^2}\mathrm{d}x = \sqrt{\pi}$. ◀

Zadržimo se za trenutak na slučajnim prmjenljivim

$$X : \begin{matrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{matrix} \quad \text{i} \quad Y : \begin{matrix} -100 & 100 \\ \frac{1}{2} & \frac{1}{2} \end{matrix}.$$

Očigledno, $EX = EY = 0$ iako su X i Y bitno različite slučajne promjenljive (i na ovom primjeru se vidi da je matematičko očekivanje manje informativno od funkcije raspodjele). Značenja slučajne promjenljive X su "bliska" očekivanoj vrijednosti, a značenja slučajne promjenljive Y su "daleko" od očekivane vrijednosti. Veličina koja mjeri stepen koncentracije slučajne promjenljive oko njenog očekivanja naziva se **disperzija**.

Definicija 6.4 *Disperzija DX slučajne promjenljive X se zadaje sa*

$$DX =: E(X - EX)^2.$$

Jednakost iz definicije disperzije potencira da je ova karakteristika mjera odstupanja slučajne promjenljive od njenog očekivanja. Neposredno računanje disperzije je ponekad jednostavnije sprovesti pomoću jedne od dvije jednakosti.

$$DX = E(X-EX)^2 = EX^2 - E(2XEX) + E(EX)^2 = EX^2 - 2EXEX + (EX)^2 = EX^2 - (EX)^2.$$

$$DX = EX(X-1) + EX - (EX)^2.$$

Evidentiraćemo svojstva disperzije.

a) $DX \geqslant 0$, $D(cX) = c^2 DX$, $D(X+c) = DX$, c je konstanta.

b) **Bijenemeova jednakost**: Ako su $X_1, \dots, X_n$ nezavisne slučajne promjenljive, tada je $D\left(\sum\limits_{i=1}^{n} X_i\right) = \sum\limits_{i=1}^{n} DX_i$. Zaista, zbog nezavisnosti imamo

$$D\left(\sum_{i=1}^{n} X_i\right) = E\left(\sum_{i=1}^{n}(X_i - EX_i)\right)^2 = E\sum_{i=1}^{n}(X_i - EX_i)^2 +$$
$$+ E\sum_{i\neq j}((X_i - EX_i)(X_j - EX_j)) = \sum_{i=1}^{n} E(X_i - EX_i)^2 +$$
$$\sum_{i\neq j} E((X_i - EX_i)(X_j - EX_j)) = \sum_{i=1}^{n} DX_i + \sum_{i\neq j}(EX_i - EX_i)(EX_j - EX_j) = \sum_{i=1}^{n} DX_i.$$

Koristimo: $(\sum\limits_{i=1}^{n} a_i)^2 = \sum\limits_{i=1}^{n} a_i^2 + \sum\limits_{i\neq j} a_i a_j$, u posljednjoj sumi je $n^2 - n$ sabiraka.

Indikator $I_A : \begin{matrix} 0 & 1 \\ 1-p & p \end{matrix}, p = P(A), q = 1-p.$

$I_A = I_A^2$, $EI_A = p$; $DI_A = EI_A^2 - (EI_A)^2 = p - p^2 = pq.$

Binomna raspodjela. $S_n : \mathcal{B}(n,p)$; $P\{S_n = k\} = \binom{n}{k} p^k q^{n-k}, k = 0, 1, ..., n.$

Pokažimo da je $S_n = I_{A_1} + I_{A_2} + ... + I_{A_n}$, gdje je $A_i, i = 1, 2, ..., n$ događaj da se u i-tom opitu ostvari uspjeh. S_n je broj ostvarenih uspjeha u seriji od n ponovljenih opita. Sumandi su jednaki 0 ili 1 u zavisnosti od toga da li je u odgovarajućem ponavljanju ostvaren neuspjeh ili uspjeh. Suma je jednaka broju sumanada koji su jednaki 1 tj. broju ostvarenih uspjeha. Dakle, jednakost je korektna. Događaji A_i su zbog nezavisnosti opita nezavisni, a samim tim su nezavisni i odgovarajući indikatori. I na kraju

$$ES_n = EI_{A_1} + EI_{A_2} + ... + EI_{A_n} = np.$$

$$DS_n = DI_{A_1} + DI_{A_2} + ... + DI_{A_n} = npq.$$

U analizi koja slijedi, koristićemo:

$$a)\ \sum_{k=1}^{\infty} kx^{k-1} = \big(\sum_{k=1}^{\infty} x^k\big)' = \left(\frac{x}{1-x}\right)' = \frac{1}{(1-x)^2},\ -1 < x < 1.$$

$$b)\ \sum_{k=2}^{\infty} k(k-1)x^{k-2} = \big(\sum_{k=2}^{\infty} x^k\big)'' = \left(\frac{x^2}{1-x}\right)'' = \frac{2}{(1-x)^3},\ -1 < x < 1.$$

Geometrijska raspodjela. $X : \mathcal{G}(p)$; $P\{X = k\} = q^k p, k = 0, 1, 2, ...$

$EX = \sum\limits_{k=0}^{\infty} kpq^k = pq \sum\limits_{k=1}^{\infty} kq^{k-1} = \frac{q}{p}$.

$DX = EX(X-1) + EX - (EX)^2 = pq^2 \sum\limits_{k=2}^{\infty} k(k-1)q^{k-2} + \frac{q}{p} - \frac{q^2}{p^2} = \frac{q}{p^2}$.

Negativna binomna raspodjela. $X : \mathcal{B}i(r, p), r \geqslant 2$;

$$P\{X = k\} = \binom{k+r-1}{r-1} q^k p^r, k = 0, 1, 2, ...$$

$X = X_1 + X_2 + ... + X_r$, gdje je $X_i, 1 \leqslant i \leqslant r$, broj neuspjeha od ostvarivanja $i-1$-og uspjeha do ostvarivanja i tog uspjeha. Promjenljive X_i su nezavisne i svaka ima $\mathcal{G}(p)$ raspodjelu pa je $EX = \frac{rq}{p}$, $DX = \frac{rq}{p^2}$.

Hipergeometrijska raspodjela. $X : H(n, m, k)$

$$P\{X = l\} = \frac{\binom{n}{l}\binom{m}{k-l}}{\binom{n+m}{k}}, \max\{0, k-m\} \leqslant l \leqslant \min\{n, k\}.$$

$$X = I_{A_1} + I_{A_2} ... + I_{A_k};\ I_{A_j} : \begin{array}{cc} 0 & 1 \\ \frac{m}{n+m} & \frac{n}{m+n} \end{array};\ I_{A_i}I_{A_j} : \begin{array}{cc} 0 & 1 \\ .... & \frac{\binom{n}{2}}{\binom{m+n}{2}} \end{array},\ i \neq j,$$

$A_j, 1 \leqslant j \leqslant k$ je događaj da se u j-tom izvlačenju izvuče bijela kuglica. Imamo:

$EX = \frac{kn}{m+n}$.

$EX^2 = E\big(\sum\limits_{i=1}^{k} I_{A_i}\big)^2 = E\sum\limits_{i=1}^{k} I_{A_i}^2 + E\sum\limits_{i\neq j} I_{A_i}I_{A_j} = \sum\limits_{i=1}^{k} EI_{A_i}^2 + \sum\limits_{i\neq j} EI_{A_i}I_{A_j} =$

$= \frac{kn}{m+n} + (k^2 - k)\frac{\binom{n}{2}}{\binom{m+n}{2}}$.

$DX = \frac{kmn(m+n-k)}{(m+n)^2(m+n-1)}$.

Puasonova raspodjela. $X : \mathcal{P}(\lambda)$; $P\{X = k\} = \frac{\lambda^k}{k!}e^{-\lambda}, k = 0, 1, 2, ...$

$EX = \sum\limits_{k=1}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \lambda \sum\limits_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda} = \lambda e^{-\lambda} \sum\limits_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda.$

$DX = EX(X-1) + EX - (EX)^2 = \sum\limits_{k=2}^{\infty} k(k-1)\frac{\lambda^k}{k!}e^{-\lambda} + \lambda - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$

Primjer 6.7 $X : \mathcal{U}(a,b)$, $DX = \int\limits_a^b \frac{x^2}{b-a}\,\mathrm{d}x - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$. ◀

Primjer 6.8 $X : \mathcal{E}(\lambda)$, $DX = \int\limits_0^{\infty} \lambda x^2 e^{-\lambda x}\mathrm{d}x - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$. ◀

Primjer 6.9 *Neka je* $X : \mathcal{N}(m, \sigma^2)$. *Imamo*

$$DX = E(X-m)^2 = \int\limits_{-\infty}^{\infty} \frac{(x-m)^2}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}\mathrm{d}x = \sigma^2. ◀$$

U slučaju $m = 0, \sigma^2 = 1$ *koristi se oznaka* $X^* : \mathcal{N}(0,1)$ *i čita: slučajna promjenljiva* X^* *ima* **standardnu normalnu raspodjelu.**

Primjer 6.10 *Iz špila od 52 karte slučajno se po modelu bez vraćanja izvlači 10 karata. Naći EW i DW, gdje je W broj izvučenih asova.*

$W : H(4, 48, 10)$.

Primjer 6.11 *Deset osoba ulazi u lift u prizemlju sedmospratne zgrade. Neka je* W *broj zaustavljanja lifta do izlaska svih putnika. Naći EW i DW.*

Neka je $A_i, i = 1,2,...,7$ događaj lift se zaustavlja na i-tom spratu.

$W = I_{A_1} + ...I_{A_7}$. $I_{A_i} : \begin{matrix} 0 & 1 \\ \frac{6^{10}}{7^{10}} & 1 - \frac{6^{10}}{7^{10}} \end{matrix}$

Primjer 6.12 *U kutiji se nalaze listice na kojima su zapisani brojevi od* 1 *do* 20. *Iz kutije se po modelu sa vraćanjem izvlači* 10 *listica. Neka je* X *broj zabilježenih brojeva. Naći* EX *i* DX.

$A_i, i = 1,2,...,20$, među izvučenim brojevima postoji broj i. $X = I_{A_1} + ... + I_{A_{20}}$.

$I_{A_i} : \begin{matrix} 0 & 1 \\ \frac{19^{10}}{20^{10}} & 1 - \frac{19^{10}}{20^{10}} \end{matrix}$

Primjer 6.13 *Deset pisama se nasumce stavlja u deset koverata. Neka jw* W *broj osoba koje dobiju odgovarajuće pismo. Izračunati* EW *i* DW.

$A_i, i = 1, 2, ..., 10$, i-ta osoba je dobila svoje pismo. $W = I_{A_1} + ... + I_{A_{10}}$.

$I_{A_i} : \begin{array}{cc} 0 & 1 \\ \frac{9}{10} & \frac{1}{10} \end{array}$; $I_{A_i} I_{A_j} : \begin{array}{cc} 0 & 1 \\ 1 - \frac{1}{9 \cdot 10} & \frac{1}{9 \cdot 10} \end{array}$

Primjer 6.14 *Dvadeeset kuglica se stavlja u deset kutija. Neka je W broj slobodnih kutija. Izračunati EW i DW.*

$A_i, i = 1, 2, ..., 10$, i-ta kutija je slobodna. $W = I_{A_1} + ... + I_{A_{10}}$.

$I_{A_i} : \begin{array}{cc} 0 & 1 \\ 1 - \frac{9^{20}}{10^{20}} & \frac{9^{20}}{10^{20}} \end{array}$

Teorema 6.2 (Čebišovljeva nejednakost.) *Ako slučajna promjenljiva X ima EX^2, tada je za svako $\varepsilon > 0$*

$$P\{|X| > \varepsilon\} \leqslant \frac{EX^2}{\varepsilon^2}.$$

Dokaz. Definišimo slučajnu promjenljivu

$$Y = \begin{cases} 0, & \text{ako je } |X| \leqslant \varepsilon \\ \varepsilon, & \text{ako je } |X| > \varepsilon \end{cases}.$$

Očigledno, $Y \leqslant |X| \Longrightarrow Y^2 \leqslant X^2$, pa je

$$EX^2 \geqslant EY^2 = \varepsilon^2 P\{|X| > \varepsilon\}. \blacklozenge$$

Posljedica 1. $P\{|X - EX| > \varepsilon\} \leqslant \frac{DX}{\varepsilon^2}$, za svako $\varepsilon > 0$.

Teorema 6.3 (Bernulijev zakon velikih brojeva.) *Neka je $S_n : \mathcal{B}(n, p)$. Tada*

$$\lim_{n \to \infty} P\left\{\left|\frac{S_n}{n} - p\right| > \varepsilon\right\} = 0 \tag{6.1}$$

Dokaz. Budući da je $E\frac{S_n}{n} = \frac{np}{n} = p$, $D\frac{S_n}{n} = \frac{npq}{n^2} = \frac{pq}{n}$, nakon primjene posljedice 1 dobijamo

$$P\left\{\left|\frac{S_n}{n} - p\right| > \varepsilon\right\} \leqslant \frac{D\frac{S_n}{n}}{\varepsilon^2} = \frac{pq}{n\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \leqslant \frac{1}{4n\varepsilon^2} \longrightarrow 0, \, n \to \infty$$

i tvrđenje implicira teorema o policajcima.$\blacklozenge$

Komentar. Iz iskustva znamo da u opitu u kome n puta bacamo novčić, n je veliko, relativna učestalost palih pisama ne odstupa "mnogo" od $\frac{1}{2}$. Da li navedena konstatacija važi za svaku seriju bacanja? Recimo, u seriji u kojoj u četvrtini bacanja padne pismo, relativna učestalost pisama je

$\frac{1}{4}$ te "mnogo" odstupa od $\frac{1}{2}$. Polaznu konstataciju treba preciznije formulisati. Nova formulacija će dati dublji smisao laičkoj interpretaciji sa početka komentara.

Neka je ϵ proizvoljan pozitivan broj i

$$W = \left\{ \omega : \left| \frac{S_n(\omega)}{n} - \frac{1}{2} \right| > \epsilon \right\}.$$

W je događaj da u seriji od n bacanja novčića, relativna frekvencija palih pisama odstupi od $\frac{1}{2}$ više od ϵ. Kako su serije jednako vjerovatne, $P(W)$ je količnik broja serija koje odgovaraju W i 2^n. Bernuli ne računa $P(W)$ već dobija ocjenu

$$P\left\{ \omega : \left| \frac{S_n(\omega)}{n} - \frac{1}{2} \right| > \epsilon \right\} \leqslant \frac{1}{4n\epsilon^2}.$$

Ocjena implicira: ako broj bacanja novčića teži beskonačnosti, tada vjerovatnoća događaja da relativna frekvencija palih pisama odstupi od $\frac{1}{2}$ više od datog broja ϵ teži 0. Upravo rečeno je precizna formulacija Bernulijevog zakona velikih brojeva. Primjenom ocjene se rješava zadatak: naći n tako da je za zadate pozitivne ϵ i ζ

$$P\left\{ \omega : \left| \frac{S_n(\omega)}{n} - \frac{1}{2} \right| > \epsilon \right\} \leqslant \zeta.$$

Rezultat se dobija nakon rješavanja nejednačine $\frac{1}{4n\epsilon^2} \leqslant \zeta$.

Uskoro ćemo naći bolju ocjenu za $P\left\{ \mid \frac{S_n}{n} - p \mid > \epsilon \right\}$.

Stepen zavisnosti između slučajnih promjenljivih se računa pomoću koeficijenta korelacije.

Definicija 6.5 **Kovarijacija** *između slučajnih promjenljivih* X *i* Y *je* $cov(X,Y) := EXY - EXEY$, *a* **koeficijent korelacije** *između slučajnih promjenljivih* X *i* Y *je*

$$\rho_{X,Y} := \frac{cov(X,Y)}{\sqrt{DXDY}} = \frac{EXY - EXEY}{\sqrt{DXDY}}.$$

Pokazuje se da je $|\rho_{X,Y}| \leqslant 1$ i $\rho_{X,Y} = \pm 1$ ako i samo ako je $Y = aX + b$. Takođe, ako su slučajne promjenljive X i Y nezavisne, tada je $EXY = EXEY$, te je $\rho_{X,Y} = 0$. Obrnuto, kao što pokazuje sljedeći primjer, ne važi.

Primjer 6.15 *Novčić se baca dva puta. Pojavi pisma pridružujemo* 0 *poena, a pojavi grba jedan poen. Neka je* X *suma poena kod bacanja, a* Y *razlika poena kod drugog i prvog bacanja. Naći* $\rho_{X,Y}$.

▶Raspodjela vektora (X,Y) je data tabelom

$Y \diagdown X$	0	1	2
-1	0	$\frac{1}{4}$	0
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	0	$\frac{1}{4}$	0

$$X : \begin{matrix} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{matrix},\ Y : \begin{matrix} -1 & 0 & 1 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{matrix},\ XY : \begin{matrix} -1 & 0 & 1 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{matrix},$$

$$EX = 1,\ EY = 0, EXY = 0,\ \rho_{X,Y} = 0,\ P\{(X,Y) = (0,0)\} \neq P\{X = 0\}P\{Y = 0\}.$$

Dakle, X i Y su zavisne slučajne promjenljive iako je $\rho_{X,Y} = 0$. ◀

Primjer 6.16 *Kocka se baca n puta. Neka je X broj palih šestica, a Y broj palih jedinica. Naći $\rho_{X,Y}$.*

▶ $X = \sum_{i=1}^{n} I_{A_i}$, A_i je događaj da u i-tom bacanju padne šestica.

$Y = \sum_{j=1}^{n} I_{B_j}$, B_j je događaj da u j-tom bacanju padne jedinica.

$EX = EY = \frac{1}{6} \cdot n,\ DX = DY = \frac{5}{36} \cdot n,$

$$\begin{aligned} EXY &= E\sum_{i=1}^{n}\sum_{j=1}^{n} I_{A_i} I_{B_j} = \sum_{i=1}^{n} EI_{A_i} I_{B_i} + \sum\sum_{i \neq j} EI_{A_i} I_{B_j} \\ &= \sum\sum_{i \neq j} EI_{A_i} EI_{B_j} = \frac{n^2 - n}{36}, \end{aligned}$$

jer su članovi u prvoj sumi jednaki nuli (ne možemo istovremeno registrovati i 1 i 6), u drugu sumu ulazi $n^2 - n$ članova (odbacujemo one kod kojih je $i = j$), a vjerovatnoća da u i-tom bacanju padne šestica, a u j-tom jedinica je $\frac{1}{36}$. Nakon ove analize ostaje da objedinimo podatke i zaključimo da je $\rho_{X,Y} = -\frac{1}{5}$. ◀

Rješenja primjera koji su označeni sa ■ su u dodatku crtezi.

Primjer 6.17 ■ *U jednakostraničnom trouglu ABC dužine stranice 1 slučajno se bira tačka T. Slučajna promjenljiva X je rastojanje od tačke T do stranice AB. Naći raspodjelu slučajne promjenljive X i EX.*

Primjer 6.18 ■ *U jednakostraničnom trouglu dužine stranice 1 slučajno se bira tačka T.*

a) Slučajna promjenljiva X je rastojanje od tačke T do najbliže stranice trougla.

b) Slučajna promjenljiva X je rastojanje od tačke T do najudaljenije stranice trougla.

Naći raspodjelu slučajne promjenljive X i EX.

Primjer 6.19 ■ *U pravilnoj kupi poluprečnika osnove R i visine H slučajno se bira tačka T. Neka je slučajna promjenljiva X rastojanje od tačke T do ose kupe. Naći raspodjelu slučajne promjenljive X i EX.*

Primjer 6.20 *Tačka T se slučajno bira u zarubljenoj kupi čije su osnove krugovi $K_1 = \{(x,y,z) : x^2+y^2 \leqslant 4,\ z=0\}$ i $K_2 = \{(x,y,z) : x^2+y^2 \leqslant 1,\ z=1\}$. Neka je W rastojanje tačke T od z ose. Naći raspodjelu slučajne promjenljive W i EW.*

Primjer 6.21 *Na intervalu $(0,1)$ slučajno se bira tačka T koja ga dijeli na dva intervala. Neka je W dužina intervala koji natkriva tačku $M(\frac{1}{3})$. Naći raspodjelu slučajne promjenljive W i EW.*

Primjer 6.22 *Na intervalu $(0,1)$ slučajno se bira tačka T. Neka je W rastojanje od tačke T do tačke $M(0,8)$. Naći raspodjelu slučajne promjenljive W i EW.*

7 Centralna granična teorema

Teorema 7.1 Klasična centralna granična teorema. *Za niz $X_n, n = 1,2,...$ nezavisnih, jednako raspodijeljenih slučajnih promjenljivih, $EX_k = a, DX_k = \sigma^2, S_n = X_1 + ... + X_n, k = 1,...,n$, važi:*

$$(\forall x \in \mathbf{R}) P\Big\{\frac{S_n - na}{\sqrt{n\sigma^2}} < x\Big\} \to \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}} dt, n \to \infty.$$

Budući da je $P\Big\{\frac{S_n-na}{\sqrt{n\sigma^2}} < x\Big\}$ funkcija raspodjele slučajne promjenljive $\frac{S_n-na}{\sqrt{n\sigma^2}}$, a $\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ funkcija raspodjele slučajne promjenljive $X^* : \mathcal{N}(0,1)$, tvrđenje teoreme može biti formulisano na sljedeći način

$$\frac{S_n - ES_n}{\sqrt{DS_n}} \xrightarrow{R} \mathcal{N}(0,1)$$

i čitamo: raspodjele niza slučajnih promjenljivih $\frac{S_n-ES_n}{\sqrt{DS_n}}$ konvergiraju ka standardnoj normalnoj raspodjeli $(m = 0, \sigma^2 = 1)$. CGT se može interpretirati i na sljedeći način: slučajna promjenljiva

$\frac{S_n - na}{\sqrt{n\sigma^2}}$ ima asimptotski standardnu normalnu raspodjelu tj. u slučaju velikih n-ova slučajna promjenljiva

$$\frac{S_n - na}{\sqrt{n\sigma^2}}$$

ima približno $\mathcal{N}(0,1)$ raspodjelu. Nakon dijeljenja sa n brojioca i imenioca u izrazu $\frac{S_n - na}{\sqrt{n\sigma^2}}$ dobijamo još jednu interpretaciju: u slučaju velikih n-ova slučajna promjenljiva

$$\left(\frac{S_n}{n} - a\right)\frac{\sqrt{n}}{\sigma}$$

ima približno $\mathcal{N}(0,1)$ raspodjelu.

POSLJEDICA 7.1 **Moavr-Laplasova teorema.** *Neka* $S_n : \mathcal{B}(n,p)$. *Tada*

$$(\forall x \in \mathbf{R}) P\left\{\frac{S_n - np}{\sqrt{npq}} < x\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt, n \to \infty,$$

$$tj \ \frac{S_n - np}{\sqrt{npq}} \xrightarrow{R} \mathcal{N}(0,1) \ ili \ \left(\frac{S_n}{n} - p\right)\sqrt{\frac{n}{pq}} \xrightarrow{R} \mathcal{N}(0,1).$$

Komentari.

1. CGT je robustna. Uslovi iz klasične CGT se mogu oslabiti a da tvrđenje teoreme važi. Slabljenje uslova se prevashodno odnosi na odbacivanje nezavisnosti i jednake raspodijeljenosti sabiraka. Nezavisnost se zamjenjuje uslovom slabe zavisnosti. CGT važi i za slučajne vektore.

2. CGT se primjenjuje u rješavanju zadataka u kojima treba odrediti $P\{S_n \in W\}, W \subset \mathbf{R}$. Koristi se činjenica da u slučaju velikog n, slučajna promjenljiva $\frac{S_n - ES_n}{\sqrt{DS_n}}$ ima približno $\mathcal{N}(0,1)$ raspodjelu. Zbog brze konvergencije, aproksimacija je odlična i za umjerene vrijednosti n.

3. Bez obzira na članove niza X_n (zbog jednostavnosti komentar vezujemo za klasičnu CGT), u slučaju velikog n slučajna promjenljiva $\frac{S_n - ES_n}{\sqrt{DS_n}}$ ima približno $\mathcal{N}(0,1)$ raspodjelu-interpretira se kao uniformnost tj. tvrđenje važi istovremeno za sve sume. Za sve nizove X_n za koje je $EX_k = a, DX_k = \sigma^2$ važi: za veliko n slučajna promjenljiva S_n ima približno $\mathcal{N}(na, n\sigma^2)$ raspodjelu. Iako slučajne promjenljive koje imaju isto očekivanje i disperziju mogu biti bitno različite, njihovi zbirovi (broj članova sume je velik) su u vjerovatnosnom smislu identični-imaju istu raspodjelu. Sve različitosti su nestale.

4. Ako su $X_1, X_2, ..., X_n$ nezavisne slučajne promjenljive i svaka ima $\mathcal{N}(m, \sigma^2)$ raspodjelu, tada $\frac{S_n}{n} : \mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$ što je ekvivalentno sa $\left(\frac{S_n}{n} - m\right)\frac{\sqrt{n}}{\sigma} : \mathcal{N}(0,1), S_n = X_1 + ...X_n$. Vidimo da je nakon osrednjavanja sačuvana normalna raspodjela, stim da je disperzija znatno manja od disperzije sumanada. Zaključujemo da su realizacije $\frac{S_n}{n}$ skoncetrisane oko m. CGT tvrdi da analogan

rezultat važi i u opštem slučaju. Naime, ako su $X_1, ..., X_n$ nezavisne i jednako raspodijeljenje slučajne promjenljive, $EX_1 = m, DX_1 = \sigma^2$, tada za veliko n, slučajna promjenljiva $\frac{S_n}{n}$ ima približno $\mathcal{N}\big(m, \frac{\sigma^2}{n}\big)$ raspodjelu što je ekvivalentno sa $\Big(\frac{S_n}{n} - m\Big)\frac{\sqrt{n}}{\sigma} : \mathcal{N}(0,1), S_n = X_1 + ...X_n$.

Primjer 7.1 *Kocka se baca* 100 *puta. Kolika je vjerovatnoća da je zbir palih brojeva između* 340 *i* 370*?*

►Sa X_k, $k = 1, 2, \ldots, 100$, označimo rezultat k-tog bacanja. Jasno

$$X_k : \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{matrix}, \; EX_k = 3,5, \; DX_k = \frac{35}{12}, \; k = 1, 2, \ldots, 100.$$

Zbir palih brojeva je $S_{100} = \sum\limits_{k=1}^{100} X_k$.

Niz vjerovatnoća u CGT brzo konvergira, 100 je dovoljno velik broj pa je opravdano tražiti rezultat kao što to radimo u redovima koji slijede. Iako koristimo aproksimaciju, zbog njene izuzetne preciznosti, nakon primjene CGT nećemo upotrijebiti simbol za približnu vrijednost već simbol jednakosti.

$$P\{340 < S_{100} < 370\} = P\left\{\frac{340-350}{\sqrt{\cdot}} < \frac{S_{100} - 100 \cdot 3,5}{\sqrt{100 \cdot \frac{35}{12}}} < \frac{370-350}{\sqrt{\cdot}}\right\}$$
$$= \Phi^*(1,17) + \Phi^*(0,59) = 0,379 + 0,222 = 0,601. \blacktriangleleft$$

Primjer 7.2 *Informacionim kanalom se prenose binarni nizovi. Zbog prisustva bijelog šuma u kanalu, vjerovatnoća pravilnog prijema odaslanog znaka je* 0,55*. Zbog toga se svaki znak šalje* n *puta, a odluka o poslanom znaku se donosi nakon utvrđivanja znaka koji se češće javlja. Naći najmanje* n *za koje je vjerovatnoća pravilne odluke* > 0.99.

►Smatraćemo da je uspjeh pravilno primljeni znak. Kod nas je $p = 0,55$, $q = 0,45$, dok $S_n : \mathcal{B}(n,p)$ predstavlja broj pravilno primljenih znakova. Pravilnu odluku donosimo ako je $S_n > \frac{n}{2}$. Sada imamo

$$P\left\{S_n > \frac{n}{2}\right\} = P\left\{\frac{S_n - 0,55n}{\sqrt{\cdot}} > \frac{0,5n - 0,55n}{\sqrt{0,55 \cdot 0,45n}}\right\} = 1 - \Phi\left(-\frac{0,05\sqrt{n}}{\sqrt{0,2475}}\right)$$
$$= 0,5 + \Phi^*\left(\frac{0,05\sqrt{n}}{\sqrt{0,2475}}\right) > 0,99 \Longrightarrow \frac{0,05\sqrt{n}}{\sqrt{0,2475}} > 2,33 \Longrightarrow n = 537. \blacktriangleleft$$

Primjer 7.3 *Strijelac pogađa metu sa vjerovatnoćom* 0,4 *i gađa u nju* 150 *puta. Naći bar jedan interval u kojem će se sa vjerovatnoćom* $> 0,8$ *nalaziti broj pogodaka.*

▶Slučajna promjenljiva $S_{150} : \mathcal{B}(150; 0,4)$ predstavlja broj pogodaka. Tražićemo interval oblika (A, B).

$$P\{A < S_{100} < B\} = P\left\{\frac{A-60}{6} < \frac{S_{150}-60}{\sqrt{150\cdot 0,4\cdot 0,6}} < \frac{B-60}{6}\right\}$$
$$= 0,8 \Longrightarrow \frac{A-60}{6} = -1,28,\ \frac{B-60}{6} = 1,28 \Longrightarrow A = 52,32,\ B = 67,68$$

te je zbog cjelobrojnosti broja pogodaka traženi interval $(52, 68)$. Interval je tim bolji što je kraći. Cilj dobijanja najkraćeg intervala se ostvaruje izborom simetričnog intervala $\left(\frac{A-60}{6}; \frac{B-60}{6}\right)$. Naime, od svih intervala za koje je vjerovatnoća da u njih "upadne" X^* konstantna, najkraći je simetričan interval. ◀

Primjer 7.4 *Koliko puta treba baciti novčić pa da sa vjerovatnoćom* $> 0,95$ *odstupanje relativne učestalosti padanja pisma od* $\frac{1}{2}$ *bude* $\leqslant 0,02$*?*

▶ Muavr–Laplasova teorema nam daje mogućnost da nađemo n za koje je sa vjerovatnoćom $> \alpha$ (prirodno je raditi sa α koje je blisko 1) odstupanje relativne frekvencije posmatranog događaja od vjerovatnoće tog događaja manje od nekog malog ε. Sa dobijenim rezultatom sebi možemo dati za pravo da u nekom smislu predviđamo budućnost. Naime, kad u nekom konkretnom vjerovatnosnom zadatku nađemo vjerovatnoću p nekog događaja A, a zatim nađemo i gore pomenuto n, tada možemo sa velikom dozom uvjerenosti tvrditi da će nakon n ponavljanja opita, relativna frekvencija događaja A biti blizu broja p.

Potražimo n u konkretnom slučaju.

$$P\left\{\left|\frac{S_n}{n} - 0,5\right| \leqslant 0,02\right\} = P\left\{\left|\frac{S_n}{n} - 0,5\right| 2\sqrt{n} \leqslant 0,04\sqrt{n}\right\} = 2\Phi^*(0,04\sqrt{n})$$
$$> 0,95 \Longrightarrow 0,04\sqrt{n} > 1,96 \Longrightarrow n > 2401.$$

Korišćenjem nejednakosti Čebišova dobijamo grubu aproksimaciju broja n.

$$0,95 < P\left\{\left|\frac{S_n}{n} - 0,5\right| \leqslant 0,02\right\} \Longrightarrow P\left\{\left|\frac{S_n}{n} - 0,5\right| > 0,02\right\} \leqslant 0,05$$
$$\Longrightarrow \frac{1}{4n(0,02)^2} \leqslant 0,05 \Longrightarrow n > 12500.$$ ◀

8 Uvod u statistiku

Razliku između vjerovatnosnih i statističkih zadataka ćemo ilustrovati konkretnim primjerom.

Vjerovatnosni zadatak. U kutiji se nalaze kuglice na kojima su zapisani brojevi $1, 2, ..., 10$. Iz kutije se po modelu sa vraćanjem izvlači 5 kuglica. Kolika je vjerovatnoća događaja A da se bar jednom izvuče kuglica sa brojem 10? $P(A) = 1 - \left(\frac{9}{10}\right)^{10}$.

Statistički zadatak. U kutiji se nalaze kuglice na kojima su zapisani brojevi od 1 do N, N je nepoznato. Po modelu sa vraćanjem izvučeno je 5 kuglica i na njima su zabilježeni brojevi $x_1, x_2, ..., x_5$. Nameće se zadatak koji je iz domena statistike i glasi: kako na osnovu registrovanih podataka ocijeniti nepoznato N? Metod maksimalne vjerodostojnosti rješava zadatak i za ocjenu N predlaže $\max\limits_{1\leqslant k\leqslant 5} x_k$. Ovakav izbor ocjene određuje konstatacija da je od svih modela, u modelu $N = \max\limits_{1\leqslant k\leqslant 5} x_k$ najveća vjerovatnoća da se registruju brojevi koje smo registrovali. Logika je, popularno govoreći, opredijeli se za model u okviru koga je najrealnije da se ostvari ono što se ostvarilo. Napravite analizu ako su zabilježeni podaci, recimo $4, 1, 3, 4, 6$. Vjerovatnoća ove petorke u modelu $N = 6$ je $\frac{1}{(6)^5}$, u modelu $N = 7$ je $\frac{1}{(7)^5}$ itd.◀

U vjerovatnoći se u zadatom vjerovatnosnom modelu traže vjerovatnoće zadatih događaja. Statistika rješava u izvjesnom smislu inverzan zadatak. U statistici se, između ostalog, razrađuju metodi kojima se na osnovu sakupljenih podataka traži adekvatan vjerovatnosni model koji te podatke produkuje. Rješavanje svih statističkih zadataka podrazumijeva sakupljanje podataka. Stoga se nerijetko može čuti da je statistika nauka koja razrađuje postupke zaključivanja na osnovu podataka.

Osnovne statističke pojmove ćemo približiti na jednostavnom modelu.

Model kutije sa listicama. U kutiji se nalazi N listica, N je poznato, na kojima su zapisani različiti i nama nepoznati brojevi $w_1 < w_2 < ... < w_N$. Neka je X broj na slučajno izvučenoj listici. Slučajnu promjenljivu X nazivamo obilježje, broj na listici obilježava listicu. Jasno,

$$* \; X : \begin{matrix} w_1 & w_2 & ... & w_N \\ \frac{1}{N} & \frac{1}{N} & ... & \frac{1}{N}. \end{matrix}$$

Kako brojeve na listicama ne znamo, matematička formalizacija nas obavezuje da konstatujemo: obilježje X ima raspodjelu koja pripada familiji raspodjela zadatih sa $*$, odgovaraju sve N-torke $w_1, ..., w_N, w_1 < ... < w_N$. Iz kutije po modelu sa vraćanjem vadimo n listica. Neka je slučajna promjenljiva X_1 broj na listici koju vadimo u prvom izvlačenju, X_2 u drugom,..., X_n u n-tom. Slučajne promjenljive $X_1, ..., X_n$ su nezavisne i raspodijeljene kao X. Slučajni vektor $(X_1, X_2, ..., X_n)$ nazivamo prost slučajni uzorak (random sample), karakterišu ga nezavisnost i jednaka raspodijeljenost komponenti. Neka su realizacije slučajnih promjenljivih $X_1, ..., X_n$ tj. opservirane-registrovane vrijednosti obilježja na izvučenim listicama $x_1, ..., x_n$. Vektor $(x_1, ..., x_n)$ nazivamo realizovani uzorak.

Razmotrimo model u kome se listice izvlače bez vraćanja. Neka je X_1' broj na listici tj. obilježje listice koju vadimo u prvom izvlačenju, X_2' u drugom,..., X_n' u n-tom, $n \leqslant N$. Promjenljive $X_1', ..., X_n'$ su zavisne i raspodijeljene su kao obilježje X, provjeriti. Slučajni vektor $(X_1', ..., X_n')$ se naziva uzorak.◀

Znamo da je sakupljanje podataka prvi korak u realizaciji mnogih naučnih projekata. **Obilježje**, označimo ga sa X, je slučajna promjenljiva koja predstavlja numeričku karakteristiku nekog fenomena iz oblasi istraživanja. Navedimo nekoliko primjera. Iznos koji osiguravajuća kompanija isplaćuje klijentu po osnovu pretrpljene štete; broj uključivanja na neki web sajt u toku sata; maksimalna godišnja temperatura u Podgorici; koeficijent inteligencije stanovnika Crne Gore.

DEFINICIJA 8.1 *Neka obilježje X ima funkciju raspodjele $F(x)$.* **Prost slučajni uzorak (random sample)** *obima n iz raspodjele $F(x)$ je slučajni vektor $(X_1, X_2, ..., X_n)$, gdje su $X_1, ..., X_n$ nezavisne slučajne promjenljive sa istom funkcijom raspodjele $F(x)$.*

Za komponente $X_1, ..., X_n$ se kaže da su nezavisne kopije obilježja X. Prost slučajni uzorak je matematički model **statističkog eksperimenta** koji se sastoji od n nezavisnih registrovanja vrijednosti (opservacija, mjerenja) obilježja X. Mi ćemo ubuduće po pravilu raditi sa prostim slučajnim uzorkom i kratko ćemo ga nazivati **uzorkom**. Prije sprovođenja eksperimenta, vrijednosti koje ćemo dobiti su slučajne promjenljive $X_1, X_2, ..., X_n$, registrovane vrijednosti zavise od slučaja, u različitim serijama "kupljenja podataka" dobijamo različite podatke-rezultate. Kada je eksperiment sproveden dobijamo vektor $(x_1, x_2, ..., x_n)$-**realizovani uzorak**, brojevi $x_1, ..., x_n$ su registrovane vrijednsti. Ilustracija svega navedenog je primjer kutije sa listicama. Istu svrhu ima i primjer sa web sajtom, biće uskoro izložen.

Često se na osnovu predznanja o izučavanom fenomenu može pretpostaviti da raspodjela obilježja X pripada nekoj familiji $\mathfrak{R}$ koju nazivamo **familija dopustivih raspodjela**. Znamo za korespondenciju između raspodjele i funkcije raspodjele slučajne promjenljive te familiju dopustivih raspodjela identifikujemo sa familijom odgovarajućih funkcija raspodjele. Navedimo dva primjera. Prema Gausu, rezultat mjerenja fizičke veličine je slučajna promjenljiva $X : \mathcal{N}(m, \sigma^2)$. Gaus je analizom rezultata mnogih mjerenja raznih fizičkih veličina, dakle eksperimentalno, potvrdio tačnost svoje tvrdnje. Tvrđenje je teorijski potkrijepio sa CGT. Naime, $X = m + \epsilon$, m je stvarna i nepoznata (zato se mjeri) vrijednost fizičke veličine, a ϵ je greška mjerenja. $E\epsilon = 0$ ako ne postoji sistematska greška koja se obično vezuje za neispravni instrument. Atmosferski procesi, mehaničke turbulencije i mnogi drugi faktori "izbacuju" instrument iz stanja stabilnosti i instrument produkuje greške. Zbog brojnosti grešaka riječ je o modelu iz CGT. Stoga, je ukupna greška $\epsilon : \mathcal{N}(0, \sigma^2)$ i $X : \mathcal{N}(m, \sigma^2)$, parametar σ^2 govori o preciznosti mjerenja.

Mi ćemo raditi sa parametarskim familijama raspodjela, koristi se oznaka $\{F(x, \theta), \theta \in \Theta\}$, θ je parametar. **Θ je skup dopustivih vrijednosti parametra** ili kraće **parametarski skup**. Funkcije raspodjele iz familije generiše parametar θ. Parametar θ može biti ili broj ili vektor. U primjeru sa mjerenjem, parametarska familija je $\{\mathcal{N}(m, \sigma^2), m \in R, \sigma^2 > 0\}$ i parametar je vektor $\theta = (m, \sigma^2)$.

Primjer 8.1 **Slučajni brojevi.** *U modelu u kom se slučajno bira broj sa $(0, \theta), \theta > 0$, parametarska familija je $\{\mathcal{U}(0; \theta),\ \theta > 0\}$, parametar θ je broj, a parametarski skup $\Theta = (0, \infty)$. Podaci tj. slučajno* *izabrani brojevi (ili brojevi generisani programom koji modelira ravnomjernu raspodjelu) stižu do statističara kome je θ nepoznato i on nakon obrade podataka dobija ocjenu za θ.*

Primjer 8.2 **Web sajt.** *U ovom primjeru ćemo koristiti rezultate lijepe stohastičke analize u koju zbog nedostatka vremena ne ulazimo. Registruju se momenti uključivanja posjetilaca na neki web sajt. Neka je obilježje X broj uključivanja u toku sata. X ima Puasonovu raspodjelu i slučajne preomjenljive koje su jednake broju uključivanja na disjunktnim vremenskim intervalima su nezavisne. Dakle familija dopustivih raspodjela je $\{\mathcal{P}(\lambda), \lambda > 0\}$, a komponente uzorka $(X_1, X_2, ..., X_n)$ su slučajne promejnljive koje su jednake brojevima uključivanja na n uzastopnih vremenskih intervala dužine jednog sata.*

Primjer 8.3 *Obilježje je dužina "života" baterije automobila na električni pogon. Familija dopustivih raspodjela je eksponencijalna. Podatke dobijamo registrovanjem dužine života baterija koje za potrebe statističke analize pratimo.*

Teorijska razmatranja u statistici obično započinju sa: obilježje X ima raspodjelu koja pripada familiji $\{F(x, \theta), \theta \in \Theta\}$.

Osnovni zadatak matematičke statistike je da se na osnovu registrovanih podataka tj. statističkog eksperimenta zaključi što je moguće više o nepoznatoj raspodjeli obilježja.

DEFINICIJA 8.2 *Neka je $f : \mathrm{R}^n \to \mathrm{R}$ i $(X_1, ..., X_n)$ uzorak. Slučajna promjenljiva $V = f(X_1, ..., X_n)$ se naziva* **statistika** *ili* **uzoračka karakteristika**.

Ako je $(x_1, ..., x_n)$ realizovani uzorak, tada je $v = f(x_1, ..., x_n)$ **realizovana statistika** V.

Primjeri statistika.

Uzoračka sredina. $f(x_1, ..., x_n) = \frac{x_1+...+x_n}{n}, \overline{X}_n = \frac{X_1+...+X_n}{n}$. Neka je $EX = a, DX = \sigma^2$. Tada je $E\overline{X}_n = a, D\overline{X}_n = \frac{\sigma^2}{n}$. Znamo (CGT)da statistika $\frac{\overline{X}_n - a}{\sigma}\sqrt{n}$ u slučaju velikog n ima približno $\mathcal{N}(0,1)$. Takođe, ako je $X : \mathcal{N}(m, \sigma^2)$ tada $\frac{\overline{X}_n - m}{\sigma}\sqrt{n} : \mathcal{N}(0,1)$.

Uzoračka disperzija. $\overline{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}_n^2 + \overline{X}_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2$. $E\overline{S}_n^2 = E(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2) = EX^2 - \frac{1}{n^2}E(\sum_{i=1}^{n} X_i^2 + \sum_{i \neq j} X_i X_j) = EX^2 - \frac{1}{n}EX^2 - \frac{n^2-n}{n^2}(EX)^2 = \frac{n-1}{n}(EX^2 - (EX)^2) = \frac{n-1}{n}\sigma^2$.

Popravljena uzoračka disperzija. $\hat{S}_n^2 = \frac{n}{n-1}\overline{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\overline{X}_n^2$; $E\hat{S}_n^2 = \sigma^2$.

Kada komponente uzorka $(X_1, ..., X_n)$ poređamo po veličini dobijamo slučajne promjenljive $Y_1 \leqslant Y_2 \leqslant ... \leqslant Y_n$ koje se nazivaju **statistike poretka**. Specijalno, **uzorački minimum:** $Y_1 = \min\limits_{1 \leqslant i \leqslant n} X_i$, **uzorački maksimum:** $Y_n = \max\limits_{1 \leqslant i \leqslant n} X_i$.

9 Tačkasta ocjena

Neka teorijska tj. stvarna funkcija raspodjele F_X obilježja X pripada parametarskoj familiji $\{F(x, \theta), \theta \in \Theta\}$. Procedura ocjenjivanja nepoznatog parametra θ podrazumijeva izbor statistike $T = T(\mathbb{X})$ čijom se realizacijom $t = T(x_1, ..., x_n)$ aproksimira parametar θ. Statistika $T(\mathbb{X})$ i njena realizacija $t = T(x_1, ..., x_n)$ se nazivaju **ocjena parametra** θ. Ocjenom parametra θ dobijamo ocjenu nepoznate funkcije raspodjele, a samim tim i ocjenu nepoznate raspodjele obilježja.

Da bi ocjena bila "dobra", poželjno je da ima neke osobine. Ocjena T kojom se ocjenjuje nepoznati parametar $\theta \in \Theta$ je **centrirana-nepristrasna** ako je $ET = \theta$ za $\forall \theta \in \Theta$. Ako su T_1 i T_2 centrirane ocjene nepoznatog parametra θ i $DT_1 < DT_2$, tada je ocjena T_1 bolja.

Primjer 9.1 *$X : \mathcal{U}(0, \theta), \theta > 0$, θ je nepoznati parametar. Nepoznati parametar θ se ocjenjuje statistikama $T_1 = 2\overline{X}_n$ i $T_2 = \frac{n+1}{n} Y_n$.*

▶Kako je $X : \mathcal{U}(0, \theta)$, $\theta > 0$, $EX = \theta/2$, $DX = \theta^2/12$ pa je

$$ET_1 = \theta,\ DT_1 = \frac{4}{n} \cdot \frac{\theta^2}{12} = \frac{\theta^2}{3n}.$$

Znači, ocjena T_1 je centrirana.

$$\begin{aligned}
F_{Y_n}(v) &= P\left\{\max_{1 \leqslant k \leqslant n} X_k < v\right\} = P\{X_1 < v, \ldots, X_n < v\} \\
&= P\{X_1 < v\} \cdots P\{X_n < v\} = \left(\frac{v}{\theta}\right)^n, \ 0 < v < \theta, \\
F(v) &= 0,\ v < 0,\ \ F(v) = 1,\ v > \theta \text{ i gustina je} \\
g(v) =& F'(v) = \begin{cases} 0, & v \notin (0, \theta) \\ n\frac{v^{n-1}}{\theta^n}, & v \in (0, \theta) \end{cases}.
\end{aligned}$$

Iskoristili smo nezavisnost komponenti i činjenicu da je maksimalna komponenta manja od v akko je svaka komponenta manja od v.

Sada imamo

$$ET_2 = \frac{n+1}{n}\int_0^\theta vn\frac{v^{n-1}}{\theta^n}dv = \theta,$$
$$DT_2 = \left(\frac{n+1}{n}\right)^2\int_0^\theta v^2 n\frac{v^{n-1}}{\theta^n}dv - \theta^2 = \frac{\theta^2}{n(n+2)}.$$

Dakle, i ocjena T_2 je centrirana.

Za $n \geqslant 2$, $DT_2 < DT_1$ te je ocjena T_2 bolja od ocjene T_1. Zamislimo da je pokrenut program koji generiše slučajne brojeve sa $(0,\theta)$, konkretnu vrijednost za θ zna samo inicijator. Program je generisao brojeve $x_1, ..., x_n$. Nepoznato θ možemo ocijeniti sa $t_1 = 2\overline{x}_n$ ili sa $t_2 = \frac{n+1}{n}y_n$. Izloženi rezultati preferiraju ocjenu t_2. ◄

U sljedećem primjeru nećemo ocjenjivati parametar koji određuje raspodjelu već nepoznato očekivanje i disperziju obilježja.

Primjer 9.2 *Neka je X obilježje, $EX = a$, $DX = \sigma^2$, a i σ^2 su nepoznati brojevi. Kako je $E\overline{X}_n = a$ to je uzoračka sredina centrirana ocjena nepoznatog očekivanja obilježja. Kako je $E\hat{S}_n^2 = \sigma^2$ to je popravljena uzoračka disperzija centrirana ocjena nepoznate disperzije obilježja.*

Budući da u svim razmatranim slučajevima ocjenjujemo realizacijom ocjene-statistike tj. brojem odnosno tačkom, govorimo o **tačkastoj ocjeni**.

Metod maksimalne vjerodostojnosti

Metod maksimalne vjerodostojnosti daje postupak kojim se dobija "dobra" ocjena nepoznatog parametra. Ideju na kojoj se zasniva metod maksimalne vjerodostojnosti približimo primjerom.

Primjer. Ocjenjujemo nepoznati broj N riba u ribnjaku. Specijalnom mrežom izvadimo n riba, markiramo ih i vratimo u ribnjak. Sačekamo neko vrijeme da bi se markirane pomiješale sa nemarkiranim ribama, a zatim ponovo vadimo n riba. Označimo sa X slučajnu promjenljivu koja je jednaka broju markiranih među izvađenim. Imamo

$$P_N\{X = r\} = \frac{\binom{n}{r}\binom{N-n}{n-r}}{\binom{N}{n}},\ r = 0, 1, ..., n.$$

Pretpostavimo da je broj registrovanih markiranih riba r_0. Primijetimo da je $P_N\{X = r_0\}$ funkcija od N. Nađimo vrijednost argumenta za koju $P_N\{X = r_0\}$ dostiže maksimum. Ta vrijednost je cio broj najbliži broju $\frac{n^2}{r_0}$. Nepoznato N ocjenjujemo tim brojem. Ovako odabranom ocjenom odabran je model u kome je najveća vjerovatnoća da se registruje upravo r_0 izvađenih markiranih riba. U ovakav način zaključivanja je ugrađena logika: odlučiću se za model u kome je nejveća vjerovatnoća (tj. najrealnije je) da se desi ono što se desilo. Ako ocjenu označimo sa $\hat{N}$ imamo $\hat{N} \approx \frac{n^2}{r_0} \Rightarrow \frac{n}{\hat{N}} \approx \frac{r_0}{n}$ što je u saglasju sa intuicijom.◄

Neka je $\mathbb{X} = (X_1, ..., X_n)$ uzorak obilježja X čija funkcija raspodjele pripada familiji

$$\mathfrak{R} = \{F(x, \theta),\, \theta \in \Theta\}.$$

Definisaśimo funkciju vjerodostojnosti u slučaju kada raspodjela obilježja pripada familiji raspodjela diskretnog tipa $\mathfrak{R} = \{p(w_k, \theta),\, \theta \in \Theta\}$, $p(w_k, \theta) = P_\theta\{X = w_k\}$.

Definicija 9.1 **Funkcija vjerodostojnosti** $L(\theta, x_1, ..., x_n) := P_\theta\{(X_1, ..., X_n) = (x_1, ..., x_n)\} = P_\theta\{X_1 = x_1, ..., X_n = x_n\} = p(x_1, \theta) \cdots p(x_n, \theta)$, $\theta \in \Theta$, *je funkcija čiji je argument* θ, $(x_1, ..., x_n)$ *je neka realizacija uzorka obilježja* X *i brojevi* $x_1, ..., x_n$ *se tretiraju kao parametri.*

Iz definicije se vidi da je $L(\theta, x_1, ..., x_n)$ vjerovatnoća sa kojom se u modelu u kome je raspodjela obilježja određena sa θ dobija realizovani uzorak $(x_1, ..., x_n)$.

Definisaśimo sada funkciju vjerodostojnosti u slučaju kada raspodjela obilježja pripada familiji raspodjela apsolutno neprekidnog tipa $\mathfrak{R} = \{g(x, \theta),\, \theta \in \Theta\}$, g je gustina.

Definicija 9.2 **Funkcija vjerodostojnosti** $L(\theta, x_1, ..., x_n) := g(x_1, \theta) \cdots g(x_n, \theta)$, $\theta \in \Theta$, *je funkcija čiji je argument* θ, $(x_1, ..., x_n)$ *je neka realizacija uzorka obilježja* X *i brojevi* $x_1, ..., x_n$ *se tretiraju kao parametri.*

Neka funkcija $L(\theta, x_1, ..., x_n)$, $\theta \in \Theta$, dostiže maksimum za $\hat{\theta} = h(x_1, ..., x_n)$. Za ocjenu $\hat{\theta} = h(X_1, ..., X_n)$ kažemo da je **ocjena maksimalne vjerodostojnosti**, a za njenu realizaciju $\hat{\theta} = h(x_1, ..., x_n)$ kojom se konkretizuje ocjena nepoznatog parametra θ koristimo isti naziv. Identičnu oznaku $\hat{\theta}$ koristimo za statistiku i realizaciju te statistike. Ocjena maksimalne vjerodostojnosti nas vodi u model u kome je vjrovatnoća dobijanja podataka koje smo sakupili najveća.

Kada u raspodjeli obilježja figuriše eksponencijalna funkcija, umjesto traženja tačke maksimuma za $L(\theta, x_1, ..., x_n)$ jednostavnije je tražiti tačku maksimuma za $\ln L(\theta, x_1, ..., x_n)$. Dobija se isti rezultat jer zbog monotonosti funkcije ln funkcije $L(\theta, x_1, ..., x_n)$ i $\ln L(\theta, x_1, ..., x_n)$ maksimum dostižu u istoj tački.

Primjer 9.3 *a)* $X : \mathcal{N}(m, 1), m \in \mathbb{R}$. *b)* $X : \mathcal{N}(m, 1), m \geqslant m_0$.

►a) $L(m, x_1, ..., x_n) = \prod\limits_{k=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_k - m)^2}{2}} = (2\pi)^{-\frac{n}{2}} e^{-\frac{\sum_{k=1}^{n}(x_k - m)^2}{2}}$, $\frac{\partial}{\partial m} \ln L(m, x_1, ..., x_n) = \sum\limits_{k=1}^{n} (x_k - m)$, $\frac{\partial}{\partial m} \ln L = 0 \Leftrightarrow m = \overline{x}_n$. Dakle, $\hat{m} = \overline{X}_n$.

b) $\ln L(m, x_1, ..., x_n) = -\frac{n}{2} m^2 + m \sum\limits_{k=1}^{n} x_k - \sum\limits_{k=1}^{n} x_k^2 - \frac{n}{2} \ln(2\pi)$, $m \geqslant m_0$. Nakon traženja tačke u kojoj upravo dobijena kvadratna funkcija dostiže maksimum, dobijamo $\hat{m} = \max\{m_0, \overline{X}_n\}$. ◄

Primjer 9.4 $X : \mathcal{E}\Big(\frac{1}{\theta}\Big)$ *tj.* $g(x,\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}, x > 0, \theta > 0.$

$\blacktriangleright$ $L(\theta, x_1, ..., x_n) = \prod\limits_{k=1}^{n} \frac{1}{\theta}e^{-\frac{x_k}{\theta}} = \frac{1}{\theta^n}e^{-\frac{\sum\limits_{k=1}^{n} x_k}{\theta}}, \frac{\partial}{\partial\theta} \ln L(\theta, x_1, ..., x_n) = 0 \Leftrightarrow \theta = \overline{x}_n \Rightarrow \hat{\theta} = \overline{X}_n.$ $\blacktriangleleft$

10 Intervali povjerenja

Neka je $\mathbb{X} = (X_1, ..., X_n)$ uzorak obilježja X čija funkcija raspodjele pripada familiji

$$\mathfrak{R} = \{F(x,\theta),\ \theta \in \Theta\}.$$

Neka su $\theta_1 = \theta_1(X_1, ..., X_n)$ i $\theta_2 = \theta_2(X_1, ..., X_n)$ statistike takve da za $\forall\theta \in \Theta$ važi:

a) $P_\theta\{\theta_1 < \theta_2\} = 1,$

b) $P_\theta\{\theta_1 < \theta < \theta_2\} = \gamma.$

Tada se interval (θ_1, θ_2) naziva **interval povjerenja za nepoznati parametar** θ, a broj γ se naziva **nivo povjerenja**. Na osnovu zakona velikih brojeva, približno $100\gamma\%$ realizovanih intervala sadrži nepoznati parametar θ.

Koristićemo oznaku z_γ gdje je $\frac{\gamma}{2} = P\{0 < X^* < z_\gamma\} = \frac{1}{\sqrt{2\pi}} \int\limits_0^{z_\gamma} e^{-\frac{t^2}{2}}dt$. Uobičajene vrijednosti za γ su $0,90; 0,95; 0,99$. z_γ se nalazi u tablici normalne raspodjele.

Primjer 10.1 $X : \mathcal{N}(m, \sigma_0^2),\ m \in \mathbb{R}.$

$\blacktriangleright$Znamo, $\frac{\overline{X}_n - m}{\sigma_0}\sqrt{n} : \mathcal{N}(0,1),\ \forall m \in \mathbb{R}$. Imamo

$$P_m\Big\{ -z_\gamma < \frac{\overline{X}_n - m}{\sigma_0}\sqrt{n} < z_\gamma\Big\} = \gamma \Leftrightarrow P_m\left\{\overline{X}_n - \frac{\sigma_0}{\sqrt{n}}z_\gamma < m < \overline{X}_n - \frac{\sigma_0}{\sqrt{n}}z_\gamma\right\} = \gamma.$$

Zaključujemo da je $I = \Big(\overline{X}_n - \frac{\sigma_0}{\sqrt{n}}z_\gamma, \overline{X}_n + \frac{\sigma_0}{\sqrt{n}}z_\gamma\Big)$. Dužina intervala je $d(I) = 2\frac{\sigma_0}{\sqrt{n}}z_\gamma$. Pogledati crtež iz priloga Interval!$\blacktriangleleft$

Primjer 10.2 $X : \mathcal{E}\Big(\frac{1}{\theta}\Big)$ *tj.* $g(x,\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}, x > 0, \theta > 0.$

$\blacktriangleright$Koristimo asimptotsku normalnost promjenljive $(\overline{X}_n - \theta)\frac{\sqrt{n}}{\theta}$.

$$P_\theta\Big\{ -z_\gamma < (\overline{X}_n - \theta)\frac{\sqrt{n}}{\theta} < z_\gamma\Big\} = \gamma \Leftrightarrow P_\theta\left\{\frac{\overline{X}_n}{1 + \frac{z_\gamma}{\sqrt{n}}} < \theta < \frac{\overline{X}_n}{1 - \frac{z_\gamma}{\sqrt{n}}}\right\} = \gamma.$$

Interval povjerenja za nepoznato θ, n veliko, je

$$I = \left(\frac{\overline{X}_n}{1 + \frac{z_\gamma}{\sqrt{n}}} ; \frac{\overline{X}_n}{1 - \frac{z_\gamma}{\sqrt{n}}} \right). \blacktriangleleft$$

11 Testiranje statističkih hipoteza

Pojam statističke hipoteze i ideju testiranja približimo kroz jedan primjer.

Primjer 1. U kutiji se nalazi 10 kuglica i znamo da je kutija napunjena po jednoj od dvije strategije:

0) Sa 9 kuglica na kojima je broj 1 i jednom kuglicom na kojoj je broj 2.

1) Sa 9 kuglica na kojima je broj 2 i jednom kuglicom na kojoj je broj 1.

Dozvoljeno nam je da iz kutije izvučemo 4 kuglice po modelu sa vraćanjem. Na osnovu izvučenih kuglica tj. brojeva na njima, treba da se odlučimo za jednu od dvije hipoteze (pretpostavke, mogućnosti):

1^0 Kutija je napunjena po strategiji 0 – govorićemo o hipotezi H_0.

2^0 Kutija je napunjena po strategiji 1 – govorićemo o hipotezi H_1.

Postupak presuđivanja u korist jedne od hipoteza, tj. postupak prihvatanja jedne od hipoteza, zvaćemo **testom**. Ako je kutija napunjena po strategiji 0, tada je zbog velike vjerovatnoće pojave broja 1, vjerovatnoća da se registruje mala suma brojeva, velika. Ako je kutija napunjena po strategiji 1, tada je zbog velike vjerovatnoće pojave broja 2, vjerovatnoća da se registruje velika suma brojeva, velika. Jasno, suma brojeva je u rasponu od 4 do 8 i kada govorimo o maloj odnosno velikoj sumi imamo u vidu male odnosno velike vrijednosti u odnosu na interval omeđen brojevima 4 i 8. Nakon ove analize, nameće se kao razuman sljedeći postupak odlučivanja: Ako je zbir izvučenih brojeva $\geqslant 7$ prihvatićemo H_1, a H_0 odbaciti. U suprotnom tj. ako je zbir izvučenih brojeva < 7 prihvatićemo H_0, a H_1 odbaciti.

Označimo sa X obilježje koje predstavlja broj na izvučenoj kuglici. U slučaju kada je važeća strategija 0, obilježje X ima raspodjelu

$$X : \begin{array}{cc} 1 & 2 \\ \frac{9}{10} & \frac{1}{10} \end{array},$$

a u slučaju kada je važeća strategija 1, obilježje X ima raspodjelu

$$X : \begin{array}{cc} 1 & 2 \\ \frac{1}{10} & \frac{9}{10} \end{array}.$$

Ove dvije raspodjele možemo tretirati kao familiju raspodjela obilježja X. U uzorku (X_1, X_2, X_3, X_4) komponente predstavljaju redom prvi, drugi, treći i četvrti izvučeni broj. U statistici je uobičajeno da se hipoteze izražavaju u terminima raspodjela. U našem primjeru H_0 je hipoteza da obilježje X ima gornju, a H_1 donju raspodjelu.

Ako je (x_1, x_2, x_3, x_4) realizovani uzorak, tada se uslov "zbir izvučenih brojeva je $\geqslant 7$" zapisuje sa $x_1 + x_2 + x_3 + x_4 \geqslant 7$. Ovaj uslov generiše oblast

$$C = \{(x_1, x_2, x_3, x_4) : x_1 + x_2 + x_3 + x_4 \geqslant 7\},\ C \subset \mathbb{R}^4.$$

Nakon uvođenja oblasti C, pravilo odlučivanja možemo ovako formulisati: Ako realizovani uzorak (x_1, x_2, x_3, x_4) "upadne" u oblast C, tada prihvatamo H_1, a ako "upadne" u C^c, tada prihvatamo H_0. Oblast C određuje postupak – pravilo odlučivanja tj. test. "Upadanje" uzorka u oblast C ne odgovara hipotezi H_0 (tada imamo onu situaciju kada je zbir izvučenih brojeva veliki tj. među brojevima dominira 2; ostvaruje se događaj čija je vjerovatnoća realizacije mala ako je tačna hipoteza H_0). Zbog toga oblast C nazivamo kritična oblast za H_0.

U postupku odlučivanja nema izričitosti. Mi samo konstatujemo da na osnovu registrovanih brojeva prednost dajemo jednoj hipotezi (prihvatamo jednu hipotezu). Naravno, postoji mogućnost greške. Grešku pravimo kada povodeći se za pravilom odlučivanja odbacimo hipotezu koja je faktički tačna. Preciznije, moguće je da odbacimo H_0 koja je tačna i samim tim prihvatimo H_1 koja je netačna. Jasno, moguća je i situacija u kojoj H_0 i H_1 imaju zamijenjene uloge.

Izračunajmo vjerovatnoću α da odbacimo tačnu hipotezu H_0.

$$\alpha = P_{H_0}\{(X_1, X_2, X_3, X_4) \in C\} = \frac{1}{10^4} + 4 \cdot \frac{9}{10^4} = 0,0037.$$

Izračunajmo vjerovatnoću β da odbacimo tačnu hipotezu H_1.

$$\beta = P_{H_1}\{(X_1, X_2, X_3, X_4) \in C^c\} = \frac{1}{10^4} + 4 \cdot \frac{9}{10^4} + 6 \cdot \frac{9^2}{10^4} = 0,0523.$$

Interesantno je vidjeti šta se dešava ako promijenimo test na taj način što za kritičnu oblast za H_0 sada uzmemo

$$D = \{(x_1, x_2.x_3, x_4) : x_1 + x_2 + x_3 + x_4 \geqslant 8\},\ D \subset \mathbb{R}^4.$$

Lako se dobija $\alpha = 0,0001$, $\beta = 0,3439$. Primijetimo, vjerovatnoća da se odbaci faktički tačna H_0 je znatno smanjena, ali je vjerovatnoća odbacivanja faktički tačne H_1 postala enormno velika. Praktično, u jednom od tri slučaja ćemo odbaciti tačnu hipotezu H_1. U ovakvoj situaciji je bolje testiranje obaviti prvim postupkom.◀

Motivisani prethodnim primjerom, možemo preći na izlaganje teorije.

Neka je X obiležje čija funkcija raspodjele vjerovatnoća pripada familiji

$$\mathfrak{R} = \{F(x,\theta),\, \theta \in \Theta\}.$$

Pretpostavka oblika $H_0(\theta \in \Theta_0)$, $\Theta_0 \subset \Theta$, zove se **statistička hipoteza**. Malo drugačije rečeno, pretpostavka se sastoji u tome da obiležje X ima raspodjelu koja pripada užoj familiji određenoj parametarskim skupom Θ_0. Ako je Θ_0 jednočlani skup, tada kažemo da je hipoteza H_0 prosta. U protivnom govorimo o složenoj hipotezi. Hipotezi H_0 ćemo suprotstaviti hipotezu $H_1(\theta \in \Theta \backslash \Theta_0)$. Hipoteza H_o se naziva nulta, a hipoteza H_1 alternativna. Postupak odlučivanja u korist jedne hipoteze, tj. postupak prihvatanja jedne od hipoteza, na osnovu realizovanog uzorka se naziva **statistički test**. Taj postupak je određen zadavanjem kritične oblasti $C \subset \mathbb{R}^n$ i sprovodi se na sljedeći način: Ako $(x_1, ..., x_n) \in C$ tada H_0 odbacujemo u korist H_1, a ako $(x_1, ..., x_n) \in C^c$ tada H_1 odbacujemo u korist H_0. Zbog upravo izloženog se kaže da kritična oblast zadaje test.

Posvetimo se slučaju kada je $\Theta = \{\theta_0, \theta_1\}$, $H_0(\theta = \theta_0), H_1(\theta = \theta_1)$, dakle obje hipoteze su proste. Pretpostavimo da je $C \subset \mathbb{R}^n$ kritična oblast koja zadaje test.

Prilikom odlučivanja, postoji mogućnost da se napravi greška.

1^o **Grešku prve vrste** pravimo kada odbacimo faktički tačnu hipotezu H_0.

2^o **Grešku druge vrste** pravimo kada odbacimo faktički tačnu hipotezu H_1.

Vjerovatnoća greške prve vrste se označava sa α i za nju, na osnovu rečenog, važi

$$\alpha = P_{H_0}\{(X_1, \ldots, X_n) \in C\}.$$

α se naziva **pragom značajnosti testa**, a C se naziva **kritična oblast veličine** α.

Vjerovatnoća greške druge vrste se označava sa β i za nju, na osnovu rečenog, važi

$$\beta = P_{H_1}\{(X_1, \ldots, X_n) \in C^c\}.$$

Primjer 11.1 *Obiljежje X ima $\mathcal{N}(m,1)$, $m \in \{0,1\}$, radimo sa uzorkom obima $n = 9$. Za testiranje $H_0(m = 0)$ protiv $H_1(m = 1)$ koristi se $\mathcal{C} = \{(x_1, ..., x_9) : \overline{x}_9 \geqslant c\}$. Naći c ako je $\alpha = 0,05$, a zatim naći β.*

► $0,05 = P_0\{\overline{X}_9 \geqslant c\} = P_0\Big\{(\overline{X}_9 - 0)\frac{\sqrt{9}}{1} \geqslant (c-0)\frac{\sqrt{9}}{1}\Big\} = P\{X^* \geqslant 3c\} \Rightarrow 3c = 1,65 \Rightarrow c = 0,55.$
$\beta = P_1\{\overline{X}_9 < 0,55\} = P_1\{(\overline{X}_9 - 1)3 < (0,55-1)3\} = P\{X^* < -1,35\} = 0,09.$ ◄

Primjer 11.2 *$X : \mathcal{E}(\theta^{-1})$, $\theta \in \{\frac{2}{3}, 1\}$, radimo sa uzorkom obima $n = 100$. Za testiranje $H_0(\theta = 1)$ protiv $H_1(\theta = \frac{2}{3})$ koristi se $\mathcal{C} = \{(x_1, ..., x_{100}) : \overline{x}_{100} \leqslant c\}$. Naći c ako je $\alpha = 0,05$, a zatim naći β.*

► $c = 0,835$, $\beta = 0,0059$, koristimo asimptotsku normalnost statistike $\left(\overline{X}_n - \theta\right)\frac{\sqrt{n}}{\theta}$. ◄

email adresa nastavnika: sstamatovic@ucg.ac.me