

PREDGOVOR

U skripti su obrađene teme koje su naznačene programom predmeta "Teorija vjerovatnoća. Statistika", jednosemestralni kurs, a koji sam nekoliko godina predavao studentima Fizike. Iako je skripta prvenstveno namijenjena studentima Fizike, smatram da može služiti studentima Tehničkih fakulteta i postdiplomcima na Građevinskom i na Mašinskom fakultetu. I studenti Matematike u skripti mogu naći koristan materijal.

Studenti Fizike i studenti Tehnike imaju dobro matematičko obrazovanje, ali ne poznaju Teoriju mjere na koju se u Vjerovatnoći često pozivamo. Stoga sam, radeći na ovom tekstu, na pojedinim mjestima bio primoran da odustanem od matematičke strogosti i neke pojmove definišem manje formalno. Moj afinitet ka izlaganju teorije sa što više primjera prisutan je u ponuđenoj skripti. Najbolji test razumijevanja matematičke teorije je umješnost u rješavanju zadataka. U skripti je ponuđeno 140 zadataka i među njima ima mnogo ispitnih zadataka.

Veliku zahvalnost dugujem recenzentima prof. dr Milojici Jaćimoviću i docentu dr Svjetlani Terzić. Njihovi savjeti su mi pomogli u pripremi završne verzije teksta.

Podgorica, maj 2000. g.

Siniša Stamatović

Dio 1

Teorija vjerovatnoća

1.1 Uvod u predmet

Posmatrajmo jednostavni opit bacanja simetričnog novčića ili popularno nazvanu igru "pismo-grb". Opit se sastoji u tome da se novčić baci na sto. Nakon što je novčić pao na sto, na njegovoj je gornjoj strani pismo (P) ili grb (G) (ovdje dogovorno smatramo da su (P) i (G) jedini mogući ishodi; razumno je isključiti ostale ishode kakav je npr. padanje novčića na bok). Zbog simetričnosti novčića, mi prije njegovog bacanja, ne znamo hoće li "pasti" pismo ili grb. Dakle, uz date pretpostavke, bacanje novčića je opit sa dva moguća ishoda; ishodi su elementi dvočlanog skupa

$$\Omega = \{P, G\},$$

ali unaprijed ne znamo koji će ishod biti ostvaren.

Slično, bacanje kocke za igru je opit kod kojeg će svaki ishod biti jedan od elemenata skupa

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

ali unaprijed ne možemo reći koji će to element biti.

Naveli smo dva tipična slučajna opita. Pod slučajnim opitom (koristi se i sinonim eksperiment) podrazumijevamo takav opit kod koga ishodi nisu jednoznačno određeni uslovima u kojima se opit realizuje.

Teorija vjerovatnoća je matematička disciplina čiji je zadatak formiranje i proučavanje matematičkog modela slučajnog eksperimenta. Podvucimo, Teorija vjerovatnoća izučava matematički model slučajnog opita, a ne prirodu samog opita.

Opit je moćan metod za spoznaju prirode. U suštini, opitom treba da ustanovimo odnos između uzroka i posljedice. Poznavanje tog odnosa omogućava istraživaču da definiše uslove opita i predvidi rezultat u situacijama kad se uslovi realizuju. Opiti se dijele na dvije grupe: determinističke i slučajne. Dugo su predmet interesovanja naučnika bili isključivo deterministički opiti. Kod determinističkog opita, rezultat je jednoznačno određen uslovima opita. Navedimo dva primjera.

1. Dozvolimo tijelu da slobodno pada vrijeme t i izmjerimo pređeni put. Pređeni put je $s = \frac{gt^2}{2}$.

2. Hemijski čista voda koja se na lokalitetu čija je nadmorska visina 0, zagrije do 100^0C , iz tečnog stanja prelazi u paru. Dakle, ako se is-pune pomenuti uslovi, kao rezultat (i to uvijek) ćemo imati promjenu agregatnog stanja vode.

Kod slučajnog opita, ostvarivanje kompleksa uslova ne daje jednoznačan rezultat.

U Vjerovatnoći se izučavaju slučajni opiti kod kojih se kompleksi uslova mogu ponoviti u tom smislu da je ponovljeni opit identičan prethodnom. Bacanje novčića na sto možemo ponoviti proizvoljno (konačno) mnogo puta. Kada smo govorili o identičnosti opita, imali smo u vidu "bacanje simetričnog novčića na sto", dakle uslove u kojima se opit sprovodi, a ne i rezultate opita.

Kod matematičkog modeliranja slučajnog opita, potrebno se dogovoriti o tome što smatramo mogućim ishodom tog opita. Kada se to uradi, onda su ti dogovorno ustanovljeni mogući ishodi jedini objekti koji služe za konstrukciju matematičkog modela tog slučajnog opita.

Rečeno pojasnimo na jednom primjeru. Vratimo se opitu u kome se baca novčić na sto. Možemo se dogovoriti, kao što smo uradili na početku, da registrujemo gornju stranu novčića kada on padne na sto. Tada su ishodi (P) i (G). Međutim, ako je daska na stolu pravougaonik dimenzija a i b , možemo se dogovoriti da kao ishod registrujemo rastojanje centra novčića od najbliže ivice pravougaonika. Tada su mogući ishodi svi brojevi od 0 do $\frac{1}{2} \min\{a, b\}$. Prirodno, možemo se dogovoriti da kao ishod registrujemo i gornju stranu novčića i rastojanje centra novčića od najbliže ivice stola. Odluka o izboru mogućih ishoda je uslovljena prirodom problema kog izučavamo.

Rezultate slučajnog opita zvaćemo **događaji**. Na primjer, ako je slučajni opit bacanje dvije kocke, "padanje istog broja" je događaj. Mi ćemo ustanoviti da je ostvaren događaj "pao je isti broj" tako što ćemo ustanoviti da je pao jedan od ishoda "(1, 1) ili (2, 2) ili (3, 3) ili (4, 4) ili (5, 5) ili (6, 6)".

Razlikovaćemo složene i elementarne događaje. U prethodnom primjeru, događaj pali su isti brojevi je složen, a događaji (1, 1) (u oba bacanja je pala jedinica), (2, 2), (3, 3), (4, 4), (5, 5) i (6, 6) su elementarni. Kod složenih događaja, skup ishoda čija realizacija povlači ostvarivanje događaja ima bar dva elementa, dok je kod elementarnih događaja taj skup jednočlan. Svaki složeni događaj se jednoznačno razlaže na elementarne događaje i svaki složeni događaj se može opisati pomoću elementarnih događaja.

Neka je A događaj vezan uz neki slučajni opit, odnosno, A je mogući ishod tog opita. Pretpostavimo da smo taj opit ponovili n puta i da se u tih n ponavljanja događaj A pojavio tačno n_A puta. Tada broj n_A zovemo **frekvencija (učestalost)** događaja A , a broj $\frac{n_A}{n}$ **relativna frekvencija**

događaja A (u tih n ponavljanja opita).

Iz rečenog slijedi da je $0 \leq \frac{nA}{n} \leq 1$.

Vratimo se opitu u kome se baca novčić i postavimo sljedeće pitanje: Kolika je vjerovatnoća da će pasti pismo? Iako je pitanje neprecizno formulisano, jer pojam vjerovatnoće još nije definisan, većina upitanih bi odgovorila da je tražena vjerovatnoća $\frac{1}{2}$. Podsvjesno se smatra da je tražena vjerovatnoća mjera mogućnosti pojavljivanja razmatranog ishoda prilikom jednog obavljanja opita. Odgovor se ne može logički obrazložiti već je zasnovan na sljedećim intuitivno – iskustvenim činjenicama:

¹⁰ Pošto u opitu imamo dva ravnopravna ishoda (zbog simetričnosti novčića), izgledi za pismo i grb su po 50%, pa je vjerovatnoća da će pasti pismo $\frac{1}{2}$.

²⁰ Iskustvo nas uči da se pri velikom broju bacanja novčića, pismo i grb pojavljuju približno jednak broj puta, pa je relativna frekvencija $\frac{nP}{n}$ blizu $\frac{1}{2}$. Stoga je razumno $\frac{1}{2}$ proglasiti za traženu vjerovatnoću pojavljivanja pisma u jednom opitu.

U Teoriji vjerovatnoća se razmatraju samo oni opiti koji imaju svojstvo tzv. statističke stabilnosti relativnih frekvencija. Pojasnimo.

Pretpostavimo da smo obavili j serija opita, pri čemu je u svakoj seriji opit ponovljen n puta, n je "velik" broj. Označimo sa f_A^i , $i = 1, 2, \dots, j$, relativnu frekvenciju događaja A u i -toj seriji. Svojstvo statističke stabilnosti relativnih frekvencija podrazumijeva približno jednake relativne frekvencije u našim serijama.

Klasična definicija vjerovatnoće a posteriori: Ako slučajni opit zadovoljava uslov statističke stabilnosti relativnih frekvencija, tada se vjerovatnoća **a posteriori** proizvoljnog događaja A iz tog opita definiše kao realni broj $P(A)$, $0 \leq P(A) \leq 1$, oko kog se grupišu, odnosno, kom teže relativne frekvencije tog događaja.

Ova definicija nije matematički precizna. Naime, postavlja se pitanje, kako provjeriti hipotezu o stabilnosti relativnih frekvencija.

Na primjer, u opitu sa bacanjem novčića, mi očekujemo da $\frac{nP}{n} \rightarrow \frac{1}{2}$, no nije isključen slučaj da novčić stalno pada na pismo (grb) i tada je $\frac{nP}{n} \rightarrow 1(0)$. Slično, nemamo argumenata da isključimo slučaj kada $\frac{nP}{n} \rightarrow \alpha \in [0, 1]$, $\alpha \neq \frac{1}{2}$ ili da $\frac{nP}{n}$ uopšte ne konvergira.

Bez obzira na navedene manjkavosti ove definicije vjerovatnoće, ona je kroz nekoliko vjekova bila važeća i u tom periodu je dobijen niz klasičnih i vrijednih vjerovatnosnih rezultata.

Primjer 1. U tabeli formata 10×10 dati su rezultati 10000 bacanja novčića grupisanih u 100 različitih serija po 100 bacanja, a u tabeli formata 10×1 rezultati grupisani u 10 serija po 1000 bacanja. Brojevi u tabelama pokazuju

broj pojavljivanja pisma.

54	46	53	55	46	54	41	48	51	53	501
48	46	40	53	49	49	48	54	53	45	485
43	52	58	51	51	50	52	50	53	49	509
58	60	54	55	50	48	47	57	52	55	536
48	51	51	49	44	52	50	46	53	41	485
49	50	45	52	52	48	47	47	47	51	488
45	47	41	51	49	59	50	55	53	50	500
53	52	46	52	44	51	48	51	46	54	497
45	47	46	52	47	48	59	57	45	48	494
47	41	51	48	59	51	52	55	39	41	484

Iz tabela se vidi da su relativne frekvencije događaja "palo je pismo" u svakoj seriji blizu $\frac{1}{2}$ i da su tim bliže što je serija veća (duža). ◀

Klasična definicija vjerovatnoće a priori: Neka imamo slučajni opit sa konačno mnogo elementarnih događaja i neka su svi ti elementarni događaji jednako mogući. Tada je vjerovatnoća proizvoljnog događaja iz tog opita jednaka količniku broja ishoda (elementarnih događaja) povoljnih za taj događaj i broja svih mogućih ishoda.

U ovoj definiciji postoje dvije slabosti. Prva je u njenoj restriktivnosti jer se odnosi samo na opite sa konačno mnogo ishoda. Definicija je kružna jer u sebi sadrži pojam jednako moguć koji defacto znači jednako vjerovatan.

Ni jedna od navedenih definicija vjerovatnoće ne može služiti kao baza aksiomske matematičke teorije vjerovatnoća.

Da bismo prešli na konstrukciju matematičke Teorije vjerovatnoća, moramo uvesti neke pojmove.

Polazni objekat u teoriji vjerovatnoća je skup Ω koji zovemo **prostor ishoda**. Elementi skupa Ω se označavaju sa ω i nazivaju se **ishodi** ili **elementarni događaji**.

Primjer 2. Novčić bacamo četiri puta. Tada možemo uzeti

$$\Omega = \{PPPP, PPPG, PPGP, \dots, GGGG\}.$$

Vidimo da je $\overline{\Omega} = 16$ ($\overline{\Omega}$ – broj elemenata skupa Ω). ◀

Primjer 3. Kocku bacamo dva puta. Tada možemo uzeti

$$\Omega = \{(1, 1), (1, 2), (1, 3), \dots, (5, 6), (6, 6)\} \text{ i jasno } \overline{\Omega} = 36. \blacktriangleleft$$

Primjer 4. Novčić se baca do prve pojave pisma. Tada možemo uzeti

$$\Omega = \{P, GP, GGP, GGGP, GGGGP, \dots\}.$$

Jasno, skup Ω je prebrojiv tj. $\overline{\Omega} = \aleph_0$. ◀

Primjer 5. Opit se sastoji u registrovanju broja kosmičkih čestica koje u toku dana padnu na neki lokalitet. Tada je prirodno uzeti

$$\Omega = \{0, 1, 2, 3, 4, 5, \dots\} = N \cup \{0\}.$$

Ω je beskonačan skup iako znamo da broj registrovanih čestica mora biti konačan. Za ovakav izbor skupa Ω smo se odlučili da bismo izbjegli neugodnosti kod ustanovljavanja broja koji treba da predstavlja maksimalni broj čestica. Ako prihvatimo da broj čestica može biti npr. 7213, nema razloga da isključimo mogućnost da taj broj bude 7214 i ovakvim rezonovanjem maksimalni broj možemo uvećati do beskonačnosti. ◀

Primjer 6. Opit se sastoji u biranju tačke sa kruga K . Tada je $\Omega = \{\omega : \omega \in K\} = K$. Opet imamo skup Ω sa beskonačno mnogo elemenata, pri čemu je $\overline{\Omega} = c$. ◀

Primjer 7. Opit se sastoji u kontinuiranom mjerenju temperature na nekom lokalitetu. Temperatura se mjeri 24 sata. Pošto je dnevni grafik temperature neprekidna funkcija, prirodno je za Ω uzeti skup neprekidnih funkcija zadatih na vremenskom intervalu od 24 sata. Ovdje svjesno povećavamo Ω , iako znamo da među neprekidnim funkcijama ima onih koje ne mogu biti grafikom dnevne temperature. Ovakav pristup je posljedica nemogućnosti da se odredi dokle ići sa prihvatanjem neprekidnih funkcija kao eventualnih grafika dnevne temperature. I ovdje imamo $\overline{\Omega} = c$. ◀

Neka je A događaj vezan uz neki slučajni opit i neka je Ω prostor ishoda opita. Kako na pitanje: da li se događaj A ostvario u opitu? odgovaramo sa "da" ili "ne" u zavisnosti od toga da li je u opitu realizovan ishod koji odgovara događaju A , to događaj A možemo identifikovati sa skupom elementarnih ishoda iz Ω čija realizacija povlači realizaciju događaja A .

Dakle, **događaj** je podskup prostora ishoda Ω . Za događaj i odgovarajući skup koristimo istu oznaku. Događaj A će se dogoditi akko se u opitu dogodi (realizuje) neki od ishoda iz A (sada A predstavlja skup).

Ω nazivamo siguran događaj i on će se dogoditi u svakom opitu (u svakom opitu će se realizovati ishod iz Ω). Prazan skup $\emptyset$ nazivamo nemoguć događaj; on se nikada neće dogoditi.

U tekstu koji slijedi, svi skupovi pripadaju istom prostoru ishoda Ω .

Za događaj A kažemo da povlači (implicira) događaj B ako je $A \subseteq B$.

Neka je $A^c = \Omega \setminus A$. Događaj A^c nazivamo suprotni događaj od A . Događaj A^c se realizuje u opitu akko se A ne realizuje.

Skup – događaj $A \cup B$ nazivamo unija događaja A i B . Događaj $A \cup B$ se realizuje akko se realizuje bar jedan od događaja A i B . Analogno se definiše

unija konačno mnogo događaja $\bigcup_{k=1}^n A_k$ i unija prebrojivo mnogo događaja $\bigcup_{n=1}^{\infty} A_n$.

Skup – događaj $A \cap B$ nazivamo presjek događaja A i B . Događaj AB (u Teoriji vjerovatnoća uobičajena oznaka za presjek) se realizuje akko se realizuju oba događaja A i B . Analogno se definišu $\bigcap_{k=1}^n A_k$ i $\bigcap_{n=1}^{\infty} A_n$.

Za događaje A i B kažemo da se uzajamno isključuju (da su disjunktne) ako je $AB = \emptyset$. U tom slučaju, događaji A i B se ne mogu istovremeno dogoditi.

Skup – događaj $A \setminus B$ nazivamo razlika događaja A i B . Kako je $A \setminus B = AB^c$, zaključujemo da se događaj $A \setminus B$ realizuje akko se A realizuje i B ne realizuje.

Za događaje – skupove važe pravila iz Teorije skupova. Na primjer, važe De Morganovi zakoni

$$\left(\bigcup_{n=1}^{\infty} A_n \right)^c = \bigcap_{n=1}^{\infty} A_n^c \text{ i } \left(\bigcap_{n=1}^{\infty} A_n \right)^c = \bigcup_{n=1}^{\infty} A_n^c.$$

Izloženu teoriju ilustriramo primjerom.

Primjer 8. Opi se sastoji od dva bacanja kocke. Jasno,

$$\Omega = \{(i, j) : i, j = 1, 2, \dots, 6\}.$$

Neka je $A = \{(i, j) : i + j \leq 3\}$, $B = \{(i, j) : j = 6\}$ i $C = \{(i, j) : j = 2, 4, 6\}$. Zadržimo se za trenutak na događaju A . Osim date karakterizacije, možemo zapisati $A = \{(1, 1), (1, 2), (2, 1)\}$ ili pak možemo koristiti zapis " A – zbir brojeva je najviše 3". Dakle, događaj se karakteriše navođenjem elementarnih ishoda koji događaju odgovaraju, ali postoje slučajevi kada karakterizaciju možemo obaviti navođenjem uslova uz koje će se događaj realizovati. Primijetimo,

$B = \{(1, 6), (2, 6), \dots, (6, 6)\}$ ili B – u drugom bacanju je pao broj 6,

$C = \{(1, 2), (2, 2), \dots, (6, 6)\}$ ili C – u drugom bacanju je pao paran broj.

Potražimo A^c i B^c . Imamo

$A^c = \{(1, 3), (1, 4), \dots, (6, 6)\}$ ili A^c – zbir brojeva je bar 4,

C^c – u drugom bacanju je pao neparan broj.

Konstatujemo, $A \cap B = \emptyset$, $A \cap C = \{(1, 2)\}$, $B \cup C = C$, $B \subset C$, $C \setminus B = \{(i, j) : j = 2, 4\}$. ◀

Preporučujemo sljedeće zadatke za vježbu.

Zadaci

1. Neka su A, B i C događaji. Opisati događaj da
 - a) se ostvario samo događaj A ,
 - b) su se ostvarili događaji A i B , a C nije ostvaren,
 - c) su se ostvarila sva tri događaja,
 - d) je ostvaren bar jedan događaj,
 - e) su ostvarena bar dva događaja,
 - f) je ostvaren jedan i samo jedan događaj,
 - g) su ostvarena dva i samo dva događaja,
 - h) nije ostvaren nijedan događaj,
 - i) su ostvarena ne više od dva događaja.
2. Strijelac gađa u metu 4 puta, pri čemu se registruju pogoci i promašaji. Opisati prostor ishoda i događaje: A – gađanje je započelo promašajem; B – rezultat svih gađanja je isti; C – cilj je pogođen tačno dva puta; D – cilj je pogođen bar dva puta.
3. Uniju $A \cup B$ možemo zapisati kao $A \cup B = A \cup (A \setminus B)$, dakle kao zbir dva disjunktna događaja. Uniju događaja A, B i C zapisati kao zbir disjunktних događaja.
4. Novčić se baca dva puta. Ako se u oba bacanja pojavi pismo, novčić se baca još jednom, a u suprotnom, jednom se baca kocka za igru. Odrediti prostor ishoda i događaj da se opit završava padanjem šestice.
5. Novčić se baca sve dok se ne pojave dva pisma uzastopno. Odrediti prostor ishoda i događaj da se opit završi prije sedmog bacanja.

1.2 Aksiomi Teorije vjerovatnoća

Izložićemo aksiome Teorije vjerovatnoća. Aksiomatiku je u Teoriju vjerovatnoća uveo Kolmogorov. Aksiomi su objavljeni u knjizi *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin, 1933. Predložena aksiomatika je korespondirala sa do tada poznatim vjerovatnosnim rezultatima. Uz to, Kolmogorovljev sistem aksioma dobro reprezentuje naše intuitivne predstave o vjerovatnoći kao graničnoj vrijednosti relativnih frekvencija. Kolmogorovljeva aksiomatika je bila odlična osnova iz koje se Teorija vjerovatnoća razvila u jednu od značajnih matematičkih disciplina. Iako su do 1933.g. do-

bijeni neki krupni vjerovatnosni rezultati, istoričari matematike smatraju da je te godine Teorija vjerovatnoća postala samostalna matematička disciplina. Svaka matematička disciplina podrazumijeva tri atributa: 1 – aksiomatiku, 2 – specifične probleme i 3 – specifične metode. Vjerovatnoća posjeduje ove attribute.

Sa $\mathcal{P}(\Omega)$ ćemo označavati partitivni skup prostora ishoda Ω .

DEFINICIJA 1.2.1 *Familija $\mathcal{A}$ podskupova prostora ishoda Ω je algebra skupova ako je:*

$$A1. \emptyset \in \mathcal{A}.$$

$$A2. A \in \mathcal{A} \implies A^c \in \mathcal{A}.$$

$$A3. A_1, A_2, \dots, A_n \in \mathcal{A} \implies \bigcup_{i=1}^n A_i \in \mathcal{A}.$$

Dakle, $\mathcal{A}$ je algebra ako je zatvorena na komplementiranje i konačne unije. Iz A1 – A3 slijedi da za algebru $\mathcal{A}$ vrijedi:

$$\Omega = \emptyset^c \in \mathcal{A}.$$

$$A_1, A_2, \dots, A_n \in \mathcal{A} \implies \bigcap_{i=1}^n A_i = \left(\bigcup_{i=1}^n A_i^c \right)^c \in \mathcal{A}.$$

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

Prema tome, algebra je zatvorena i na konačne presjeke i skupovne razlike.

DEFINICIJA 1.2.2 *Familija $\mathcal{F}$ podskupova prostora ishoda Ω ($\mathcal{F} \subset \mathcal{P}(\Omega)$) je σ -algebra skupova ako je:*

$$F1. \emptyset \in \mathcal{F}.$$

$$F2. A \in \mathcal{F} \implies A^c \in \mathcal{F}.$$

$$F3. A_i \in \mathcal{F}, i \in \mathbb{N} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

Kao i u slučaju algebre, pokazuje se da $\Omega \in \mathcal{F}$ i da je σ -algebra $\mathcal{F}$ zatvorena na prebrojive presjeke i skupovne razlike.

Postoje dvije trivijalne σ -algebre i to su $\mathcal{P}(\Omega)$ i $\{\emptyset, \Omega\}$.

Očigledno, svaka σ -algebra je algebra. Navešćemo primjer algebre koja nije σ -algebra.

Primjer 1. Neka je Ω beskonačan skup i $\mathcal{A}$ familija svih podskupova skupa Ω koji su konačni ili imaju konačne komplemente. Tada je $\mathcal{A}$ algebra, ali nije σ -algebra.

Primjer 2. Neka je $\mathcal{F}$ familija svih podskupova koji su prebrojivi ili imaju prebrojive komplemente. Tada je $\mathcal{F}$ σ -algebra.

Iz $F1 - F3$ slijedi da je svaka σ -algebra istovremeno i algebra.

DEFINICIJA 1.2.3 Neka je $\mathcal{F}$ σ -algebra na prostoru ishoda Ω . Ureden par $(\Omega, \mathcal{F})$ naziva se **mjerljivi prostor**.

DEFINICIJA 1.2.4 Neka je $(\Omega, \mathcal{F})$ mjerljivi prostor. Funkcija $P : \mathcal{F} \rightarrow \mathbb{R}$ je **vjerovatnoća** (na $\mathcal{F}$, na Ω) ako važi:

P1. $P(A) \geq 0, A \in \mathcal{F}; P(\Omega) = 1.$

P2. $A_i \in \mathcal{F}, i \in \mathbb{N} \text{ i } A_i \cap A_j = \emptyset, i \neq j \implies P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$

DEFINICIJA 1.2.5 Uredena trojka $(\Omega, \mathcal{F}, P)$, gdje je $\mathcal{F}$ σ -algebra na Ω i P vjerovatnoća na $\mathcal{F}$, naziva se **vjerovatnosni prostor**.

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

Neka je $(\Omega, \mathcal{F}, P)$ vjerovatnosni prostor. Elemente σ -algebre $\mathcal{F}$ nazivamo **događaji**, a broj $P(A), A \in \mathcal{F}$, naziva se **vjerovatnoća događaja A** .

Važan problem u teoriji vjerovatnoća je konstrukcija vjerovatnosnog prostora, tj. σ -algebre događaja i vjerovatnosne mjere na toj σ -algebri. U odnosu na ovaj problem postoji bitna razlika između slučaja kada je skup Ω konačan ili prebrojiv i slučaja kada je skup Ω neprebrojiv. Naime, u prvom slučaju na skupu $\mathcal{P}(\Omega)$ je uvijek moguće zadati vjerovatnoću, dok u slučaju kada je skup Ω neprebrojiv to nije moguće. Objašnjenje ovog fakta zahtijeva šire i dublje poznavanje matematičke analize od onog koje se od nas traži, pa nam stoga ostaje da ovaj fakt prihvatimo bez obrazloženja.

Dokazaćemo dvije teoreme koje slijede iz definicije vjerovatnosnog prostora. U tim teoremama će biti evidentirana neka značajna svojstva vjerovatnoće.

Teorema 1.2.1 Neka je $(\Omega, \mathcal{F}, P)$ vjerovatnosni prostor. Tada vrijedi:

(a) $P(\emptyset) = 0.$

- (b) Ako su $A_1, \dots, A_n \in \mathcal{F}$ međusobno disjunktni događaji, tada je
- $$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$
- (svojstvo **konačne aditivnosti vjerovatnoće**).
- (c) $A, B \in \mathcal{F}$ i $A \subseteq B \implies P(A) \leq P(B)$
 (svojstvo **monotonosti vjerovatnoće**).
- (d) $A_n \in \mathcal{F}, n \in N, A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$ i
- $$A = \bigcup_{n=1}^{\infty} A_n \implies P(A) = \lim_{n \rightarrow \infty} P(A_n)$$
- (**neprekidnost vjerovatnoće u odnosu na rastući niz događaja**).
- (e) $A_n \in \mathcal{F}, n \in N, A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots$ i
- $$A = \bigcap_{n=1}^{\infty} A_n \implies P(A) = \lim_{n \rightarrow \infty} P(A_n)$$
- (**neprekidnost vjerovatnoće u odnosu na opadajući niz događaja**).
- (f) $A_n \in \mathcal{F}, n \in N \implies P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n)$
 (σ -**poluaditivnost vjerovatnoće**).
- (g) $A, B \in \mathcal{F} \implies P(A \cup B) = P(A) + P(B) - P(AB)$.
- (h) $A \in \mathcal{F} \implies P(A^c) = 1 - P(A)$
 (**vjerovatnoća suprotnog događaja A^c**).
- (i) $A \in \mathcal{F} \implies 0 \leq P(A) \leq 1$.

Dokaz.

- (a) U aksiomi P2 uzmimo $A_1 = \Omega$ i $A_i = \emptyset, i \geq 2$. Dobijamo

$$1 = 1 + P(\emptyset) + P(\emptyset) + \dots \implies P(\emptyset) = 0$$

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

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

- (d) Stavimo $B_1 = A_1, B_n = A_n \setminus A_{n-1}, n \geq 2$. Tada je $B_n \in \mathcal{F}, n \in N$ i $B_i \cap B_j = \emptyset, i \neq j$. Osim toga vrijedi $A_n = \bigcup_{i=1}^n B_i, A = \bigcup_{n=1}^{\infty} B_n$, pa imamo

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

Sada (d) slijedi iz definicije sume konvergentnog reda.

- (e) Stavimo $C_n = A_1 \setminus A_n, n \in N$. Tada je $C_n \in \mathcal{F}, C_n \subseteq C_{n+1}, n \in$

N i $A_1 \setminus A = \bigcup_{n=1}^{\infty} C_n$. Sada imamo $P(C_n) = P(A_1) - P(A_n)$, pa iz (d) slijedi

$$P(A_1) - P(A) = P(A_1 \setminus A) = \lim_{n \rightarrow \infty} P(C_n) = P(A_1) - \lim_{n \rightarrow \infty} P(A_n).$$

Iz posljednje jednakosti slijedi (e).

(f) Neka je $B_1 = A_1$, $B_n = A_n \setminus \bigcup_{k=1}^{n-1} A_k$, $n \geq 2$. Tada je B_n , $n \in N$, niz

disjunktnih događaja, $B_n \subseteq A_n$, $n \in N$ i $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$. Zato imamo

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(B_n) \leq \sum_{n=1}^{\infty} P(A_n).$$

(g) Imamo $A \cup B = A \cup (B \setminus A)$, $(A \cap B) \cup (B \setminus A) = B$. Odavde, zbog (b), dobijamo

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

Odavde slijedi (g).

(h) $A \cup A^c = \Omega \implies P(A) + P(A^c) = 1$.

(i) Slijedi neposredno iz P1, (a) i (c).

Teorema 1.2.2 (Silvesterova formula.) *Neka je $(\Omega, \mathcal{F}, P)$ vjerovatnosni prostor i $A_i \in \mathcal{F}$, $i = 1, 2, \dots, n$. Tada važi*

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i \leq n} P(A_i) - \sum_{i < j \leq n} P(A_i A_j) \\ &\quad + \sum_{i < j < k \leq n} P(A_i A_j A_k) - \dots + (-1)^{n+1} P\left(\bigcap_{i=1}^n A_i\right). \end{aligned} \quad (1)$$

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

Dokaz ćemo sprovesti indukcijom po n . Za $n = 1$ tvrđenje je očigledno, a za $n = 2$ to je tvrđenje (g) iz teoreme 1.2.1. Pretpostavimo da tvrđenje vrijedi za sve familije od najviše $n - 1$ događaja. Stavimo

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

Tada vrijedi

$$P\left(\bigcup_{i=1}^n A_i\right) = P(B \cup A_n) = P(B) + P(A_n) - P(BA_n). \quad (2)$$

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

$$P(B) = \sum_{i < n} P(A_i) - \sum_{i < j < n} P(A_i A_j) + \dots + (-1)^n P\left(\bigcup_{i=1}^{n-1} A_i\right) \quad (3)$$

i

$$\begin{aligned} P(BA_n) &= \sum_{i < n} P(C_i) - \sum_{i < j < n} P(C_i C_j) + \dots + (-1)^n P\left(\bigcap_{i=1}^{n-1} C_i\right) \\ &= \sum_{i < n} P(A_i A_n) - \sum_{i < j < n} P(A_i A_j A_n) + \dots + (-1)^n P\left(\bigcap_{i=1}^n A_i\right). \end{aligned} \quad (4)$$

Iz (2), (3) i (4) slijedi (1).

1.3 Minimalna σ -algebra i Borelova σ -algebra

U Vjerovatnoći se često pojavljuje potreba da se razmatra σ -algebra događaja koja sadrži skupove iz neke kolekcije $\mathcal{C}$ podskupova prostora ishoda Ω , pri čemu sama kolekcija $\mathcal{C}$ nije σ -algebra podskupova iz Ω . Jasno, σ -algebra $\mathcal{P}(\Omega)$ sadrži svaku kolekciju. Ako je $\mathcal{W} = \{\mathcal{C}_\lambda, \lambda \in \Lambda\}$ kolekcija svih σ -algebri koje sadrže $\mathcal{C}$, tada je $\bigcap_{\lambda \in \Lambda} \mathcal{C}_\lambda = \sigma[\mathcal{C}]$, σ -algebra koja sadrži $\mathcal{C}$ i ona ima svojstvo da je sadržana u svakoj drugoj σ -algebri koja sadrži $\mathcal{C}$. σ -algebra $\sigma[\mathcal{C}]$ se naziva **minimalna σ -algebra generisana kolekcijom $\mathcal{C}$** .

Upoznajmo se sa jednim važnim primjerom σ -algebre.

Neka je $\Omega = R$ i neka je $\mathcal{I} = \{(a, b) : a, b \in R, a < b\}$. Minimalna σ -algebra generisana kolekcijom $\mathcal{I}$ naziva se **Borelova σ -algebra**, a njeni elementi se nazivaju **Borelovi skupovi**. Za upravo uvedenu σ -algebru, koristi se oznaka $\mathcal{B}$. σ -algebru $\mathcal{B}$ generišu i neke druge kolekcije. Da bismo bili određeniji, navešćemo teoremu koja se dokazuje u kursovima Matematičke analize.

Teorema 1.3.1 *Neka je $\mathcal{A}_1$ – kolekcija otvorenih skupova na R ; $\mathcal{A}_2$ – kolekcija zatvorenih skupova na R ; $\mathcal{A}_3 = \{[a, b) : a, b \in R, a < b\}$; $\mathcal{A}_4 = \{(a, b] : a, b \in R, a < b\}$; $\mathcal{A}_5 = \{[a, b] : a, b \in R, a < b\}$; $\mathcal{A}_6 = \{(-\infty, a) : a \in R\}$; $\mathcal{A}_7 = \{(a, \infty) : a \in R\}$. Tada je $\sigma[\mathcal{A}_1] = \sigma[\mathcal{A}_2] = \sigma[\mathcal{A}_3] = \sigma[\mathcal{A}_4] = \sigma[\mathcal{A}_5] = \sigma[\mathcal{A}_6] = \sigma[\mathcal{A}_7] = \sigma[\mathcal{I}] = \mathcal{B}$.*

σ -algebra $\mathcal{B}$ je veoma bogata. Primjera radi, u nju ulaze: singletoni (tj. jednočlani skupovi), skup racionalnih brojeva, skup iracionalnih brojeva.

Podsjetimo se nekih pojmova iz višedimenzionalne analize. U prostoru R^n postoji prirodno uopštenje intervala. Neka su $a, b \in R^n, a = (a_1, \dots, a_n), b = (b_1, \dots, b_n)$, pri čemu je $a_i \leq b_i, i = 1, 2, \dots, n$. Tada stavljamo:

$$\begin{aligned}(a, b) &= \{(x_1, \dots, x_n) \in R^n : a_i < x_i < b_i, i = 1, 2, \dots, n\}, \\ [a, b) &= \{(x_1, \dots, x_n) \in R^n : a_i \leq x_i < b_i, i = 1, 2, \dots, n\}, \\ (-\infty, a) &= \{(x_1, \dots, x_n) \in R^n : x_i < a_i, i = 1, 2, \dots, n\}.\end{aligned}$$

Neka je $\Omega = R^n$. Minimalna σ -algebra generisana kolekcijom intervala (koristi se i termin paralelopiped) $(a, b), a, b \in R^n$ se naziva **Borelova σ -algebra na R^n** i označava sa $\mathcal{B}^n$.

DEFINICIJA 1.3.1 *Funkcija $f : R^n \longrightarrow R$ je **Borelova** ako je za svako $S \in \mathcal{B}, f^{-1}(S) \in \mathcal{B}^n$.*

Primjer 1. Slučajno biramo tačku iz intervala $(0, 1)$. Kako svaka tačka iz $(0, 1)$ može biti izabrana, to je $\Omega = (0, 1)$. Slučajnom biranju tačke damo i sljedeći smisao: Ako je I interval sa $(0, 1)$, tada je vjerovatnoća da slučajno izabrana tačka pripadne I jednaka $m(I)$, gdje je $m(I)$ oznaka za dužinu intervala I . Vidimo da vjerovatnoća ne zavisi od položaja intervala I na $(0, 1)$. Interval I možemo identifikovati sa događajem da tačka pripadne intervalu I . Proglašavajući intervale za događaje, dobijamo kolekciju $\mathcal{I}$ intervala – događaja. Međutim ova kolekcija nije σ -algebra jer je recimo, $\bigcap_{n=2}^{\infty} (1/2 - 1/n, 1/2 + 1/n) = \{1/2\}$, a jednočlani skup $\{1/2\}$ nije interval. Na σ -algebri $\mathcal{P}(\Omega)$ ne postoji funkcija koja posjeduje sva svojstva vjerovatnoće, te rješenje za σ -algebru događaja treba tražiti na potezu između kolekcije $\mathcal{I}$ i kolekcije $\mathcal{P}(\Omega)$. $\sigma[\mathcal{I}]$ je kolekcija Borelovih skupova sa $(0, 1)$ i na njoj postoji funkcija P (u Analizi se funkcija P naziva Lebegova mjera i predstavlja uopštenje dužine) koja svakom skupu iz $\sigma[\mathcal{I}]$ dodjeljuje realni broj, a uz to ima sva svojstva vjerovatnoće. Na kolekciji $\mathcal{I}$ funkcija P se svodi na dužinu odgovarajućeg intervala. Vjerovatnosni prostor kojim se modelira naš opit je $((0, 1), \mathcal{B}_{(0,1)}, P)$, gdje je $\mathcal{B}_{(0,1)}$ kolekcija Borelovih skupova sa $(0, 1)$, a P je

Lebegova mjera. Kako je dužina, recimo intervala $(1/4, 1/3)$ jednaka $1/12$ to je vjerovatnoća događaja da slučajno izabrana tačka bude iz ovog intervala jednaka $1/12$. Kako je "dužina", recimo, jednočlanog skupa $\{1/3\}$ jednaka 0 to je vjerovatnoća slučajnog izbora ove tačke jednaka 0. Dakle, izbor tačke $1/3$ nije isključen, ali je vjerovatnoća događaja da bude izabrana baš tačka $1/3$ jednaka 0. ◀

1.4 Elementi kombinatorike

Ovaj izlet u odnosu na naznačeni kurs (naš cilj je izlaganje osnova Teorije vjerovatnoća) je urađen da bismo se upoznali sa nekim pojmovima iz kombinatorike. Kombinatorne ideje su često prisutne u vjerovatnoći.

Teorema 1.4.1 (Princip proizvoda.) *Neka su $A_1, \dots, A_n$ konačni skupovi i neka je $T \subseteq A_1 \times \dots \times A_n$ skup uređenih n -torki $(a_1, \dots, a_n)$ koji je zadat na sljedeći način: prva komponenta a_1 može se birati na k_1 različitih načina (tj. među k_1 različitih elemenata skupa A_1); za svaku već izabranu prvu komponentu drugu komponentu a_2 možemo birati na k_2 različitih načina itd. Za svaki izbor komponenta $a_1, \dots, a_{n-1}$ n -tu komponentu a_n možemo birati na k_n različitih načina. Tada je $\overline{T} = k_1 k_2 \dots k_n$.*

Dokaz Teoremu ćemo dokazati matematičkom indukcijom po n . Za $n = 1$ tvrdjenje teoreme je očigledno. Neka je $n = 2$. Pretpostavimo da prvu komponentu biramo iz skupa $\{b_1, \dots, b_{k_1}\} \subseteq A_1$, pri čemu su $b_1, \dots, b_{k_1}$ međusobno različiti. Ako je prva komponenta izabrana, recimo da je to b_i , pretpostavimo da drugu komponentu biramo iz skupa $\{c_1, \dots, c_{k_2}\} \subseteq A_2$, pri čemu su $c_1, \dots, c_{k_2}$ međusobno različiti. Pošto su svi uređeni parovi (b_i, c_j) međusobno različiti, to postoji biunivoka korespondencija između svih uređenih parova (b_i, c_j) ($i = 1, \dots, k_1, j = 1, \dots, k_2$) i svih uređenih parova prirodnih brojeva (i, j) ($i = 1, \dots, k_1, j = 1, \dots, k_2$). Uređeni parovi prirodnih brojeva (ako se paru pridruži tačka u ravni) generišu pravougaoni roj tačaka. Tačaka po osnovici pravougaonika ima k_1 , a po visini k_2 , te je ukupan broj tačaka iz roja $k_1 k_2 = \overline{T}$. Pretpostavimo da tvrdjenje vrijedi za $n > 2$. Uređenu $n + 1$ -torku $(b_{j_1}, c_{j_2}, \dots, x_{j_n}, y_{j_{n+1}})$, možemo tretirati kao uređen par $((b_{j_1}, c_{j_2}, \dots, x_{j_n}), y_{j_{n+1}})$. Kako po induktivnoj pretpostavci, svih izbora za $(b_{j_1}, c_{j_2}, \dots, x_{j_n})$ ima $k_1 k_2 \dots k_n$, zaključujemo da svih uređenih $n + 1$ -torki ima $k_1 k_2 \dots k_n k_{n+1}$.

Komentar. Principu proizvoda možemo dati jedno slobodnije tumačenje. Naime, ako prvu stvar možemo napraviti na k_1 načina, drugu na k_2 načina, n -tu na k_n načina, tada se sve stvari mogu napraviti na ukupno $k_1 k_2 \dots k_n$ načina.

Primjer 1. Kocka za igru se baca 10 puta i bilježe se rezultati bacanja. Odrediti prostor ishoda Ω i naći $\overline{\Omega}$.

► Ishodi eksperimenta su desetorke, pri čemu su elementi desetorki iz skupa $S = \{1, 2, 3, 4, 5, 6\}$. Prvu komponentu desetorke možemo napraviti na 6 načina (prvi broj je iz skupa S), drugu takođe na 6 načina itd. Iz teoreme 1.4.1 slijedi da je $\overline{\Omega} = 6^{10}$. ◀

DEFINICIJA 1.4.1 *Neka je $A = \{a_1, \dots, a_n\}$ i $r \in N$, $1 \leq r \leq n$. **Varijacija r -tog reda** u skupu A je svaka uređena r -torka $a_{i_1}, \dots, a_{i_r}$ međusobno različitih elemenata skupa A .*

Odredimo broj $V_n^{(r)}$ svih varijacija reda r u n -članom skupu.

Prvu komponentu možemo izabrati na n načina. Kada je prva komponenta izabrana, drugu možemo izabrati na $n - 1$ načina. Kada je izabrano j , $1 \leq j < r$, komponenti, sljedeću možemo izabrati na $n - j$ načina. Iz teoreme 1.4.1 slijedi

$$V_n^{(r)} = n(n-1) \cdots (n-r+1). \quad (1)$$

Primjer 2. Na koliko načina može biti obavljeno izvlačenje brojeva u igri LOTO (Izvlači se 7 brojeva iz skupa $S = \{1, 2, \dots, 39\}$)?

► Prvi broj može biti izvučen ("napravljen") na 39 načina, drugi na 38 načina, ..., sedmi na 33 načina. Dakle, postoji $39 \cdot 38 \cdots 33$ načina. ◀

DEFINICIJA 1.4.2 **Permutacija** u skupu $A = \{a_1, \dots, a_n\}$, ($\overline{A} = n$), je svaka varijacija n -tog reda u skupu A .

Sa P_n se označava broj svih permutacija u n -članom skupu. Iz (1) slijedi

$$P_n = V_n^{(n)} = n(n-1) \cdots 1 = n! \quad (2)$$

Neka je $A = \{a, b, c, d\}$. Tada su varijacije trećeg reda u skupu A sljedeće:

(a,b,c)	(a,b,d)	(a,c,d)	(b,c,d)
(a,c,b)	(a,d,b)	(a,d,c)	(b,d,c)
(b,a,c)	(b,a,d)	(c,a,d)	(c,b,d)
(b,c,a)	(b,d,a)	(c,d,a)	(c,d,b)
(c,a,b)	(d,a,b)	(d,a,c)	(d,b,c)
(c,b,a)	(d,b,a)	(d,c,a)	(d,c,b)

Neka je $A = \{a, b, c\}$. Tada su permutacije u skupu A sljedeće:

$$\begin{array}{ccc} (a,b,c) & (a,c,b) & (b,a,c) \\ (b,c,d) & (c,a,b) & (c,b,a) \end{array}$$

DEFINICIJA 1.4.3 *Neka je $A = \{a_1, \dots, a_n\}$, $(\overline{A} = n)$, $r \in N$, $1 \leq r \leq n$. **Kombinacija r -tog reda** u skupu A je svaki r -člani podskup skupa A .*

Broj svih kombinacija r -tog reda u n -članom skupu se označava sa $C_n^{(r)}$.

Da bismo odredili $C_n^{(r)}$, primijenimo sljedeću proceduru.

(a) Izaberimo r različitih elemenata (bez obzira na uređenje) iz n -članog skupa A . Ovim biramo r -člani podskup od A , a ovaj izbor možemo obaviti na $C_n^{(r)}$ načina.

(b) U svakom izabranom r -članom podskupu od A , elemente možemo poredati na r načina.

Opisanom procedurom formiramo varijacije r tog reda u n -članom skupu. Iz teoreme 1.4.1 dobijamo $C_n^{(r)} \cdot r! = V_n^{(r)}$, odakle slijedi

$$C_n^{(r)} = \frac{n!}{r!(n-r)!} = \binom{n}{r}.$$

Neka je $A = \{a, b, c, d, e\}$. Tada su kombinacije trećeg reda u skupu A sljedeće:

$$\begin{array}{ccccc} (a,b,c) & (a,b,d) & (a,b,e) & (a,c,d) & (a,c,e) \\ (a,d,e) & (b,c,d) & (b,c,e) & (b,d,e) & (c,d,e) \end{array}$$

Primjer 3. Koliko kombinacija treba uplatiti u igri LOTO da bi naš dobitak sedmice bio izvjestan?

► Očigledno, traženi broj je $\binom{39}{7}$. Objasnimo razliku među rezultatima iz primjera 2 i primjera 3. Sedmorke, recimo $(2, 1, 3, 4, 5, 6, 7)$ i $(7, 6, 5, 4, 3, 2, 1)$ opisuju dva različita procesa izvlačenja, ali oba izvlačenja odgovaraju igraču koji je na tiketu zaokružio brojeve $(1, 2, 3, 4, 5, 6, 7)$. ◀

Predlažemo čitaocu da obrati pažnju na sljedeće primjere. Naša ideja je da kod čitaoca kultiviramo kombinatorni način razmišljanja.

Primjer 4. Školske sveske se proizvode u 3 različite veličine, 6 različitih boja i 4 različite debljine. Koliko ima vrsta školskih svesaka?

► Primjenjujući princip proizvoda dobijamo da je traženi broj $3 \cdot 6 \cdot 4 = 72$. ◀

Primjer 5. Na koliko načina šest osoba može formirati red pred prodavnicom?

► Traženi broj načina je $6!$. ◀

Primjer 6. U ravni je zadato n , $n \geq 3$, tačaka, od kojih nikoje 3 ne leže na istoj pravoj. Naći broj trouglova sa tjemima u zadatim tačkama.

► Broj traženih trouglova jednak je broju načina na koje iz skupa tjemena možemo izdvojiti tri tjemena, a taj broj je $\binom{n}{3}$. ◀

Primjer 7. Na šahovskom turniru svaki igrač je odigrao po jednu partiju sa svakim igračem. Odigrano je ukupno 55 partija. Koliko je šahista učestvovalo na turniru?

► Neka je n broj šahista. Broj odigranih partija jednak je broju parova koje možemo obrazovati iz n -članog skupa šahista. Sada imamo $\binom{n}{2} = 55$, odakle slijedi da je $n = 11$. ◀

Primjer 8. Od 30 studenata sa prve, 25 sa druge, 20 sa treće i 15 sa četvrte godine studija treba formirati delegaciju u kojoj će biti 5 studenata sa prve, 4 sa druge, 3 sa treće i 2 sa četvrte godine studija. Na koliko se načina može formirati delegacija?

► Predstavnike sa prve godine možemo odabrati na $\binom{30}{5}$, sa druge na $\binom{25}{4}$, sa treće na $\binom{20}{3}$, sa četvrte na $\binom{15}{2}$ načina pa je na osnovu principa proizvoda traženi broj načina $\binom{30}{5}\binom{25}{4}\binom{20}{3}\binom{15}{2}$. ◀

Primjer 9. Na koliko se načina može razmjestiti 8 topova na šahovsku tablu tako da ne postoji par topova koji se međusobno napadaju?

► Jasno, iz svakog reda i svake kolone na tabli mora biti zauzeto tačno po jedno polje (na njemu će biti smješten top). Iz prvog reda se može izabrati jedno od 8 polja, iz drugog reda jedno od 7 polja (ne odgovara polje iz kolone u kojoj je prvoizabrano polje), iz trećeg reda jedno od 6 polja itd. Na osnovu principa proizvoda slijedi da je $8 \cdot 7 \cdot \dots \cdot 1 = 8!$ traženi broj razmještaja. ◀

Primjer 10. Na koliko se načina 10 teniskih loptica može podijeliti između 4 dječaka?

► Smatraćemo da se loptice ne razlikuju te će svaka podjela biti određena brojevima loptica koje su dječaci dobili. Uračunava se i varijanta kada jedan ili više dječaka ne dobiju lopticu. Loptice ćemo simbolički označavati sa *. Svaku podjelu možemo i to jednoznačno, predstaviti nizom u koji ulaze 3 crte | i 10 zvjezdica *. U tom nizu zvjezdice do prve | predstavljaju loptice koje u podjeli dobija prvi dječak (moguće je da do prve | nema zvjezdica što odgovara slučaju kada prvi dječak nije dobio lopticu), zvjezdice između prve | i druge | predstavljaju loptice koje je dobio drugi dječak itd. Primjera radi, nizu

|*|***|

odgovara podjela u kojoj je prvi dječak dobio 2 loptice, drugi jednu, treći 7 loptica, a četvrti dječak nije dobio lopticu. Pošto između skupa podjela i nizova napravljenih od $10*$ i $3|$ postoji biunivoka korespondencija, zaključujemo da traženih podjela ima $\binom{13}{3}$ – ovo je broj nizova od $10*$ i $3|$. Naime, kad obrazujemo niz – trinaestorku, od mogućih 13 pozicija biramo 3 na koje

postavljamo $|$, a na preostalih 10 pozicija postavljamo $*$. Ovih biranja ima $C_{13}^{(3)} = \binom{13}{3}$. ◀

Zadaci

1. Koliko ima petocifrenih brojeva u čijem zapisu ne učestvuju cifre 0, 1 i 2?
2. Koliko ima petocifrenih brojeva u čijem zapisu ne učestvuje 0 i nijedna cifra se ne ponavlja?
3. Koliko ima sedmocifrenih brojeva u čijem se zapisu tri puta pojavljuje cifra 3, dva puta cifra 2 i dva puta cifra 1?
4. Koliko se različitih riječi može dobiti premještanjem slova iz riječi STATISTIKA?
5. Koliko ima nepodudarnih trouglova čije su dužine stranica iz skupa $S = \{7, 8, \dots, 13\}$?
6. Na koliko se načina iz kompleta od 32 karte može izabrati 6 karata tako da među njima budu zastupljene sve boje?
7. Na koliko se načina 8 topova može rasporediti na šahovsku tablu dimenzija 12×12 , tako da se nikoja dva međusobno ne napadaju?
8. Na koliko se načina 30 studenata može podijeliti u četiri grupe, tako da u dvije bude po 10 studenata, a u dvije po 5 studenata?
9. Na kružnici je dato n različitih tačaka, $n \geq 3$. Odrediti broj različitih konveksnih mnogouglova sa tjemenuima u datim tačkama.
10. Košarkašku ekipu sačinjava 5 bekova, 4 centra i 3 krila. Na koliko se načina može sastaviti petorka ako u njoj moraju da igraju bar dva beka i bar jedan centar?
11. Na koliko načina figura na šahovskoj tabli iz donjeg lijevog ugla može doći u gornji desni ugao, potezima desno ili gore za po jedno polje?
12. Na koliko se načina n osoba može rasporediti na m stolica, $m > n$?
13. U konveksnom n -touglu ($n \geq 4$) povučene su sve dijagonale, pri čemu se nikoje tri ne sijeku u jednoj tački. Koliko ima presječnih tačaka (Tjemena n -tougla se ne računaju kao presječne tačke.)?

1.5 Diskretni prostor vjerovatnoća

Razmotrimo slučaj kada je prostor ishoda $\Omega = \{\omega_1, \omega_2, \dots\}$ najviše prebrojiv. Za σ -algebru $\mathcal{F}$ ćemo uzeti $\mathcal{P}(\Omega)$. Elementarnim ishodima $\omega_1, \omega_2, \dots$ pridružimo redom vjerovatnoće $p_1, p_2, \dots$ tako da važi

$$(a) \quad p_1 \geq 0, p_2 \geq 0, p_3 \geq 0, \dots$$

$$(b) \quad p_1 + p_2 + p_3 + \dots = 1.$$

Vjerovatnoću proizvoljnog događaja $A = \{\omega_{j_1}, \omega_{j_2}, \dots\}$ definišemo kao zbir vjerovatnoća ishoda povoljnih za događaj A tj.

$$P(A) = p_{j_1} + p_{j_2} + \dots \quad (1)$$

Lako se provjerava da je na ovaj način korektno definisana vjerovatnoća na $\mathcal{F}$. Trojka $(\Omega, \mathcal{F}, P)$ čiji su elementi upravo zadati, naziva se **diskretni prostor vjerovatnoća**.

Neka je $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, $p_1 = p_2 = \dots = p_n = \frac{1}{n}$ i neka je $A = \{\omega_{j_1}, \dots, \omega_{j_k}\} \in \mathcal{P}(\Omega)$. Iz (1) dobijamo

$$P(A) = \frac{k}{n} = \frac{\overline{A}}{\overline{\Omega}}. \quad (2)$$

Formulom (2) data je, već pominjana, klasična definicija vjerovatnoće a priori. U konstrukciji smo koristili model koji se opire na aksiomatiku Teorije vjerovatnoća.

Formula (2) se primjenjuje u situacijama kada opit ima konačno mnogo ishoda za koje se pretpostavlja da su simetrični (ravnopravni). "Simetrija" je najčešće posljedica prirode slučajnog opita čiji su ishodi reprezentovani skupom Ω . Na primjer, ako bacamo kocku, smatraćemo da je ona geometrijski simetrična tj. da svaki broj ima vjerovatnoću $\frac{1}{6}$.

Pošto je $\overline{A}$ broj elemenata skupa A tj. broj povoljnih ishoda za događaj A , a $\overline{\Omega}$ broj mogućih ishoda u opitu, to se formula (2) može pročitati na sljedeći način: "vjerovatnoća je količnik broja povoljnih i broja mogućih ishoda".

Vratimo se za trenutak na **primjer 8** iz odjeljka 1.1. Kako imamo 36 ravnopravnih (simetričnih) ishoda, prirodno je svakom ishodu dodijeliti istu vjerovatnosnu mjeru, tj. vjerovatnoću koja iznosi $1/36$. Takođe, prirodno je za događaje proglasiti sve podskupove prostora ishoda Ω , te je $\mathcal{F} = \mathcal{P}(\Omega)$. Na osnovu gore izloženog, vjerovatnoća proizvoljnog događaja iz $\mathcal{F}$ je količnik broja ishoda koji odgovaraju događaju i broja 36. Recimo, $P(A) = 3/36$. Na opisani način je modeliran naš slučajni opit. Preciznije, formirana je trojka $(\Omega, \mathcal{F}, P)$ koja predstavlja matematičku formalizaciju opita.

Modeli razmještanja r kuglica u n kutija. Pretpostavimo da r različitih kuglica treba razmjestiti u n kutija, pri čemu svaka kutija može primiti proizvoljno mnogo kuglica. Ako je i_1 redni broj kutije u koju se smješta prva kuglica, i_2 redni broj kutije u koju se smješta druga kuglica, ..., i_r redni broj kutije u koju se smješta r -ta kuglica, tada razmještanju odgovara r -torka $(i_1, i_2, \dots, i_r)$. Na osnovu principa proizvoda, zaključujemo da

ovakvih r -torki, a samim tim i razmještanja ima $n \cdot n \cdots n = n^r$ (svaku koordinatu možemo napraviti na n načina).

U fizičkom modelu u kojem se kutije interpretiraju kao energetske nivoe, a kuglice kao čestice, Maksvel i Bolcman su formulisali pretpostavku da su svi razmještaži (znamo da ih ima n^r) jednako vjerovatni. Model sa ovakvom pretpostavkom je poznat kao Maksvel–Bolcmanov. Nađimo vjerovatnoću da se u Maksvel–Bolcmanovom modelu

a) na $1, 2, \dots, n$ -tom nivou nalazi redom $n_1, n_2, \dots, n_n$ ($n_1 + n_2 + \dots + n_n = r$) čestica,

b) na prvom nivou nalazi k , $0 \leq k \leq r$, čestica.

►a) n_1 čestica za prvi nivo možemo izdvojiti na $\binom{n}{n_1}$ načina. Iz preostalih $n - n_1$ čestica za drugi nivo možemo izdvojiti n_2 čestica na $\binom{n-n_1}{n_2}$ načina itd. Na osnovu principa proizvoda dobijamo da je tražena vjerovatnoća

$$p = \frac{\binom{n}{n_1} \binom{n-n_1}{n_2} \dots \binom{n-n_1-\dots-n_{n-1}}{n_n}}{n^r}. \blacktriangleleft$$

►b) k čestica za prvi nivo možemo izdvojiti na $\binom{n}{k}$ načina, a preostalih $r - k$ čestica možemo razmjestiti na preostalih $n - 1$ nivoa na $(n - 1)^{r-k}$ načina. Na osnovu principa proizvoda dobijamo da je tražena vjerovatnoća

$$p = \frac{\binom{n}{k} (n - 1)^{r-k}}{n^r}. \blacktriangleleft$$

Ako kuglice uzajamno ne razlikujemo i ako svaka kutija može primiti proizvoljno mnogo kuglica, tada razmještanja r kuglica u n kutija ima $\binom{n+r-1}{n-1}$. Rezultat se dobija primjenom analize iz primjera 8 odjeljka 1.4. Treba samo uočiti analogiju na relaciji loptica – kuglica i dječak – kutija. U modelu sa r identičnih čestica i n energetskih nivoa, Boze i Ajnštajn formulišu pretpostavku o jednakoj vjerovatnoći svih razmještaja.

U Boze–Ajnštajnovom modelu, izračunajmo vjerovatnoće koje su već nađene za Maksvel–Bolcmanov model.

►a) Kako je podjelu od $n_1, n_1, \dots, n_n$ čestica po nivoima moguće obaviti samo na jedan način, tražena vjerovatnoća je

$$p = \frac{1}{\binom{n+r-1}{n-1}}. \blacktriangleleft$$

►b) Pošto na prvi nivo smjestimo k čestica, preostalih $r - k$ čestica na preostalih $n - 1$ nivoa možemo smjestiti na $\binom{r-k+n-2}{n-2}$ načina. Dakle, tražena vjerovatnoća je

$$p = \frac{\binom{r-k+n-2}{n-2}}{\binom{n+r-1}{n-1}}. \blacktriangleleft$$

Neka je $0 \leq r \leq n$, neka su kuglice identične i neka svaka kutija može primiti najviše jednu kuglicu. Tada je broj razmještaja $\binom{n}{r}$, zato što iz skupa od n kutija treba odabrati r kutija u koje će biti ubačena po jedna kuglica.

Uz već opisanu fizičku interpretaciju, Fermi–Dirakova pretpostavka se sastoji u tome da su sva ova razmještanja r čestica na n nivoa, jednako vjerovatna.

Primjer 1. 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 $p_A = \frac{4 \cdot 2 + 2 \cdot 6}{36} = \frac{5}{9}$ (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 otpočinu 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. (Zadatak o rasijanom dekanu.) Na svečanoj podjeli 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 $W = A_1 \cup A_2 \cup \dots \cup A_n$. Da bismo našli $P(W)$ primijenimo 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 $P(A_1 A_2)$. 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 $n \rightarrow \infty$, dobijeni zbir teži ka $1 - e^{-1} \approx 0,6322$. Zbog brze konvergencije niza – zbira, u slučaju kada je $n \geq 20$, 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 formulirati u mnogo varijanti. Navedimo jednu. Složili smo 100 listova, zaduvala 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 3. Komplet u kome 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. ◀

Preporučujemo čitaocu da uradi sljedeće zadatke. Smatramo da će se tokom rada kod čitaoca razviti kombinatorno–vjerovatnosni način razmišljanja. Sigurni smo da će rješavajući ponuđene zadatke, čitalac još dublje

proniknuti u teoriju.

Zadaci

1. Slučajno se formira niz od 5 cifara. Kolika je vjerovatnoća da su u tom nizu sve cifre različite?

2. n kuglica se razmješta u n kutija. Kolika je vjerovatnoća da se u svakoj kutiji nalazi kuglica?

3. Svaka od 50 država ima po 2 predstavnika u senatu. Na jednoj sjednici je prisutno 50 senatora.

a) Kolika je vjerovatnoća da je na toj sjednici zastupljena država A?

b) Kolika je vjerovatnoća da su zastupljene sve države?

4. U kutiji se nalazi 10 B (bijelih) i 8 C (crnih) kuglica. Iz kutije se odjednom (model bez vraćanja) vadi 6 kuglica. Naći vjerovatnoću događaja da se među njima nalaze tačno 4 B.

5. U kutiji se nalazi 5 kuglica različite boje. Iz kutije se 25 puta sa vraćanjem vadi kuglica. Naći vjerovatnoću događaja da među izvađenim kuglicama bude po 5 kuglica svake boje.

6. Nedirigovani ždrijeb 32 atletičara dijeli u četiri kvalifikacione grupe. Kolika je vjerovatnoća da četiri najbolja atletičara budu u različitim grupama?

7. Iz grupe od 10 bračnih parova, slučajno se bira 8 osoba. Kolika je vjerovatnoća da među njima ne postoji bračni par?

8. Naći vjerovatnoću događaja da su u kompletu od 52 karte a) prve četiri karte asovi, b) prva i posljednja karta asovi?

9. Svaki od n gostiju iz n pari cipela slučajno bira za sebe po dvije cipele. Kolika je vjerovatnoća događaja da svaki gost izabere po jednu desnu i jednu lijevu cipelu (ne obavezno iz istog para)?

10. Iz kompleta od 32 karte se slučajno vadi (bez vraćanja) 6 karata. Kolika je vjerovatnoća da su među izvađenim kartama a) tačno 3 boje, b) bar 2 asa?

11. Neka osoba u svežnju ima n ključeva, a među njima samo jedan otvara vrata. U namjeri da otključa vrata, proba jedan po jedan ključ. Kolika je vjerovatnoća da će mu do izbora pravog ključa biti potrebno k pokušaja?

12. Kolika je vjerovatnoća da dva bacanja tri kocke daju isti rezultat ako se a) bacaju iste kocke b) različite kocke?

13. (Paradoks de Mere.) Pokazati da je vjerovatnije dobiti bar jednu jedinicu pri bacanju 4 kocke, nego bar dvije jedinice pri 24 bacanja kocke.

14. Novčić se baca deset puta. Izračunati vjerovatnoću da se bar pet puta za redom pojavi pismo.

15. U kutiji se nalazi 10 kuglica numerisanih brojevima od 1 do 10. Na slučajan način izvlačimo (model bez vraćanja) 5 kuglica. Izračunati vjerovatnoću da broj 5 bude drugi po veličini među izvađenim brojevima.

16. 10 žena i 10 muškaraca sjedaju na 20 mjesta koja čine red u pozorištu. Kolika je vjerovatnoća da će sve žene biti jedna pored druge?

17. n kuglica se na slučajan način razmješta u n kutija (kuglice razlikujemo i svaka kutija može primiti proizvoljno mnogo kuglica). Kolika je vjerovatnoća da tačno jedna kutija ostane prazna?

18. Kocka za igru se baca 20 puta. Kolika je vjerovatnoća da se među palim brojevima pojavljuju tačno dva broja?

19. Iz kutije u kojoj se nalazi m bijelih i n crnih kuglica, $m > n$, vadi se jedna po jedna kuglica (model bez vraćanja). Kolika je vjerovatnoća nastupanja momenta kada će broj izvađenih crnih kuglica biti jednak broju izvađenih bijelih kuglica?

20. Baca se 10 kockica. Odrediti vjerovatnoću da a) ne padne ni jedna šestica, b) padnu tačno 3 šestice?

21. Novčić se baca 20 puta. Kolika je vjerovatnoća da se pojavi jednak broj pisama i grbova? Kolika je vjerovatnoća da se pojavi više grbova nego pisama?

22. U voz sa 10 vagona ulazi, slučajno birajući vagon, 7 putnika. Kolika je vjerovatnoća da su putnici ušli u različite vagone?

23. U kutiji se nalazi 100 kuglica i među njima je 5B i 95C kuglica. Po modelu bez vraćanja izvlači se 96 kuglica. Kolika je vjerovatnoća da su među izvučenim kuglicama bar tri kuglice bijele boje?

24. Kocka se baca deset puta. Naći vjerovatnoću da se šestica pojavi pet puta, petica tri puta i četvorka dva puta.

25. Iz skupa $\{1, 2, \dots, 500\}$ slučajno se bira broj. Kolika je vjerovatnoća da izabrani broj nije djeljiv ni sa 2, ni sa 3, ni sa 5?

26. Kocka se baca tri puta. Kolika je vjerovatnoća da među palim brojevima postoji broj koji je jednak zbiru dva ostala broja?

27. Za test je pripremljeno 10 pitanja i za svako pitanje su ponuđena tri odgovora od kojih je samo jedan tačan. Kolika je vjerovatnoća da kandidat koji nasumce zaokružuje odgovore, ima bar pet tačnih odgovora?

28. U igri je učestvovalo 10 dječaka. Oni su izgubili 5 klikera, a zatim su svi izgubljeni klikeri pronađeni. Kolika je vjerovatnoća da su klikere pronašla trojica dječaka?

1.6 Geometrijska vjerovatnoća

Klasična definicija vjerovatnoće obuhvata samo opite sa konačno mnogo ravnopravnih ishoda. Još nekoliko vjekova prije pojave Kolmogorovljeve aksiomatike, analiza nekih vjerovatnosnih modela u kojima je prostor ishoda neprebrojiv, nametnula je potrebu da se definicija vjerovatnoće proširi. Pomenuti modeli su uglavnom geometrijske prirode, tako da je prvo uopštenje bila geometrijska vjerovatnoća. I u definiciju geometrijske vjerovatnoće je ugrađen pojam "jednake vjerovatnoće" nekih događaja. Upoznaćemo se sa zadatkom čije je rješavanje dovelo do proširenja pojma vjerovatnoće.

Slika 1: Geometrijska vjerovatnoća.

Neka je G oblast u ravni i $g \subset G$ podoblast. Slučajno se bira tačka u oblasti G . Kolika je vjerovatnoća da je izabrana tačka iz g ? Frazi "slučajno se bira tačka iz G " pridajemo sljedeći smisao: izabrana tačka može biti bilo koja tačka iz G , a vjerovatnoća događaja da izabrana tačka pripada nekoj fiksiranoj podoblasti, proporcionalna je površini podoblasti i ne zavisi od položaja podoblasti u oblasti. U ovakvim uslovima, prirodno je vjerovatnoću događaja da je izabrana tačka iz podoblasti g definisati kao

$$p = \frac{\text{mes } g}{\text{mes } G}. \quad (1)$$

Na ovaj način je definisana geometrijska vjerovatnoća. Kako je oblast G iz ravni, mes – mjera je površina. Naravno, izloženo se može odnositi i na oblasti na pravoj i u prostoru. U slučaju kada rješavamo zadatak na pravoj, mes je dužina, a kada rješavamo zadatak u prostoru, mes je zapremina.

Navešćemo nekoliko već klasičnih primjera iz geometrijske vjerovatnoće.

Primjer 1. (Zadatak o susretu.) Dvije osobe (imenujmo ih sa A i B) se dogovore da se nađu na nekom mjestu između 12^h i 13^h . Osoba koja dođe na mjesto susreta čeka dok se ne pojavi druga osoba, ali najviše 20 minuta i onda odlazi. Kolika je vjerovatnoća da će do susreta doći?

► Neka je x vrijeme koje prođe od 12^h do dolaska na mjesto susreta osobe A, a y vrijeme koje prođe od 12^h do dolaska na mjesto susreta osobe B. Ako za

mjernu jedinicu uzmemo jedan sat, prirodno je prostor ishoda identifikovati sa

$$\Omega = \{(x, y) : x, y \in [0, 1]\}.$$

Označimo sa W događaj da će doći do susreta. Do susreta će doći ako je

$$0 \leq |x - y| \leq \frac{1}{3} \Leftrightarrow 0 \leq \max\{x, y\} - \min\{x, y\} \leq \frac{1}{3},$$

pa je

$$W = \left\{ (x, y) : 0 \leq |x - y| \leq \frac{1}{3} \right\}, \quad W \subset \Omega.$$

Iz formule (1) dobijamo (W je osjenčeni dio na slici)

$$p = \frac{\text{mes } W}{\text{mes } \Omega} = \frac{5}{9}.$$

Slika 2: Primjer 1.

Postoje mnoga uopštenja ovog zadatka. Preporučujemo čitaocu da izračuna vjerovatnoću susreta tri osobe koje dolaze na utvrđeno mjesto, a vrijeme čekanja je 10 minuta. Rješavanje predloženog zadatka traži sprovođenje analize u prostoru. ◀

Primjer 2. (Bifonov problem igle.) Na ravan koja je išrafirana paralelnim pravima, rastojanje između pravih je $2a$, bačena je igla dužine $2l$, $l < a$. Kolika je vjerovatnoća da igla zasiječe neku pravu?

► Označimo sa x rastojanje od centra igle do najbliže "paralele", a sa α ugao između igle i "paralele". Prirodno je prostor ishoda identifikovati sa

$$\Omega = \{(x, \alpha) : x \in [0, a], \alpha \in [0, \pi]\}.$$

Igla će zasjeći neku pravu ako je $x \leq l \sin \alpha$, pa ako sa A označimo događaj da će igla zasjeći neku pravu, imamo

$$A = \{(x, \alpha) : x \leq l \sin \alpha\}, \quad A \subset \Omega.$$

Iz formule (1) dobijamo (A je osjenčeni dio na slici)

$$p = \frac{\text{mes } A}{\text{mes } \Omega} = \frac{1}{2\pi} \int_0^\pi l \sin \alpha \, d\alpha = \frac{2l}{a\pi}.$$

Slika 3: Primjer 2.

Formulisaćemo zadatak koji se nadovezuje na Bifonov problem.

Kolika je vjerovatnoća da pravilni šestougao dužine stranice a koji se baca na ravan opisanu u Bifonovom problemu, zasiječe neku pravu? ◀

Primjer 3. Štap dužine 1 se lomi na tri dijela. Kolika je vjerovatnoća događaja da se od ta tri dijela može sastaviti (sklopiti) trougao? A oštrougli trougao?

► Tačke u kojima dolazi do lomljenja označimo sa C i D (C je bliža tački A). Neka je $x = |CD|$ i $y = |DB|$. Prirodno je prostor ishoda identifikovati sa

$$\Omega = \{(x, y) : x, y \in [0, 1], x + y \leq 1\}.$$

Od duži AC , CD i DB možemo sastaviti trougao ako je

$$\begin{aligned} x + y > 1 - x - y &\Leftrightarrow x + y > \frac{1}{2}, \\ x + 1 - x - y > y &\Leftrightarrow y < \frac{1}{2}, \\ y + 1 - x - y > x &\Leftrightarrow x < \frac{1}{2}. \end{aligned}$$

Znači $A = \{(x, y) : x + y > \frac{1}{2}, y < \frac{1}{2}, x < \frac{1}{2}\}$, $A \subset \Omega$, je događaj čiju vjerovatnoću tražimo. Očigledno, $p = \frac{1}{4}$. U slučaju oštrouglog trougla, odgovarajuća oblast je određena uslovima

$$\begin{aligned} x^2 + y^2 &< (1 - x - y)^2, \\ x^2 + (1 - x - y)^2 &< y^2, \\ y^2 + (1 - x - y)^2 &< x^2 \end{aligned}$$

Slika 4: Primjer 3.

i na slici ?? je omeđena krivim linijama.

Primjenom integralnog računa se dobija da je tražena vjerovatnoća $p_O = 3 \ln 2 - 2 \approx 0,082$. Znači, vjerovatnoća za tupougli trougao je $p_T \approx 0,168$, a za pravougli $p_P = 0$ (pravouglom trouglu odgovara kriva linija sa slike ??, a njena površina (mjera) je 0) .◀

Zadaci

1. U segmentu $[-1, 1]$ slučajno se biraju dva broja p i q . Kolika je vjerovatnoća da kvadratna jednačina $x^2 + px + q = 0$ ima realna rješenja?
2. U jednakokraničnom trouglu se slučajno bira tačka. Kolika je vjerovatnoća da je izabrana tačka bliža stranici nego visini trougla?
3. U segmentu $[0, 1]$ slučajno se biraju tri tačke. Kolika je vjerovatnoća da je tačka koja je izabrana kao treća po redu, između dvije prvoizabrane?
4. U segmentu $[0, 1]$ slučajno se biraju tri tačke i označavaju sa A , B i C . Kolika je vjerovatnoća događaja da se od duži OA , OB i OC može sastaviti trougao?
5. Na kružnici se slučajno biraju tri tačke. Kolika je vjerovatnoća da je trougao određen izabranim tačkama, oštrogli?

1.7 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 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 ≤ 6 . Konstatujemo

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

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 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 $P(A|B)$. Pošto je izvjesno da je B 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 $\text{mes } K = 1$.

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

$$P_B(A) = P(A|B) = \frac{P(AB)}{P(B)}, \quad A \in \mathcal{F}. \quad (1)$$

Lako se provjerava da je P_B vjerovatnoća na $\mathcal{F}$. Nazivamo je uslovna vjerovatnoća uz uslov B . Druga jednakost u formuli (1) služi kao obrazac

Slika 5: Primjer 2.

za računanje uslovne vjerovatnoće. Pošto su nam primjeri 1,2 u svježem sjećanju, jasno je odakle potiče motiv za drugu jednakost u (1).

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

Kako je P_B vjerovatnoća na $\mathcal{F}$, iz teoreme 1.2.1 slijedi

Teorema 1.7.1 *Neka je $(\Omega, \mathcal{F}, P)$ prostor vjerovatnoće i neka je $B \in \mathcal{F}$, $P(B) > 0$. Tada je*

- (a) $P(\emptyset|B) = 0$.
- (b) $P(A^c|B) = 1 - P(A|B)$, $A \in \mathcal{F}$.
- (c) $A_1, A_2 \in \mathcal{F} \implies P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 A_2|B)$.
- (d) $A_1, A_2 \in \mathcal{F}$, $A_1 \subseteq A_2 \implies P(A_1|B) \leq P(A_2|B)$.
- (e) $A_n \in \mathcal{F}, n \in \mathbb{N} \implies P\left(\bigcup_{n=1}^{\infty} A_n|B\right) \leq \sum_{n=1}^{\infty} P(A_n|B)$.

Iz jednakosti (1) slijedi da uz uslove $P(A) > 0$, $P(B) > 0$ važi **formula množenja vjerovatnoća**

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

DEFINICIJA 1.7.1 *Konačna ili prebrojiva familija $H_i, i = 1, 2, \dots$, događaja u prostoru vjerovatnoće $(\Omega, \mathcal{F}, P)$ je **potpuni sistem događaja** ako je $P(H_i) > 0$ za svako i , $H_i H_j = \emptyset, i \neq j$ i $\bigcup_i H_i = \Omega$.*

Sljedeće dvije teoreme imaju veliku teorijsku i operativnu težinu.

Teorema 1.7.2 (Formula potpune vjerovatnoće.) *Neka je $H_i, i = 1, 2, \dots$ potpuni sistem događaja u vjerovatnosnom prostoru $(\Omega, \mathcal{F}, P)$. Tada za svako $A \in \mathcal{F}$ vrijedi*

$$P(A) = \sum_i P(A|H_i)P(H_i).$$

Dokaz. Za $A \in \mathcal{F}$ vrijedi

$$\begin{aligned} P(A) &= P(A\Omega) = P\left(A \bigcup_i H_i\right) = P\left(\bigcup_i (AH_i)\right) \\ &= \sum_i P(AH_i) = \sum_i P(A|H_i)P(H_i). \blacklozenge \end{aligned}$$

Teorema 1.7.3 (Bajesova formula.) *Neka je $H_i, i = 1, 2, \dots$, potpuni sistem događaja u vjerovatnosnom prostoru $(\Omega, \mathcal{F}, P)$ i neka je $A \in \mathcal{F}$, $P(A) > 0$. Tada za svako i vrijedi*

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{\sum_j P(A|H_j)P(H_j)}.$$

Dokaz. Iz formule proizvoda vjerovatnoća i formule potpune vjerovatnoće slijedi

$$P(H_i|A) = \frac{P(AH_i)}{P(A)} = \frac{P(A|H_i)P(H_i)}{\sum_j P(A|H_j)P(H_j)}. \blacklozenge$$

Pošto događaji iz sistema $H_i, i = 1, 2, \dots$, čine jednu potpunu kolekciju mogućih događaja, govorićemo da su H_i **hipoteze**.

Razmotrićemo slučaj kada je

$$P(A|B) = P(A), \quad (2)$$

tj. uslovna vjerovatnoća događaja A pri uslovu B jednaka je bezuslovnoj. Izvjesnost realizacije događaja B se ne reflektuje na vjerovatnoću događaja A te možemo konstatovati da A ne zavisi od B . Iz jednakosti (??) i (??) slijedi

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

Znači, ni B ne zavisi od A . Ova analiza je urađena uz uslove da je $P(A) > 0$ i $P(B) > 0$.

Obje jednakosti $P(A|B) = P(A)$ i $P(B|A) = P(B)$ su ekvivalentne sa jednakošću $P(AB) = P(A)P(B)$. Sada imamo dovoljno motiva za sljedeću definiciju.

DEFINICIJA 1.7.2 *Događaji A i B su nezavisni ako je*

$$P(AB) = P(A)P(B). \quad (3)$$

Prilikom definisanja pojma nezavisnosti događaja, odbačeni su ograničavajući uslovi $P(A) > 0$ i $P(B) > 0$.

Svi pojmovi koje smo do uvođenja pojma nezavisnosti pomenuli, imaju svoje analogone u Teoriji mjere. Nezavisnost je izvorno vjerovatnosni pojam. Pojavom pojma nezavisnosti, vjerovatnoća se oslobodila od okrilja Teorije mjere (Teorija mjere je oblast Matematičke analize).

Kod rješavanja praktičnih vjerovatnosnih zadataka, nezavisnost se rijetko ustanovljava provjerom važenja jednakosti (??). Obično se nezavisnost 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 (??) (pogledati primjer 5).

U vjerovatnoći se pojavljuje potreba za razmatranjem nezavisnosti više događaja. Daćemo strogo formalnu definiciju pojma potpune nezavisnosti kompleksa od konačno mnogo događaja.

DEFINICIJA 1.7.3 *Događaji $A_1, A_2, \dots, A_n$ su potpuno nezavisni ako za svaki broj $k \in \{2, 3, \dots, n\}$ i svaki izbor indeksa $1 \leq j_1 < j_2 < \dots < j_k \leq n$, važi jednakost*

$$P(A_{j_1} A_{j_2} \dots A_{j_k}) = P(A_{j_1}) P(A_{j_2}) \dots P(A_{j_k}).$$

Izloženu teoriju ćemo ilustrovati primjerima.

Primjer 1. Profesor je za ispit pripremio 15 cedulja. Studentu odgovara 10 cedulja (zna odgovore na pitanja koja se nalaze na tim ceduljama) i ne odgovara mu 5 cedulja (ne zna odgovore). Što je vjerovatnije: da student izvuče "dobru" cedulju kada izvlači cedulju kao prvi po redu ili kao drugi po redu?

► Neka je A događaj da student izvuče "dobru" cedulju kada cedulju izvlači kao prvi. Očigledno, $P(A) = \frac{10}{15} = \frac{2}{3}$. Neka je B događaj da izvuče "dobru" cedulju kada cedulju izvlači kao drugi po redu. Neka je H_1 : prvoizvučena cedulja je "dobra", a H_2 : prvoizvučena cedulja je "loša". Primjenom formule potpune vjerovatnoće dobijamo

$$P(B) = P(B|H_1)P(H_1) + P(B|H_2)P(H_2) = \frac{2}{3} \cdot \frac{9}{14} + \frac{1}{3} \cdot \frac{10}{14} = \frac{2}{3}.$$

Znači, vjerovatnoće su iste. Vjerovatnoća (provjerite) da student izvuče "dobru" cedulju kada je izvlači kao posljednji je takođe $\frac{2}{3}$.

Našem modelu možemo dati i sljedeću interpretaciju. Dva žetona na kojima su brojevi 1 i 2 spuštaju se na dvije cedulje. Nevažno je da li se žetoni spuštaju prvo jedan pa drugi ili pak istovremeno, efekat je isti (kao što je u modelu bez vraćanja kada recimo izvlačimo nekoliko kuglica, sve jedno da li smo ih izvukli jednu po jednu ili sve odjednom). Cedulja na koju je spušten žeton sa brojem 1 će biti prvoizvađena, a ona na koju je spušten žeton sa brojem 2 će biti drugoizvađena. Jasno, žeton sa npr. brojem 2 će se spustiti na jednu od 15 cedulja, a vjerovatnoća da će to biti "dobra" cedulja je očigledno $\frac{2}{3}$. Na ovaj način je nađena vjerovatnoća da je drugoizvađena cedulja "dobra". Kod traženja vjerovatnoće da izvučemo "dobru" cedulju kada je izvlačimo kao posljednji po redu, primjenjujemo isti način rasuđivanja, s tim što spuštamo 15 žetona numerisanih brojevima od 1 do 15. ◀

Primjer 2. 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

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

Dakle, vjerovatnoće prije i poslije prebacivanja su iste. ◀

Primjer 3. 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

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

Primjer 4. 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 najvjerojatniji 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

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

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 $k \leq 32$ i $P(A|H_k) > P(A|H_{k+1})$ za $k \geq 33$. Oдавde zaključujemo da je najvjerojatniji 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 5. 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?

► $P(A) = \frac{4}{32} = \frac{1}{8}$, $P(B) = \frac{8}{32} = \frac{1}{4}$, $P(AB) = \frac{1}{32}$ (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 džoker? ◀

Primjer 6. 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 . Imamo

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

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

Primjer 7. Kocka se baca dva puta. Izdvojimo događaje

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

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

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

Primjer 8. 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 $p_A = \frac{13}{16}$. ◀

Zadaci

1. U prvoj kutiji se nalazi 10 B, 5 C i jedna zelena kuglica, a u drugoj 5 B i 10 C kuglica. Slučajno se bira jedna kutija i iz nje se po modelu bez vraćanja vade dvije kuglice. Ako su obje izvađene kuglice bijele boje, kolika je vjerovatnoća da je izabrana prva kutija?

2. U prvoj kutiji se nalazi 8B i 7C, a u drugoj 6B i 9C kuglica. Iz prve kutije se u drugu prebacuju dvije kuglice. Nakon toga se iz druge kutije vadi po modelu bez vraćanja 5 kuglica. Naći vjerovatnoću da se među njima nalaze tačno 3 bijele kuglice.

3. Postoje tri kutije. U jednoj je 5 B i 6 C kuglica, u drugoj 6 C i 7 B, a u trećoj su samo bijele kuglice. Neko slučajno bira kutiju i iz nje izvlači jednu kuglicu. Kolika je vjerovatnoća da je izvučena bijela kuglica?

4. U uslovima iz zadatka **3.** je izvučena bijela kuglica. Kolika je vjerovatnoća da je izvučena iz a) prve, b) druge, c) treće kutije?

5. U kutiji se nalazi 5 bijelih, 8 crvenih i 12 plavih kuglica. Po modelu a) sa vraćanjem, b) bez vraćanja, iz kutije se izvlači jedna po jedna kuglica. Kolika je vjerovatnoća da će crvena kuglica biti izvučena prije plave?

6. U prvoj kutiji su 3B i 4C kuglice, a u drugoj su 3C i 4B kuglice. Iz prve kutije se vadi jedna kuglica, iz druge kutije se vade dvije kuglice i tri

izvađene kuglice se prebacuju u treću kutiju koja je bila prazna. Zatim se, iz treće kutije vadi kuglica. Kolika je vjerovatnoća da je bijela?

7. Kocka za igru se baca dva puta. Ako je u oba bacanja pao broj iste parnosti, izračunati vjerovatnoću da je zbir palih brojeva manji od pet.

8. Iz kompleta od 32 karte su odjednom izabrane 2 karte. Ako je jedna od izabranih karata kupa, izračunati vjerovatnoću da su obje karte kupe.

9. Na dva polja šahovske table su postavljene dame. Kolika je vjerovatnoća da se te dvije dame napadaju?

10. Bacaju se tri kocke. Kolika je vjerovatnoća da je pala bar jedna šestica ako su pali različiti brojevi?

11. Izvjesno je da se pri bacanju 10 kocaka pojavila bar jedna šestica. Kolika je vjerovatnoća da su pale bar dvije šestice?

12. Osobe A i B igraju šah sve dok jedan od igrača ne ostvari 6 pobjeda. Rezultati remija se ne unose u rezultat (drugim riječima, nakon remija saldi igrača se ne uvećavaju za pola poena). Vjerovatnoće pobjede igrača A, igrača B i remija su redom p, q, r . Naći vjerovatnoću događaja

a) Igrač A je pobijedio rezultatom 6 : 0.

b) Igra je trajala tačno n partija.

13. Na teniskom turniru učestvuje 128 igrača i u svakom meču pobjeđuje bolje rangirani igrač. Ždrijeb je nedirigovan i igra se po kup sistemu. Kolika je vjerovatnoća da se u finalu sastanu dva najbolje rangirana igrača?

1.8 Slučajne veličine

Konstatovali smo da je vjerovatnosni prostor $(\Omega, \mathcal{F}, P)$ matematički model slučajnog opita. Već smo pominjali opite u kojima su prostore ishoda Ω obrazovali neki "nematematički" objekti. U prvom odjeljku smo radili sa opitima u kojima su elementi prostora ishoda bili strane novčića ili nizovi strana. Prirodna je bila potreba da se sa skupa Ω pređe na, u matematici najbolje izučeni, skup realnih brojeva R . Taj prelazak je bio ostvaren uz posredovanje funkcije koja se naziva slučajna veličina.

DEFINICIJA 1.8.1 *Neka je $(\Omega, \mathcal{F}, P)$ vjerovatnosni prostor. Funkcija $X : \Omega \rightarrow R$ je slučajna veličina (na Ω) ako je $X^{-1}(B) \in \mathcal{F}$ za proizvoljnu skup $B \in \mathcal{B}$.*

Sa X^{-1} označavamo inverznu funkciju dok je sa $\mathcal{B}$ označena Borelova σ -algebra.

Iz teoreme koja slijedi (nećemo je dokazivati) zaključujemo da se slučajna veličina može definisati tako što će se uslov $X^{-1}(B) \in \mathcal{F}$, zamijeniti sa nekim od pet navedenih.

Teorema 1.8.1 *Neka je $(\Omega, \mathcal{F}, P)$ vjerovatnosni prostor i $X : \Omega \longrightarrow R$. Svaki od navedenih uslova potreban je i dovoljan da bi preslikavanje X bilo slučajna veličina.*

- a) $X^{-1}(G) \in \mathcal{F}$ za svaki otvoreni skup $G \subset R$.
- b) $X^{-1}((-\infty, x)) = \{X < x\} \in \mathcal{F}$, za svako $x \in R$.
- c) $X^{-1}((-\infty, x]) = \{X \leq x\} \in \mathcal{F}$, za svako $x \in R$.
- d) $X^{-1}([x, \infty)) = \{X \geq x\} \in \mathcal{F}$, za svako $x \in R$.
- e) $X^{-1}((x, \infty)) = \{X > x\} \in \mathcal{F}$, za svako $x \in R$.

Sljedećom teoremom ćemo zadati jednu značajnu σ -algebru.

Teorema 1.8.2 *Neka je X slučajna veličina zadata na vjerovatnosnom prostoru $(\Omega, \mathcal{F}, P)$. Tada je familija*

$$\sigma(X) = \{X^{-1}(B) : B \in \mathcal{B}\}$$

σ -algebra.

Dokaz.

a) $\emptyset = X^{-1}(\emptyset) \in \sigma(X)$.

b) Neka je $A \in \sigma(X)$. Tada postoji skup $B \in \mathcal{B}$ takav da je $A = X^{-1}(B)$.

Sada imamo

$$A^c = (X^{-1}(B))^c = X^{-1}(B^c) \in \sigma(X).$$

c) Neka je A_n , $n \in N$, niz događaja iz $\mathcal{F}$. Zbog toga što je X slučajna veličina, za svako $n \in N$ postoji skup $B_n \in \mathcal{B}$ takav da je $A_n = X^{-1}(B_n)$. Sada imamo

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} X^{-1}(B_n) = X^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right) \in \sigma(X).$$

Ovim je dokazano da je $\sigma(X)$ σ -algebra. $\sigma(X)$ je podalgebra σ -algebre $\mathcal{F}$ i naziva se **σ -algebra generisana slučajnom veličinom X** . ♦

Neka je X slučajna veličina zadata na vjerovatnosnom prostoru $(\Omega, \mathcal{F}, P)$ i neka je $P_X : \mathcal{B} \longrightarrow [0, 1]$ funkcija definisana sa

$$P_X(B) = P\{\omega : X(\omega) \in B\} = P(X^{-1}(B)). \quad (1)$$

Pokažimo da je $P_X(\cdot)$ vjerovatnoća na mjerljivom prostoru $(R, \mathcal{B})$.

a) $P_X(B) \geq 0$, $B \in \mathcal{B}$ i $P_X(R) = P(X^{-1}(R)) = P(\Omega) = 1$.

b) Neka je B_i , $i \in N$, $B_i \cap B_j = \emptyset$, $i \neq j$. Sada je

$$P_X\left(\bigcup_i B_i\right) = P\left(X^{-1}\left(\bigcup_i B_i\right)\right) = P\left(\bigcup_i X^{-1}(B_i)\right) = \sum_i P_X(B_i).$$

DEFINICIJA 1.8.2 *Funkcija $P_X : \mathcal{B} \rightarrow R$ definisana pomoću jednakosti (??) naziva se raspodjela vjerovatnoća slučajne veličine X ili kraće raspodjela slučajne veličine X . Vjerovatnosni prostor $(R, \mathcal{B}, P_X)$ naziva se fazni prostor slučajne veličine X .*

Primjer 1. Novčić se baca dva puta. Prostor ishoda je

$$\Omega = \{PP, PG, GP, GG\}.$$

Slučajna veličina X koja predstavlja broj palih pisama data je sa:

$$X(GG) = 0, \quad X(PG) = X(GP) = 1, \quad X(PP) = 2.$$

Slučajna veličina Y koja predstavlja razliku broja palih pisama i grbova data je sa:

$$Y(PP) = 2, \quad Y(PG) = Y(GP) = 0, \quad Y(GG) = -2.$$

Kao što vidimo iz ovog primjera, na jednom vjerovatnosnom prostoru možemo zadati više slučajnih veličina. ◀

Primjer 2. Neka je $(\Omega, \mathcal{F}, P)$ vjerovatnosni prostor,

$$\Omega = \{a, b, c\}, \quad \mathcal{F} = \{\{a, b\}, \{c\}, \Omega, \emptyset\}, \quad P(\{a, b\}) = \frac{1}{2}, \quad P(\{c\}) = \frac{1}{2}.$$

Neka je preslikavanje $X : \Omega \rightarrow R$ zadato sa

$$X(a) = X(c) = 0, \quad X(b) = 1.$$

Tada je $X^{-1}(1) = \{b\} \notin \mathcal{F}$. Dakle, X nije slučajna veličina. Ovim primjerom je pokazano da postoje preslikavanja iz Ω u R koja nisu slučajne veličine. ◀

Primjer 3. Baca se kocka. Zadaćemo slučajne veličine:

a) $X(1) = X(2) = X(3) = X(4) = X(5) = X(6) = 0,$

b) $Y(1) = Y(3) = Y(5) = 0, \quad Y(2) = Y(4) = Y(6) = 1,$

c) $Z(1) = 1, \quad Z(2) = 2, \quad Z(3) = 3, \quad Z(4) = 4, \quad Z(5) = 5, \quad Z(6) = 6.$

Imamo

$$\sigma(X) = \{\emptyset, \Omega\}, \quad \sigma(Y) = \{\{1, 3, 5\}, \{2, 4, 6\}, \Omega, \emptyset\}, \quad \sigma(Z) = \mathcal{P}(\Omega). \quad \blacktriangleleft$$

Primjer 4. Neka je $A \in \mathcal{F}$ i neka je

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

Slučajna veličina I_A se naziva **indikator događaja A** . ◀

Primjer 5. Neka je $\Omega = A_1 \cup A_2 \cup \dots$ razbijanje skupa Ω na međusobno disjunktne događaje $A_1, A_2, \dots$ i neka su $x_1, x_2, \dots$ realni brojevi. Funkcija $X : \Omega \rightarrow R$ data sa

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

naziva se **diskretna slučajna veličina**. Ako je $\omega \in A_k$, tada je $X(\omega) = x_k$. Ako razbijanje sadrži konačno mnogo događaja, tada se odgovarajuća diskretna slučajna veličina naziva **prosta slučajna veličina**.

Označimo sa p_k vjerovatnoću da slučajna veličina X uzme vrijednost x_k tj. $p_k = P\{X = x_k\} = P\{\omega : X(\omega) = x_k\}$. Raspodjela vjerovatnoća P_X slučajne veličine X određena je nizom parova brojeva $(x_1, p_1), (x_2, p_2), \dots$. Zaista, za svaki Borelov skup B važi $P_X(B) = \sum_k p_k$, pri čemu se sumiranje vrši po svim indeksima k za koje je $x_k \in B$. Kako je $P_X(R) = 1$, to je $\sum_k p_k = 1$. Standardni način bilježenja raspodjele vjerovatnoća diskretne slučajne veličine X je dat tabelom

$$X : \begin{array}{ccccccc} x_1 & x_2 & x_3 & \cdot & \cdot & \cdot & \cdot \\ p_1 & p_2 & p_3 & \cdot & \cdot & \cdot & \cdot \end{array} \quad \blacktriangleleft$$

Primjer 6. U kutiji se nalaze dvije kuglice na kojima je zapisan broj 1 i dvije kuglice na kojima je zapisan broj 2. Slučajno se, po modelu bez vraćanja, vade dvije kuglice. Naći raspodjelu vjerovatnoća slučajne veličine X koja je jednaka zbiru brojeva na izvađenim kuglicama.

► Kuglice označimo sa $1^1, 1^2, 2^1, 2^2$ (gornji indeks označava prvu odnosno drugu kuglicu sa odgovarajućim brojem). Prostor ishoda je

$$\Omega = \{(1^1, 1^2), (1^1, 2^1), (1^1, 2^2), (1^2, 2^1), (1^2, 2^2), (2^1, 2^2)\},$$

pa je

$$X : \begin{array}{ccc} 2 & 3 & 4 \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array}.$$

Primjer 7. Izvodi se serija od n nezavisnih opita. Ta serija ponovljenih opita se tretira kao novi opit. U inicijalnom – polaznom opitu fiksiramo neki događaj A i proglasimo ga uspjehom. Neka je $p(A) = p$. Događaj A^c se naziva neuspjeh i njegovu vjerovatnoću označavamo sa q . Nakon završetka pojedinačnog opita, konstatovaćemo kao njegov ishod uspjeh ili neuspjeh pa ćemo kao ishod serije imati n -torku čiji su elementi uspjesi A i neuspjesi A^c . Upravo opisani model je poznat kao **Bernulijeva shema**. Označimo

sa S_n slučajnu veličinu koja je u Bernulijevoj shemi jednaka broju uspješno završenih opita tj. onih opita u kojima se realizovao događaj A (koristi se i fraza: S_n prebrojava uspjehe). Tada važi

$$P\{S_n = k\} = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, \dots, n.$$

Objasnićemo kako je dobijena gornja jednakost. Događaju $\{\omega : S_n(\omega) = k\}$ odgovaraju sve n -torke – ishodi u kojima ima k simbola A . Takvih n -torki ima $\binom{n}{k}$, a vjerovatnoća svake takve n -torke je zbog nezavisnosti opita jednaka $p^k q^{n-k}$. Za slučajnu veličinu S_n kažemo da je raspodijeljena po **binomnom** zakonu sa parametrima n i p i koristimo zapis $S_n : \mathcal{B}(n, p)$. Imamo,

$$\sum_{k=0}^n P\{S_n = k\} = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1.$$

Uradićemo primjer u kojem se pojavljuje konkretna Bernulijeva shema.

Kocka se baca deset puta. Naći ćemo raspodjelu slučajne veličine X koja je jednaka broju pojavljivanja brojeva djeljivih sa tri.

Opit možemo modelirati kao Bernulijevu shemu sa 10 ponavljanja polaznog opita (baca se kocka i registruje broj nakon njenog padanja) i uspjehom koji se identifikuje sa događajem $A = \{3, 6\}$. Kako je vjerovatnoća uspjeha $p = \frac{1}{3}$, možemo konstatovati $X : \mathcal{B}(10, \frac{1}{3})$, iza čega stoji

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

Primjer 8. U kutiji je 10 bijelih i 12 crnih kuglica. Slučajno se a) po modelu sa vraćanjem, b) po modelu bez vraćanja, vadi 5 kuglica. Naći raspodjelu slučajne veličine X koja je jednaka broju izvađenih bijelih kuglica.

► a) $X : \mathcal{B}(5, \frac{5}{11})$ jer je vjerovatnoća vađenja bijele kuglice u svakom vađenju $\frac{5}{11}$.

b) Iz skupa od 10 bijelih kuglica, k kuglica možemo izvaditi na $\binom{10}{k}$ načina, a iz skupa od 12 crnih kuglica, $5 - k$ kuglica možemo izvaditi na $\binom{12}{5-k}$ načina. Nakon primjene principa proizvoda dobijamo

$$P\{X = k\} = \frac{\binom{10}{k} \binom{12}{5-k}}{\binom{22}{5}}, \quad k = 0, 1, \dots, 5. \blacktriangleleft$$

Primjer 9. U Maksvel–Bolcmanovom modelu n čestica se razmješta na n energetskih nivoa. Naći raspodjelu slučajne veličine X koja je jednaka broju čestica na prvom nivou.

► Smještanje čestice na prvi nivo možemo tretirati kao uspjeh, vjerovatnoća uspjeha je $\frac{1}{n}$, a smještanje čestice na neki drugi nivo kao neuspjeh. Svako smještanje čestice na neki od nivoa možemo tretirati kao opit, a smještanja svih n čestica kao seriju od n opita. Sada imamo

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

Preporučujemo čitaocu da izračuna graničnu vrijednost ove vjerovatnoće u slučaju kad $n \rightarrow \infty$. Nakon kraćeg računanja se dobija

$$\binom{n}{k} \frac{1}{n^k} \left(1 - \frac{1}{n}\right)^{n-k} \longrightarrow \frac{e^{-1}}{k!}, \quad k = 0, 1, 2, \dots$$

Broj $\frac{e^{-1}}{k!}$ možemo tretirati kao vjerovatnoću da se u čisto teorijskom modelu (fiktivnom) sa prebrojivo mnogo čestica i nivoa, na prvom nivou nalazi k čestica. ◀

Primjer 10. Za slučajnu veličinu X sa raspodjelom

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

kažemo da ima **Puasonovu** raspodjelu sa parametrom λ i koristimo zapis $X : \mathcal{P}(\lambda)$. Puasonova raspodjela je rasprostranjena u praksi. Suptilna analiza pokazuje da je po Puasonovom zakonu raspodijeljena slučajna veličina koja prebrojava automobile koji u toku nekog vremenskog intervala prođu kroz neku raskrnicu. Takođe, po Puasonovom zakonu je raspodijeljena slučajna veličina koja prebrojava kosmičke čestice koje padnu na neki lokalitet u toku, recimo jednog sata. ◀

Primjer 11. Izvodi se serija opita sve dok se ne ostvari prvi uspjeh. Vjerovatnoća uspjeha je p . Naći raspodjelu slučajne veličine X koja je jednaka broju izvedenih opita.

► Zbog nezavisnosti opita imamo

$$P\{X = n\} = q^{n-1}p, \quad n = 1, 2, 3, \dots$$

Za slučajnu veličinu X kažemo da ima **geometrijsku** raspodjelu i koristimo zapis $X : \mathcal{G}(p)$. Primjera radi, neka je slučajna veličina X , u opitu u kome se baca kocka, jednaka broju bacanja do prve pojave (do prvog padanja) šestice. Imamo $X : \mathcal{G}(\frac{1}{6})$. ◀

DEFINICIJA 1.8.3 Funkcija raspodjele vjerovatnoća slučajne veličine X (ili kratko **funkcija raspodjele slučajne veličine**) 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 12. Funkcija raspodjele slučajne veličine X zadate u primjeru 6 data je sa

$$F(x) = \begin{cases} 0, & x \leq 2 \\ 1/6, & 2 < x \leq 3 \\ 5/6, & 3 < x \leq 4 \\ 1, & x > 4 \end{cases}.$$

Slika 6: Primjer 12.

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

Navešćemo svojstva funkcije raspodjele.

a) $F(x)$ je monotonno rastuća funkcija. Dokažimo. Neka su x_1 i x_2 proizvoljni realni brojevi takvi da je $x_1 < x_2$. Imamo

$$\{X < x_1\} \subseteq \{X < x_2\} \implies F(x_1) = P\{X < x_1\} \leq P\{X < x_2\} = F(x_2).$$

$$\text{b) } F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0, \quad F(\infty) = \lim_{x \rightarrow \infty} F(x) = 1.$$

Zbog monotonosti funkcije $F(x)$ dovoljno je dokazati da je $\lim_{n \rightarrow \infty} F(-n) = 0$ i $\lim_{n \rightarrow \infty} F(n) = 1$. Događaji $A_n = \{X < -n\}$, $n \in N$, čine monotonno opadajući niz pa iz neprekidnosti vjerovatnoće u odnosu na opadajući niz događaja slijedi

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} F(-n) = F(-\infty).$$

Međutim, $\bigcap_{n=1}^{\infty} A_n = \emptyset$ jer ne postoji ishod ω takav da je $X(\omega)$ manji od svih negativnih cijelih brojeva. Kako je $P(\emptyset) = 0 \implies F(-\infty) = 0$.

Slično se dokazuje druga jednakost.

c) Za $\forall x \in \mathbb{R}$ važi $F(x-0) = F(x)$ tj. funkcija $F(x)$ je neprekidna sa lijeve strane u svim tačkama. Dokažimo ovo tvrđenje.

Neka je x proizvoljni realni broj i neka je x_n proizvoljni niz takav da $x_n \uparrow x$. Očigledno, za niz događaja $A_n = \{X < x_n\}$ važi $A_n \uparrow$ i $\{X < x\} = \bigcup_{n=1}^{\infty} A_n$. Na osnovu neprekidnosti vjerovatnoće u odnosu na monotono rastući niz događaja, imamo

$$F(x) = P\left\{\bigcup_{n=1}^{\infty} A_n\right\} = \lim_{n \rightarrow \infty} P\{X < x_n\} = \lim_{n \rightarrow \infty} F(x_n) = F(x-0).$$

Formulisaćemo teoremu koja daje karakterizaciju klase funkcija raspodjele. Dokaz izostavljamo zbog toga što se bazira na aparatu koji nismo upoznali u kursevima Matematike.

Teorema 1.8.3 *Neka je $F(x)$:*

1⁰ *monotono rastuća funkcija,*

2⁰ *neprekidna sa lijeve strane,*

3⁰ $F(-\infty) = 0, F(\infty) = 1$.

Tada postoji vjerovatnosni prostor $(\Omega, \mathcal{F}, P)$ i slučajna veličina X na njemu takva da je $F_X(x) = F(x)$.

DEFINICIJA 1.8.4 *Za slučajnu veličinu X , zadatu na vjerovatnosnom prostoru $(\Omega, \mathcal{F}, P)$, kažemo je apsolutno neprekidna ako postoji nenegativna funkcija $\varphi(t)$ takva da je*

$$F_X(x) = \int_{-\infty}^x \varphi(t) dt \text{ za } \forall x \in \mathbb{R}.$$

Funkcija $\varphi(t)$ se naziva **gustina slučajne veličine X** ili prosto **gustina od X** . Jasno, $\int_{-\infty}^{\infty} \varphi(t) dt = F(\infty) = 1$. U kursevima Matematičke analize se dokazuje da u svim tačkama neprekidnosti funkcije $\varphi(x)$ važi $F'(x) = \varphi(x)$.

Neka je $[a, b)$ proizvoljni segment na realnoj pravoj i neka je X slučajna veličina sa funkcijom raspodjele $F(x)$. Pošto je

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

dobijamo

$$P\{X \in [a, b)\} = P\{X < b\} - P\{X < a\} = F(b) - F(a). \quad (2)$$

Jednakost (2) se u slučaju kada je veličina X apsolutno neprekidna svodi na

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

Takođe, pokazuje se da za proizvoljni Borelov skup B važi

$$P\{X \in B\} = \int_B \varphi(t) dt.$$

Znači, raspodjela apsolutno neprekidne slučajne veličine je određena njenom gustinom (ili zbog već pomenute veze funkcije raspodjele i gustine, funkcijom raspodjele te slučajne veličine). Raspodjela slučajne veličine jednoznačno određuje funkciju raspodjele. Važi i obrnuto, tj. funkcija raspodjele jednoznačno određuje raspodjelu vjerovatnoća slučajne veličine. Zbog toga se, u opštem slučaju, smatra da je raspodjela zadata ako je zadata funkcija raspodjele.

Slika 7: Gustina raspodjele.

Primjenom teoreme o srednjoj vrijednosti integrala dobijamo aproksimativnu jednakost

$$P\{t_0 < X < t_0 + \Delta t\} \approx \varphi(t_0) \Delta t,$$

gdje je Δt dovoljno mali broj. Dakle, gustina u tački t_0 simulira vjerovatnoću "upadanja" slučajne veličine u mali interval sa lijevim krajem u tački t_0 . Određenije, tamo gdje je vrijednost $\varphi(t_0)$ veća, vjerovatnoća "upadanja" u interval je veća. Zbog toga, gustinu možemo prihvatiti kao neku vrstu uopštenja vjerovatnoće.

Primjer 13. Slučajno se bira broj iz intervala (a, b) . Jasno, $\Omega = \{\omega : \omega \in (a, b)\} = (a, b)$. Neka je $X : \Omega \rightarrow R$, $X(\omega) = \omega$. U slobodnijoj interpretaciji

možemo reći da slučajna veličina X opisuje opit slučajnog izbora tačke iz intervala (a, b) . Imamo,

$$F(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x \leq b \\ 1, & x \geq b \end{cases} \quad \text{i} \quad \varphi(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b) \\ 0, & x \notin (a, b) \end{cases}.$$

Za slučajnu veličinu X se kaže da ima **ravnomjernu** (**uniformnu**) raspodjelu na intervalu (a, b) i koristi se notacija $X : \mathcal{U}(a, b)$. ◀

Slika 8: Primjer 14.

Primjer 14. Pretpostavimo da se u tački (m, a) nalazi izvor koji zrači α čestice. Razmotrimo čestice koje sijeku x osu. Označimo sa φ ugao koji takve čestice grade sa y osom. Pretpostavimo da je $\varphi : \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$ i označimo sa X tačku presjeka α čestice i x ose.

Za funkciju raspodjele slučajne veličine X dobijamo

$$F_X(x) = P\{X < x\} = P\left\{\varphi < \arctg \frac{x-m}{a}\right\} = \frac{1}{2} + \frac{1}{\pi} \cdot \arctg \frac{x-m}{a}.$$

Odgovarajuća gustina je

$$\varphi(x) = \frac{1}{\pi} \cdot \frac{a}{(x-m)^2 + a^2}.$$

Za slučajnu veličinu X kažemo da ima **Košijevu** raspodjelu i koristimo zapis $X : \mathcal{C}(m, a)$. ◀

Primjer 15. Slučajna veličina X ima **eksponencijalnu** raspodjelu sa parametrom λ , ako je njena gustina $\varphi(x) = \lambda e^{-\lambda x}$, $x > 0$, $\lambda > 0$. Koristimo notaciju $X : \mathcal{E}(\lambda)$. ◀

Primjer 16. Slučajna veličina X ima **normalnu** raspodjelu sa parametrima m i σ^2 , ako je njena gustina

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

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

Diskretne i apsolutno neprekidne slučajne veličine ne iscrpljuju klasu slučajnih veličina. Ove dvije vrste slučajnih veličina se najčešće susreću u praksi i u našem tekstu ćemo raditi isključivo sa njima.

Zadaci

1. Neka je X slučajna veličina i neka je $F(x)$ odgovarajuća funkcija raspodjele. Dokazati da za proizvoljni realni broj t_0 važi $P\{X = t_0\} = F(t_0 + 0) - F(t_0)$.

2. Baca se kocka do prve pojave šestice, ali najviše šest puta. Naći raspodjelu vjerovatnoća slučajne veličine X koja je jednaka broju bacanja.

3. Kocka čije su strane obojene, dijeli se na n^3 , $n \geq 2$, jednakih malih kocki. Neka je X broj obojenih strana slučajno izabrane male kocke. Naći raspodjelu vjerovatnoća slučajne veličine X .

4. (**Banahov zadatak.**) Osoba A kupuje dvije kutije šibica (u svakoj se nalazi po 50 šibica) i jednu kutiju stavlja u lijevi, a jednu u desni džep. Kada joj zatreba šibica, vadi jednu kutiju, uzme šibicu i kutiju vrati u isti džep. Naći raspodjelu vjerovatnoća slučajne veličine X koja je jednaka broju šibica u drugoj kutiji u trenutku kada osoba A prvi put izvadi praznu kutiju.

5. Baca se kocka i kada padne šestica vi gubite jedan dinar, a kada padne broj različit od 6 vi dobijate jedan dinar. Naći raspodjelu slučajne veličine X koja predstavlja vaš "dobitak" nakon n bacanja (jasno, "dobitak" može biti i negativan).

6. Na slučajan način 10 osoba ulazi u 4 vagona. Naći raspodjelu slučajne veličine X koja je jednaka broju zauzetih vagona.

1.9 Slučajni vektori

U praksi su česte situacije kada nekom elementarnom ishodu treba dodijeliti nekoliko brojnih karakteristika. Ova okolnost glavni je motiv za uvođenje pojma slučajnog vektora.

DEFINICIJA 1.9.1 *Neka je $(\Omega, \mathcal{F}, P)$ vjerovatnosni prostor i neka je $X : \Omega \rightarrow R^n$. Kažemo da je X **n-dimenzionalna slučajna veličina** (ili kraće, **slučajni vektor**) (na Ω) ako je $X^{-1}(B) \in \mathcal{F}$ za svako $B \in \mathcal{B}^n$.*

Važi teorema analogna teoremi ??.

Teorema 1.9.1 *Neka je $(\Omega, \mathcal{F}, P)$ vjerovatnosni prostor i $X : \Omega \longrightarrow R^n$. Svaki od navedenih uslova potreban je i dovoljan da bi preslikavanje X bilo slučajni vektor.*

- a) $X^{-1}(G) \in \mathcal{F}$ za svaki otvoreni skup $G \subset R^n$.
- b) $X^{-1}((-\infty, x)) = \{X < x\} \in \mathcal{F}$ za svako $x \in R^n$.
- c) $X^{-1}((-\infty, x]) = \{X \leq x\} \in \mathcal{F}$ za svako $x \in R^n$.
- d) $X^{-1}([x, \infty)) = \{X \geq x\} \in \mathcal{F}$ za svako $x \in R^n$.
- e) $X^{-1}((x, \infty)) = \{X > x\} \in \mathcal{F}$ za svako $x \in R^n$.

Neka je $X : \Omega \longrightarrow R^n$ i $\pi_k : R^n \longrightarrow R$, $k = 1, 2, \dots, n$, k -ta projekcija, $\pi_k(x) = \pi_k(x_1, \dots, x_n) = x_k$. Stavimo

$$X_k = \pi_k(X) : \Omega \longrightarrow R.$$

Tada je $X = (X_1, \dots, X_n)$ tj. $X(\omega) = (X_1(\omega), \dots, X_n(\omega))$, $\omega \in \Omega$.

Teorema 1.9.2 *Neka je $X : \Omega \longrightarrow R^n$, $X = (X_1, \dots, X_n)$. Preslikavanje X je slučajni vektor akko je X_k slučajna veličina za svako $k = 1, 2, \dots, n$.*

Dokaz. ($\Leftarrow$) Neka je za svako k , X_k slučajna veličina. Tada za $\forall x = (x_1, \dots, x_n) \in R^n$ vrijedi

$$X^{-1}((-\infty, x)) = \bigcap_{k=1}^n X_k^{-1}((-\infty, x_k)) \in \mathcal{F}.$$

($\Rightarrow$) Neka je X slučajni vektor. Za proizvoljne $k = 1, 2, \dots, n$ i $x_k \in R$ vrijedi

$$X_k^{-1}((-\infty, x_k)) = X^{-1}((-\infty, b)) \in \mathcal{F},$$

gdje je $b = (\infty, \dots, x_k, \dots, \infty)$. ♦

Neka je X slučajni vektor koji je zadat na vjerovatnosnom prostoru $(\Omega, \mathcal{F}, P)$ i neka je $P_X : \mathcal{B}^n \longrightarrow [0, 1]$ funkcija definisana sa

$$P_X(B) = P\{\omega : X(\omega) \in B\} = P(X^{-1}(B)). \quad (1)$$

Identično kao i u jednodimenzionalnom slučaju, pokazuje se da je $P_X(\cdot)$ vjerovatnoća na mjerljivom prostoru $(R^n, \mathcal{B}^n)$.

DEFINICIJA 1.9.2 *Funkcija $P_X : \mathcal{B}^n \longrightarrow [0, 1]$ koja je definisana pomoću jednakosti (??) zove se raspodjela vjerovatnoća slučajnog vektora X ili kraće raspodjela slučajnog vektora X . Vjerovatnosni prostor $(R^n, \mathcal{B}^n, P_X)$ zove se fazni prostor slučajnog vektora X .*

DEFINICIJA 1.9.3 **Funkcija raspodjele slučajnog vektora** X je funkcija $F_X = F : R^n \rightarrow [0, 1]$ zadana sa

$$F_X(x) = F(x_1, \dots, x_n) = P\{\omega : X(\omega) < x\} = P\{X_1 < x_1, \dots, X_n < x_n\},$$

gdje je $x = (x_1, \dots, x_n) \in R^n$.

Pojmovi diskretnosti i apsolutne neprekidnosti slučajne veličine se prenose u teoriju slučajnih vektora pa imamo diskretne slučajne vektore i apsolutno neprekidne slučajne vektore.

Raspodjela diskretnog slučajnog vektora $X = (X_1, \dots, X_n)$ zadaje se vjerovatnoćama

$$\begin{aligned} p(x_1, \dots, x_n) &= P\{(X_1, \dots, X_n) = (x_1, \dots, x_n)\} = \\ &= P\{X_1 = x_1, \dots, X_n = x_n\}, \end{aligned}$$

gdje je $(x_1, \dots, x_n)$ vektor iz (najviše prebrojivog) skupa mogućih vrijednosti slučajnog vektora. Za dvodimenzionalne slučajne vektore postoji prosta tabelarna prezentacija raspodjele vjerovatnoća (**primjer 1**).

Slučajni vektor $X = (X_1, \dots, X_n)$ je apsolutno neprekidan ako postoji nenegativna funkcija $\varphi(t_1, \dots, t_n)$, naziva se **gustina**, takva da za funkciju raspodjele vektora X i svaku n -torku $(x_1, \dots, x_n) \in R^n$ važi

$$F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} \varphi(t_1, \dots, t_n) dt_1 \dots dt_n.$$

U svim tačkama neprekidnosti gustine $\varphi(t_1, \dots, t_n)$ važi

$$\frac{\partial^n F(t_1, \dots, t_n)}{\partial t_1 \dots \partial t_n} = \varphi(t_1, \dots, t_n).$$

Takođe, za svaki skup $B \in \mathcal{B}^n$ važi

$$P\{X \in B\} = \int \dots \int_B \varphi(t_1, \dots, t_n) dt_1 \dots dt_n.$$

Oдавde zaključujemo da je raspodjela apsolutno neprekidnog slučajnog vektora određena njegovom gustinom.

Kako je svaka komponenta slučajnog vektora slučajna veličina, otvoreno je pitanje kako izračunati odgovarajuću raspodjelu vjerovatnoća. Raspodjela vjerovatnoća komponente se naziva **marginalna raspodjela**. Objasnićemo

kako se računa marginalna raspodjela u slučaju dvodimenzionalnog slučajnog vektora.

Neka je (X, Y) diskretni slučajni vektor sa raspodjelom $p(x_i, y_j)$, $i, j = 1, 2, \dots$. Potražimo raspodjelu komponente X . Imamo

$$\begin{aligned} p(x_i) &= P\{X = x_i\} = P\{\{X = x_i\} \cap \Omega\} \\ &= P\{X = x_i \cap \sum_j \{Y = y_j\}\} = \sum_j p(x_i, y_j). \end{aligned} \quad (2)$$

Analogno, $q(y_j) = P\{Y = y_j\} = \sum_i p(x_i, y_j)$.

Neka je (X, Y) apsolutno neprekidni slučajni vektor sa gustinom $\varphi(u, v)$. Tada je,

$$F(x) = P\{X < x\} = P\{X < x, Y \in R\} = \int_{-\infty}^x \int_{-\infty}^{\infty} \varphi(u, v) du dv$$

te je X apsolutno neprekidna slučajna veličina sa gustinom

$$f(x) = \int_{-\infty}^{\infty} \varphi(x, v) dv.$$

Po analogiji, komponenta Y ima gustinu $g(y) = \int_{-\infty}^{\infty} \varphi(u, y) du$.

DEFINICIJA 1.9.4 *Slučajne veličine $X_1, X_2, \dots, X_n$ su nezavisne ako su za svaku kolekciju skupova $S_1 \in \mathcal{B}, \dots, S_n \in \mathcal{B}$ u potpunosti nezavisni događaji $X_1^{-1}(S_1), \dots, X_n^{-1}(S_n)$.*

Zbog ozbiljnosti aparata koji se u dokazu koristi, sljedeću teoremu nećemo dokazivati.

Teorema 1.9.3 *a) Neka su $F_1, \dots, F_n$ redom, funkcije raspodjela slučajnih veličina $X_1, \dots, X_n$ i neka je $F : R^n \rightarrow [0, 1]$ funkcija raspodjele slučajnog vektora $(X_1, \dots, X_n)$. Slučajne veličine $X_1, \dots, X_n$ su nezavisne akko za svaku n -torku $(x_1, \dots, x_n) \in R^n$ važi*

$$F(x_1, \dots, x_n) = F_1(x_1) \cdots F_n(x_n).$$

b) Diskretne slučajne veličine $X_1, \dots, X_n$ su nezavisne akko za svaku n -torku $(x_1, \dots, x_n) \in R^n$ važi

$$P\left(\bigcap_{k=1}^n \{X_k = x_k\}\right) = \prod_{k=1}^n P\{X_k = x_k\}.$$

c) Neka su $f_1, \dots, f_n$ gustine raspodjela slučajnih veličina $X_1, \dots, X_n$ i neka je $f : R^n \rightarrow R^+$ gustina raspodjele slučajnog vektora $(X_1, \dots, X_n)$. Slučajne veličine $X_1, \dots, X_n$ su nezavisne akko za svaku n -torku $(x_1, \dots, x_n) \in R^n$ važi jednakost

$$f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n).$$

Primjer 1. 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 veličine?

►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,

$$\Omega = \{(1, 1, 1), (1, 1, 2), (1, 2, 1), \dots, (3, 3, 3)\}, \bar{\Omega} = 27.$$

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

$Y \backslash X$	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 veličina X uzima odgovarajuće vrijednosti se dobijaju sumiranjem elemenata iz kolone u kojoj se nalazi odgovarajuća vrijednost (ovo je posljedica primjene formule (??)). Kad se radi sa slučajnom veličinom Y , ulogu kolone preuzima vrsta.

Pokažimo da su komponente X i Y zavisne slučajne veličine. 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 2. Slučajno se bira tačka iz oblasti G , $G \subset \mathbb{R}^2$. Neka je X apscisa, a Y ordinata slučajno izabrane tačke. Dobijamo

$$P\{(X, Y) \in g\} = \frac{\text{mes}(g \cap G)}{\text{mes } G} = \int \int_g \varphi(u, v) du dv,$$

gdje je g Borelov skup, a

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

Znači, (X, Y) je apsolutno neprekidni vektor sa gustinom $\varphi(u, v)$. Koristi se zapis $(X, Y) : \mathcal{U}(G)$ i kaže da je vektor (X, Y) uniformno (ravnomjerno) raspodijeljen u oblasti G . ◀

Primjer 3. Slučajne veličine X i Y su nezavisne, $X : \mathcal{P}(\lambda)$, $Y : \mathcal{P}(\mu)$. Naći raspodjelu slučajne veličine $Z = X + Y$.

► Zbog nezavisnosti komponenti X i Y , nakon primjene formule potpune vjerovatnoće dobijamo

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

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

Primjer 4. Slučajni vektor (X, Y) ima gustinu

$$\varphi(u, v) = \begin{cases} 1, & 0 < v < 2u, 0 < u < 1 \\ 0, & \text{inače} \end{cases}.$$

Naći gustine komponenti X i Y . Da li su X i Y nezavisne slučajne veličine?

Slika 9: Primjer 4.

$$\begin{aligned}
\blacktriangleright f(x) &= \int_0^{2x} \varphi(u, v) dv = \int_0^{2x} 1 dv = 2x, \quad 0 < x < 1, \\
f(x) &= \int_{-\infty}^{\infty} 0 dv = 0, \quad x \notin (0, 1), \\
g(y) &= \int_{\frac{y}{2}}^1 \varphi(u, y) du = \int_{\frac{y}{2}}^1 1 du = 1 - \frac{y}{2}, \quad 0 < y < 2, \\
g(y) &= 0, \quad y \notin (0, 2).
\end{aligned}$$

Kako je, recimo, $\varphi(\frac{1}{2}, \frac{1}{2}) = 1 \neq f(\frac{1}{2})g(\frac{1}{2}) = 1 \cdot \frac{3}{4} = \frac{3}{4}$, zaključujemo da su komponente X i Y zavisne. $\blacktriangleleft$

Primjer 5. Slučajno se bira tačka T na hipotenuzi BC jednakokrakog pravouglog trougla ABC dužine katete 1. Naći raspodjelu slučajne veličine $Y = |AT|$ ($| \cdot |$ je oznaka za dužinu duži).

$\blacktriangleright$ Neka je X apscisa tačke T . Jasno, $X : \mathcal{U}(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$. Iz pravouglog trougla OAT dobijamo $Y = |AT| = \sqrt{\frac{1}{2} + X^2}$, pa je funkcija raspodjele $F(y)$ slučajne promjenljive Y ,

$$\begin{aligned}
F(y) &= P\{Y < y\} = P\{X^2 < y^2 - \frac{1}{2}\} = P\left\{|X| < \sqrt{y^2 - \frac{1}{2}}\right\} \\
&= P\left\{-\sqrt{y^2 - \frac{1}{2}} < X < \sqrt{y^2 - \frac{1}{2}}\right\} = \sqrt{2}\sqrt{y^2 - \frac{1}{2}}, \quad \frac{\sqrt{2}}{2} < y < 1, \\
F(y) &= 0, \quad y < \frac{\sqrt{2}}{2}, \quad F(y) = 1, \quad y > 1, \quad \text{dok je gustina } g(y) = F'(y). \quad \blacktriangleleft
\end{aligned}$$

Slika 10: Primjer 5.

Primjer 6. Na stranici AB , kvadrata $ABCD$ dužine stranice 1, slučajno se bira tačka E , a na stranici CD slučajno se bira tačka F . Naći raspodjelu slučajne veličine $Z = |EF|$.

Slika 11: Primjer 6.

► $X = |AE|$, $Y = |DF| \implies Z = \sqrt{1 + (Y - X)^2}$. Slučajne veličine X i Y su nezavisne i svaka ima $\mathcal{U}(0, 1)$ raspodjelu. Da bismo dobili gustinu vektora (X, Y) treba pomnožiti gustine komponenti X i Y . Dobijamo

$$\varphi(u, v) = \begin{cases} 1, & (u, v) \in (0, 1) \times (0, 1) \\ 0, & \text{inače} \end{cases} \quad . \text{ Dakle, } (X, Y) : \mathcal{U}((0, 1) \times (0, 1)).$$

$$\begin{aligned} F(z) &= P\{Z < z\} = P\left\{X - \sqrt{z^2 - 1} < Y < X + \sqrt{z^2 - 1}\right\} \\ &= P\{(X, Y) \in G\} = 1 - z^2 + 2\sqrt{z^2 - 1}, \quad 1 < z < \sqrt{2}, \end{aligned}$$

G je osjenčeni dio na slici. Takođe, $F(z) = 0, z < 1$, $F(z) = 1, z > \sqrt{2}$ i gustina $g(z) = F'(z)$. ◀

Borelova funkcija slika slučajni vektor u slučajnu veličinu. Preciznije, važi

Teorema 1.9.4 *Neka je $X = (X_1, \dots, X_n)$ slučajni vektor koji je zadat na vjerovatnosnom prostoru $(\Omega, \mathcal{F}, P)$ i neka je $f : R^n \longrightarrow R$ Borelova funkcija. Tada je $Y : \Omega \longrightarrow R$, $Y(\omega) = f(X_1(\omega), \dots, X_n(\omega))$ slučajna veličina.*

Dokaz. Neka je $S \in \mathcal{B}$. Zbog borelovosti funkcije f imamo

$$Y^{-1}(S) = (X_1, \dots, X_n)^{-1}(f^{-1}(S)) \in \mathcal{F}. \blacklozenge$$

Upravo dokazana teorema se često prećutno primjenjuje. Standardne funkcije su Borelove i one, na osnovu teoreme (??), slikaju inicijalne slučajne veličine ili inicijalne slučajne vektore u slučajne veličine.

Primjer 7. X i Y su nezavisne slučajne veličine i svaka ima $\mathcal{U}(0, 1)$ raspodjelu. Naći raspodjelu slučajne veličine $Z = X + Y$.

Slika 12: Primjer 7.

► Vektor (X, Y) ima $\mathcal{U}((0, 1) \times (0, 1))$ raspodjelu (objašnjenje je dato u primjeru 6).

$$F(z) = P\{Z < z\} = P\{(X, Y) \in V_1\} = \frac{z^2}{2}, \quad 0 < z < 1,$$

$$F(z) = P\{(X, Y) \in V_2\} = z - \frac{z^2}{4}, \quad 1 < z < 2, \text{ pa je}$$

$$\varphi(z) = \begin{cases} 0, & z \notin (0, 2) \\ z, & z \in (0, 1) \\ 1 - \frac{z}{2}, & z \in (1, 2) \end{cases}.$$

Primjer 8. Slučajne veličine $X_1, \dots, X_n$ su nezavisne i svaka ima $\mathcal{U}(0, 1)$ raspodjelu. Naći raspodjele slučajnih veličina $V = \max_{1 \leq k \leq n} X_k$ i $W = \min_{1 \leq k \leq n} X_k$.

$$\begin{aligned} \text{► } F(v) &= P\left\{\max_{1 \leq k \leq n} X_k < v\right\} = P\{X_1 < v, \dots, X_n < v\} \\ &= P\{X_1 < v\} \cdots P\{X_n < v\} = v^n, \quad 0 < v < 1, \\ F(v) &= 0, \quad v < 0, \quad F(v) = 1, \quad v > 1 \text{ i gustina je} \\ f(v) &= \begin{cases} 0, & v \notin (0, 1) \\ nv^{n-1}, & v \in (0, 1) \end{cases}. \end{aligned}$$

Iskoristili smo nezavisnost komponenti i činjenicu da je maksimalna komponenta manja od v akko je svaka komponenta manja od v . Potražimo raspodjelu slučajne veličine W .

$$\begin{aligned} G(v) &= P\left\{\min_{1 \leq k \leq n} X_k < v\right\} = 1 - P\left\{\min_{1 \leq k \leq n} X_k \geq v\right\} \\ &= 1 - P\{X_1 \geq v, \dots, X_n \geq v\} = 1 - P\{X_1 \geq v\} \cdots P\{X_n \geq v\} \\ &= 1 - (1 - v)^n, \quad v \in (0, 1), \quad G(v) = 0, \quad v < 0, \quad G(v) = 1, \quad v > 1, \\ g(v) &= G'(v) = \begin{cases} n(1 - v)^{n-1}, & v \in (0, 1) \\ 0, & \text{inače} \end{cases}. \end{aligned}$$

Koristili smo nezavisnost komponenti i činjenicu da je minimalna komponenta veća od v akko su sve komponente veće od v .

Ovaj model ima prostu i zanimljivu geometrijsku interpretaciju. Ako n puta slučajno biramo tačku iz intervala $(0, 1)$, slučajne veličine $X_1, \dots, X_n$ predstavljaju rastojanja odabranih tačaka od 0. Slučajna veličina V predstavlja rastojanje od 0 do tačke koja je najudaljenija od 0, a slučajna veličina W predstavlja rastojanje od 0 do tačke koja je najbliža tački 0. ◀

Zadaci

1. Slučajno se bira polje na šahovskoj tabli. Neka je X broj susjednih bijelih polja, a Y broj susjednih crnih polja. Naći raspodjelu vektora (X, Y) . Ispitati nezavisnost komponenti.

2. Bacaju se dvije kocke i registruju brojevi na njima. Neka je X broj na prvoj kocki, a Y veći broj. Naći raspodjelu vektora (X, Y) . Ispitati nezavisnost komponenti.

3. U kutiji se nalaze četiri kuglice koje su numerisane brojevima 1, 2, 3, 4. Iz kutije se po modelu bez vraćanja vade kuglice sve dok se ne izvuče kuglica sa parnim brojem. Neka je X zbir izvučenih brojeva, a Y broj izvlačenja. Naći raspodjelu vektora (X, Y) . Ispitati nezavisnost komponenti.

4. Iz kompleta od 32 karte slučajno se a) po modelu bez vraćanja, b) po modelu sa vraćanjem, vade dvije karte. Neka je X broj izvučenih dama, a Y broj izvučenih krsta. Naći raspodjelu vektora (X, Y) . Ispitati nezavisnost komponenti.

5. Na stranama pravilnog tetraedra su zapisani redom brojevi 1, 2, 3 i 4. Tetraedar se baca dva puta i registruju se brojevi na stranama na koje tetraedar padne. Neka je X broj koji se registruje kod prvog bacanja, a Y broj koji se registruje kod drugog bacanja. Neka je $Z = \min\{X, Y\}$, $V =$

$\max\{X, Y\}$. Naći raspodjelu vektora (Z, V) . Da li su Z i V nezavisne slučajne veličine?

6. Slučajna veličina X ima $\mathcal{E}(1)$ raspodjelu. Naći raspodjele slučajnih veličina $[X]$ (cio dio od X) i $\{X\}$ (razlomljeni dio od X).

7. Neka je X slučajna veličina sa neprekidnom funkcijom raspodjele F_X . Naći funkciju raspodjele slučajne veličine $Y = F_X(X)$.

8. X i Y su nezavisne slučajne veličine i svaka ima $\mathcal{G}(p)$ raspodjelu. Naći $P\{X = Y\}$. Naći raspodjele slučajnih veličina $Z = \min(X, Y)$ i $W = \max(X, Y)$.

9. Poluprečnik kruga je slučajna veličina X koja ima $\mathcal{U}(2, 3)$ raspodjelu. Naći raspodjelu površine kruga (tj. naći funkciju raspodjele ili gustinu).

10. Iz intervala $(0, 1)$ slučajno se biraju dvije tačke. Naći raspodjelu rastojanja između tih tačaka.

11. Slučajni vektor (X, Y) ima $\mathcal{U}((0, 1) \times (0, 1))$ raspodjelu. Naći raspodjelu slučajne veličine $Z = XY$.

12. Slučajni vektor (X, Y) ima $\mathcal{U}(K)$ raspodjelu gdje je $K = \{(x, y) : |x| + |y| < 1\}$. Naći raspodjele komponenti X i Y . Da li su X i Y nezavisne slučajne veličine?

13. Tačka T se slučajno bira na hipotenuzi BC pravouglog trougla ABC , $|AB| = 1$, $|AC| = 2$. Naći raspodjelu slučajne veličine $W = |AT|^2$.

14. Slučajne veličine X i Y su nezavisne i imaju redom $\mathcal{U}(0, 1)$ i $\mathcal{U}(0, 2)$ raspodjelu. Naći raspodjelu slučajne veličine $Z = |X - Y|$.

15. Novčić se baca pet puta. Neka je X broj palih grbova, a Y najveći broj uzastopnih pojavljivanja grbova. Odrediti raspodjelu vektora (X, Y) . Da li su X i Y nezavisne slučajne veličine?

1.10 Matematičko očekivanje

Kompletna vjerovatnosna informacija o nekoj slučajnoj veličini je sadržana u njenoj funkciji raspodjele. Kada znamo funkciju raspodjele, mi možemo izračunati vjerovatnoću da slučajna veličina "upadne" u neki interval ili neki drugi Borelov skup sa realne prave. Prirodno, možemo da izračunamo i vjerovatnoću sa kojom slučajna veličina uzima neku konkretnu brojnu vrijednost. U slobodnijoj interpretaciji, kad imamo funkciju raspodjele slučajne veličine mi znamo koje vrijednosti uzima ta slučajna veličina i sa kojim vjerovatnoćama.

Često je dobro imati neku sumarnu informaciju o slučajnoj veličini. Među tim informacijama – karakteristikama, najvažnije je matematičko očekivanje.

Navešćemo primjer u kome će biti sprovedena neformalna analiza. Zaključci će se oslanjati na intuiciju.

Primjer 1. Pretpostavimo da smo kocku bacili N puta i neka je jedinica pala N_1 puta, dvojka N_2 puta, ..., šestica N_6 puta. Potražimo srednju vrijednost palih brojeva. Imamo

$$\frac{1N_1 + 2N_2 + \dots + 6N_6}{N}.$$

Kada $N \rightarrow \infty$, tada će, tako nam govori intuicija, $\frac{N_i}{N} \rightarrow \frac{1}{6}$, $i = 1, 2, \dots, 6$ (tj. relativna učestalost svakog od mogućih ishoda teži ka $\frac{1}{6}$). Znači, srednja vrijednost palih brojeva konvergira ka broju $1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = 3,5$. Broj 3,5 predstavlja granicu srednje vrijednosti realizacija našeg opita i dobijen je u rezultatu sumiranja proizvoda mogućih ishoda (možemo govoriti o značenjima slučajne veličine X – broj na kocki) i odgovarajućih vjerovatnoća. Mi ćemo, motivisani načinom na koji je izračunat broj 3,5 u našem primjeru, dati definiciju matematičkog očekivanja. Tačni smisao broju 3,5 i uopšte matematičkom očekivanju će dati Hinčinov zakon velikih brojeva. ◀

DEFINICIJA 1.10.1 *Neka je X prosta slučajna veličina sa raspodjelom $p(x_k)$, $k = 1, 2, \dots, n$. Matematičko očekivanje EX slučajne veličine X definišaćemo pomoću jednakosti*

$$EX = \sum_{k=1}^n x_k p(x_k).$$

DEFINICIJA 1.10.2 *Neka je X diskretna slučajna veličina sa raspodjelom $p(x_k)$, $k = 1, 2, \dots$. Matematičko očekivanje EX slučajne veličine X definišaćemo pomoću jednakosti*

$$EX = \sum_{k=1}^{\infty} x_k p(x_k),$$

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

Da bismo objasnili smisao uslova apsolutne konvergencije reda kojim zadajemo matematičko očekivanje, uradićemo jedan primjer iz Analize.

Primjer 2. Ako u razvoju $\ln(1+x) = \sum_{n=1}^{\infty} \frac{(-x)^{n+1}}{n}$, $-1 < x \leq 1$, stavimo $x = 1$, dobijamo

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

Potražimo

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

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. ◀

Već je rečeno da će tačni smisao matematičkom očekivanju dati Hinčinov zakon velikih brojeva. Tamo će biti pokazano da je matematičko očekivanje granica srednje vrijednosti realizacija slučajne veličine (u kom smislu granica, biće tamo objašnjeno). Kada red $\sum_{k=1}^{\infty} x_k p(x_k)$ uslovno konvergira, tada se članovi mogu preindeksirati tako da novoformirani red ima sumu koja je različita od polazne. Mi indeksiranjem nekoj vrijednosti slučajne veličine dodijelimo indeks 1, nekoj indeks 2 itd., a preindeksiranje znači drugačiju podjelu indeksa. Indeksiranjem smo pripremili članove za sumiranje, a suma treba da predstavlja granicu srednje vrijednosti realizacija. Dakle, kada bi red $\sum_{k=1}^{\infty} x_k p(x_k)$ uslovno konvergirao, granica srednje vrijednosti realizacija (tj. matematičko očekivanje) slučajne veličine bi zavisila od toga kako smo indeksirali njene vrijednosti. Imali bismo patološku situaciju da ista karakteristika (tj. granica srednje vrijednosti realizacija) zavisi od čisto formalnog ređanja vrijednosti slučajne veličine. U slučaju kada red apsolutno konvergira, suma ne zavisi od načina indeksiranja i zahtjevom da red $\sum_{k=1}^{\infty} x_k p(x_k)$ apsolutno konvergira obezbjeđuje se stabilnost sume koja predstavlja granicu srednje vrijednosti realizacija.

Iz definicije (??) je jasno da svaka prosta slučajna veličina ima očekivanje (često se prefiks matematičko izostavlja). Navešćemo primjer diskretne slučajne veličine koja nema očekivanje.

Primjer 3. Slučajna veličina X ima raspodjelu

$$X : \begin{array}{cccccccccccc} \cdot & \cdot & \cdot & -3 & -2 & -1 & 1 & 2 & 3 & \cdot & \cdot & \cdot \\ & & & \frac{3}{3^2\pi^2} & \frac{3}{2^2\pi^2} & \frac{3}{\pi^2} & \frac{3}{\pi^2} & \frac{3}{2^2\pi^2} & \frac{3}{3^2\pi^2} & \cdot & \cdot & \cdot \end{array}$$

Pošto red $\frac{3}{\pi^2}(1 - 1 + \frac{1}{2} - \frac{1}{2} + \frac{1}{3} - \frac{1}{3} + \dots)$ ne konvergira apsolutno (ne konvergira čak ni uslovno), možemo konstatovati da slučajna veličina X nema očekivanje. ◀

Teorema 1.10.1 *Neka su X i Y diskretne slučajne veličine. Tada je $E(X+Y) = EX + EY$ (u teoremama sa očekivanjem, podrazumijeva se da slučajne veličine koje se pominju imaju očekivanje).*

Dokaz. Neka su $a_1, a_2, \dots$ značenja slučajne veličine X , a $p_1, p_2, \dots$ vjerovatnoće sa kojima se ona realizuju i neka su $b_1, b_2, \dots$ značenja slučajne veličine Y , a $q_1, q_2, \dots$ vjerovatnoće sa kojima se ona realizuju. Značenja slučajne veličine $X+Y$ su $a_n + b_k$, $k, n = 1, 2, \dots$. Označimo sa p_{nk} vjerovatnoću sa kojom X uzima vrijednost a_n , a Y vrijednost b_k . Sada je

$$\begin{aligned} E(X+Y) &= \sum_{n,k=1}^{\infty} (a_n + b_k) p_{nk} = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} (a_n + b_k) p_{nk} \\ &= \sum_{n=1}^{\infty} a_n \sum_{k=1}^{\infty} p_{nk} + \sum_{k=1}^{\infty} b_k \sum_{n=1}^{\infty} p_{nk}. \end{aligned}$$

Iz formule potpune vjerovatnoće imamo $\sum_{k=1}^{\infty} p_{nk} = p_n$ i $\sum_{n=1}^{\infty} p_{nk} = q_k$, pa je

$$\sum_{n=1}^{\infty} a_n \sum_{k=1}^{\infty} p_{nk} = \sum_{n=1}^{\infty} a_n p_n = EX, \quad \sum_{k=1}^{\infty} b_k \sum_{n=1}^{\infty} p_{nk} = \sum_{k=1}^{\infty} b_k q_k = EY. \blacklozenge$$

Teorema 1.10.2 *Neka su X i Y nezavisne diskretne slučajne veličine. Tada je $EXY = EXEY$.*

Dokaz Pretpostavke o značenjima slučajnih veličina X i Y kao i o vjerovatnoćama uzimanja tih značenja su iste kao i u dokazu teoreme ???. Zbog nezavisnosti važi $p_{nk} = p_n q_k$. Sada je

$$EXY = \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_k b_n p_k q_n = \sum_{k=1}^{\infty} a_k p_k \sum_{n=1}^{\infty} b_n q_n = EXEY. \blacklozenge$$

Jednostavno se dokazuju sljedeća svojstva matematičkog očekivanja diskretnih slučajnih veličina.

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

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

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

S4) $E(a_1 X_1 + \dots + a_n X_n) = a_1 EX_1 + \dots + a_n EX_n$.

Primjer 4. $EI_A = 0 \cdot q + 1 \cdot p = p$, gdje je $p = P(A)$. ◀

Primjer 5. Neka je $S_n : \mathcal{B}(n, p)$. Tada je

$$\begin{aligned} ES_n &= \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-1-(k-1))!} p^{k-1} q^{n-1-(k-1)} \\ &= np(p+q)^{n-1} = np. \end{aligned}$$

Ako sa A_k označimo događaj da se u k -tom opitu ostvari uspjeh, tada je $S_n = \sum_{k=1}^n I_{A_k}$, pa je $ES_n = \sum_{k=1}^n EI_{A_k} = np$. ◀

Primjer 6. Koliki je očekivani broj bacanja kocke do prve pojave šestice?

►Označimo sa X slučajnu veličinu koja je jednaka broju bacanja kocke do prve pojave šestice. Znamo da je $X : \Gamma(\frac{1}{6})$. Podsjetimo se, $\sum_{n=1}^{\infty} nx^{n-1} =$

$$\frac{1}{(1-x)^2}, \quad -1 < x < 1. \text{ Sada je } EX = \sum_{n=1}^{\infty} \frac{1}{6} \left(\frac{5}{6}\right)^{n-1} n = 6. \quad \blacktriangleleft$$

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

►Neka je X_1 slučajna veličina koja predstavlja prvoizvučeni broj, neka je X_2 slučajna veličina koja predstavlja drugoizvučeni broj itd. Imamo

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

Jednaka raspodijeljenost slučajnih veličina $X_i, i = 1, 2, \dots, 7$, 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 1 odjeljka 1.7. Tražena suma je $W = X_1 + \dots + X_7$, pa je

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

Zahvaljujući aditivnosti očekivanja, očekivanje slučajne veličine W smo našli bez prethodnog traženja raspodjele ove slučajne veličine. Ista ideja se koristi i u sljedećem primjeru. ◀

Primjer 8. Ravan je išrafrana pravima koje obrazuju kvadrate dužine stranice $2a$. Na ravan se baca igla dužine $2l$ ($l < a$). Naći očekivani broj zasječenih pravih.

►Neka je X broj zasječenih horizontalnih pravih, a Y broj zasječenih vertikalnih pravih. Pošto je vjerovatnoća zasijecanja (Bifonov problem igle) $\frac{2l}{a\pi}$, imamo

$$X, Y : \begin{array}{cc} 0 & 1 \\ 1 - \frac{2l}{a\pi} & \frac{2l}{a\pi} \end{array},$$

Slika 13: Primjer 8.

pa je traženo očekivanje $E(X + Y) = EX + EY = \frac{4l}{a\pi}$. ◀

DEFINICIJA 1.10.3 *Neka je X apsolutno neprekidna slučajna veličina sa gustinom $\varphi(x)$. Matematičko očekivanje EX slučajne veličine X definisaćemo pomoću jednakosti*

$$EX = \int_{-\infty}^{\infty} x\varphi(x)dx,$$

ako ovaj integral apsolutno konvergira. Ako integral ne konvergira apsolutno, tada kažemo da slučajna veličina 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. Svojstva očekivanja koja važe u diskretnom slučaju se prenose i na apsolutno neprekidni slučaj.

Sljedeća teorema ima veliki teorijski i praktični značaj.

Teorema 1.10.3 *Neka je X apsolutno neprekidna slučajna veličina sa gustinom $\varphi(x)$ i neka je $f : R \rightarrow R$ Borelova funkcija. Tada je*

$$Ef(X) = \int_{-\infty}^{\infty} f(x)\varphi(x)dx.$$

Dakle, očekivanje slučajne veličine $Y = f(X)$ možemo dobiti bez prethodnog traženja njene funkcije raspodjele. Preporučujemo čitaocu da vodeći računa o analogijama integral – suma i gustina – vjerovatnoća, formuliše prethodnu teoremu i za diskretne slučajne veličine.

Neka je (X, Y) apsolutno neprekidni slučajni vektor sa gustinom $\varphi(x, y)$ i neka je $f : R^2 \rightarrow R$ Borelova funkcija. Iz teoreme ?? slijedi da je $Z =$

$f(X, Y)$ slučajna veličina. Analogno kao u teoremi ??, imamo

$$EZ = \int \int_{R^2} f(x, y) \varphi(x, y) dx dy. \quad (1)$$

Dali smo definicije matematičkog očekivanja za diskretne i apsolutno neprekidne slučajne veličine. U opštem slučaju, očekivanje se identifikuje sa integralom koji se naziva Lebegov integral. Samo pomenimo da Lebegov integral predstavlja uopštenje Rimanovog integrala.

Primjer 9. Slučajna veličina X ima $\mathcal{C}(0, 1)$ raspodjelu. Kako integral $\int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)} dx$ ne konvergira, to možemo konstatovati da slučajna veličina X nema očekivanje. ◀

Primjer 10. $X : \mathcal{U}(a, b)$, $EX = \int_a^b \frac{x}{a+b} dx = \frac{a+b}{2}$. ◀

Primjer 11. $X : \mathcal{E}(\lambda)$, $EX = \int_0^{\infty} \lambda x e^{-\lambda x} dx = \frac{1}{\lambda}$. ◀

Primjer 12. $X : \mathcal{N}(m, \sigma^2)$, $EX = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}} dx =$
 $= \frac{m}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{v^2}{2}} dv = m$, gdje je $v = \frac{x-m}{\sigma}$. Koristili smo poznati rezultat
 $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$. ◀

Primjer 13. Na kružnici poluprečnika 1 slučajno se biraju tačke A i B . Naći EX gdje je $X = |AB|$.

Slika 14: Primjer 13.

▶ $\alpha : \mathcal{U}(0, \pi)$, $X = |AB| = 2 \sin \frac{\alpha}{2}$, $EX = \int_0^{\pi} \frac{2}{\pi} \sin \frac{\alpha}{2} d\alpha = \frac{4}{\pi}$. ◀

Primjer 14. U kvadratu dužine stranice 1 slučajno se bira tačka T . Naći raspodjelu slučajne veličine Z koja je jednaka rastojanju tačke T od najbliže stranice kvadrata. Naći EZ .

► Zbog simetričnosti, ne smanjujući opštost, možemo pretpostaviti da izabrana tačka pripada trouglu ABC . Označimo sa (X, Y) slučajni vektor koji predstavlja koordinate tačke T . Gustina ovog vektora je $\varphi(x, y) = 8$, $(x, y) \in \Delta ABC$. Očigledno $Z = Y$. Potražimo gustinu komponente Y . Imamo

$$\varphi_Y(y) = 8 \int_y^{1/2} dx = 4 - 8y, \quad 0 < y < \frac{1}{2}, \quad \text{pa je } EZ = \int_0^{1/2} y(4 - 8y)dy = \frac{1}{6}.$$

Očekivanje smo mogli dobiti i primjenom formule (??). Naime,

$$EZ = \iint_{\Delta ABC} 8y dx dy = \int_0^{1/2} dx \int_0^x 8y dy = \frac{1}{6}. \blacktriangleleft$$

Slika 15: Primjer 14.

Zadržimo se za trenutak na slučajnim veličinama

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

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

DEFINICIJA 1.10.4 *Disperzija DX slučajne veličine X se zadaje pomoću jednakosti*

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

Jednakost iz definicije disperzije potencira da je ova karakteristika mjera odstupanja slučajne veličine od njenog očekivanja. Neposredno računanje disperzije je jednostavnije sprovesti pomoću jednakosti do koje se elementarno dolazi.

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

Evidentiraćemo svojstva disperzije.

a) $DX \geq 0$ i $D(cX) = c^2 DX$, c je konstanta.

b) Ako su $X_1, \dots, X_n$ nezavisne slučajne veličine, tada je $D\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n DX_i$. Zaista, zbog nezavisnosti imamo

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

Primjer 15. $DI_A = p - p^2 = pq$ jer je $I_A^2 : \begin{smallmatrix} 0 & 1 \\ q & p \end{smallmatrix}$. ◀

Primjer 16. Neka je $S_n : \mathcal{B}(n, p)$. Već je konstatovano $S_n = \sum_{k=1}^n I_{A_k}$, pa je $DS_n = \sum_{k=1}^n DI_{A_k} = npq$. ◀

Primjer 17. Neka je $X : \mathcal{N}(m, \sigma^2)$. Imamo

$$DX = E(X - m)^2 = \int_{-\infty}^{\infty} \frac{(x - m)^2}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \sigma^2. \quad \blacktriangleleft$$

Teorema 1.10.4 (Čebišovljeva nejednakost.) *Ako slučajna veličina X ima EX^2 , tada je za svako $\varepsilon > 0$*

$$P\{|X| \geq \varepsilon\} \leq \frac{EX^2}{\varepsilon^2}.$$

Dokaz. Definišimo slučajnu veličinu

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

Očigledno, $Y \leq |X| \implies Y^2 \leq X^2$, pa je

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

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

Nejednakost se dobija kada se u Čebišovljevoj nejednakosti X zamijeni sa $X - EX$. $\blacklozenge$

Stepen zavisnosti između slučajnih veličina se računa pomoću koeficijenta korelacije.

DEFINICIJA 1.10.5 **Koeficijent korelacije između slučajnih veličina X i Y je**

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

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

Primjer 18. 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

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

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

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

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

Primjer 19. 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.

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

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

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 $\rho_{X,Y} = -\frac{1}{5}$. ◀

Primjer 20. U kutiji se nalazi 10 bijelih i 12 crnih kuglica. Iz kutije se a) po modelu sa vraćanjem, b) po modelu bez vraćanja, vadi 5 kuglica. Neka slučajna veličina X predstavlja broj izvađenih bijelih kuglica. Naći EX i DX .

► a) $X : \mathcal{B}(5, \frac{5}{11})$, $EX = 5 \cdot \frac{5}{11} = \frac{25}{11}$, $DX = 5 \cdot \frac{5}{11} \cdot \frac{6}{11} = \frac{150}{121}$.

b) $X = \sum_{i=1}^5 I_{A_i}$, gdje je A_i događaj da se u i -tom izvlačenju izvuče bijela kuglica. Zbog aditivnosti očekivanja, imamo $EX = 5 \cdot \frac{5}{11} = \frac{25}{11}$. Takođe,

$$EX^2 = \sum_{i=1}^5 E I_{A_i}^2 + \sum_{i \neq j} E I_{A_i} I_{A_j} = 5 \cdot \frac{5}{11} + 20 \cdot \frac{10}{22} \cdot \frac{9}{21},$$

pa nakon kraćeg računanja dobijamo $DX = \frac{850}{847}$. ◀

Zadaci

- Slučajna veličina X ima $\mathcal{P}(\lambda)$ raspodjelu. Naći EX i DX .
- Naći disperziju slučajne veličine a) $X : \Gamma(p)$, b) $X : \mathcal{U}(a, b)$, c) $X : \mathcal{E}(\lambda)$.
- Novčić se baca tri puta. Neka je X broj palih pisama, a Y broj palih grbova. Naći $\rho_{X,Y}$.
- U jednoj igri se dobija poen sa vjerovatnoćom p , a gubi sa vjerovatnoćom q . Neka je S_n broj osvojenih poena nakon n nezavisnih igara. Naći raspodjelu slučajne veličine S_n . Naći ES_n i DS_n .

5. Novčić se baca do prve pojave pisma. Ako je izveden paran broj bacanja, igrač gubi b dinara, a ako je izveden neparan broj bacanja, dobija a dinara. Odrediti a i b tako da igra bude "fer", tj. da očekivani dobitak igrača bude 0.

6. Čovjek u mraku pokušava da otključa vrata. U džepu ima četiri ključa od kojih samo jedan otključava vrata. Naći očekivani broj pokušaja ako se ključevi uzimaju slučajno a) sa vraćanjem u džep, b) bez vraćanja u džep.

7. Jednačina $ax^3 + bx^2 + cx + d = 0$ (*) ima slučajne koeficijente koji nezavisno uzimaju vrijednosti 1 ili -1 sa vjerovatnoćom $\frac{1}{2}$. Naći očekivani broj realnih korijena jednačine (*).

8. Strijelac koji posjeduje n metaka pogađa cilj sa vjerovatnoćom p , a gađanja su nezavisna. U metu gađa sve dok je ne pogodi ili dok ne potroši sve metke. Naći očekivani broj gađanja.

9. Tri kuglice se na slučajan način razmještaju u kutije A, B i C. Neka je X broj kuglica u kutiji A, a Y broj zauzetih kutija. Naći $\rho_{X,Y}$.

10. Slučajno se biraju podskupovi A i B skupa $S = \{1, 2, \dots, n\}$ i to na sljedeći način: svaki element skupa S pripada sa vjerovatnoćom p skupu A i nezavisno od pripadnosti skupu A pripada skupu B sa vjerovatnoćom q , $p + q = 1$. Pripadnosti različitih elemenata skupovima A i B su takođe nezavisne. Odrediti očekivani broj elemenata skupa $A \cap B$.

11. U kutiji se nalaze dva defektna i osam ispravnih proizvoda. Proizvodi se vade jedan po jedan sve dok se ne izvade oba defektna proizvoda. Naći očekivani broj izvlačenja.

12. n kuglica numerisanih brojevima $1, 2, \dots, n$ je poređano u niz. Neka je slučajna veličina X jednaka broju kuglica kod kojih se broj na kuglici poklapa sa rednim brojem kuglice u nizu. Naći EX .

13. Neka je a fiksirani broj sa $(0, 1)$, neka je $X : \mathcal{U}(0, 1)$ i $Y = |X - a|$. Naći $\rho_{X,Y}$ i ono a za koje je $\rho_{X,Y} = 0$.

14. U krugu poluprečnika r se slučajno bira tačka. Naći očekivano rastojanje od izabrane tačke do centra kruga.

15. Štap dužine 1 se lomi na dva dijela koji predstavljaju stranice pravougaonika. Naći raspodjelu slučajne veličine X koja predstavlja površinu pravougaonika. Naći EX .

16. Na intervalu $(0, 1)$ se slučajno bira n tačaka. Naći očekivani položaj tačke koja je a) najbliža 0, b) najbliža 1. Naći očekivano rastojanje između krajnjih tačaka.

17. Tačka se kreće konstantnom brzinom od 1cm/sec . Iz ishodišta krene pravolinijski i nakon jednog sekunda promijeni pravac kretanja za ugao α u odnosu na postojeći pravac. Zatim nastavi sa pravolinijskim kretanjem i nakon jednog sekunda ponovo mijenja pravac za ugao α u odnosu na pos-

tojeći pravac. Zatim se pravolinijski kreće jedan sekund i nakon njegovog isteka kretanje se završava. Ako je $\alpha : \mathcal{U}(0, \pi/2)$, naći očekivano rastojanje završnog položaja tačke od njenog ishodišta.

18. U jednakokraničnom trouglu dužine stranice 1 slučajno se bira tačka T . Naći očekivano rastojanje tačke T od a) najbliže, b) najudaljenije stranice trougla.

19. U kutiji se nalazi k bijelih i j crnih kuglica, a kraj kutije je još dovoljno mnogo bijelih i crnih kuglica. Izvodi se niz sljedećih eksperimenata: U svakom koraku se iz kutije slučajno bira jedna kuglica sa vraćanjem i u kutiju se stavlja još m kuglica iste boje kao što je izabrana kuglica.

a) Odrediti vjerovatnoću p_n da se u n -tom koraku iz kutije izabere crna kuglica.

b) Odrediti matematičko očekivanje broja X_n crnih kuglica u kutiji poslije n -tog koraka.

20. Slučajna veličina X uzima vrijednosti sa intervala (a, b) . Dokazati da je

$$DX \leq \frac{1}{4}(b-a)^2.$$

1.11 Karakteristične funkcije

Do sada izložena teorija pokazuje da se u Vjerovatnoći često koristi aparat Matematičke analize. Ovak odjeljak ćemo posvetiti karakterističnoj funkciji koja je, vidjećemo iz definicije, Furijeova transformacija funkcije raspodjele slučajne veličine. Rješenja mnogih krupnih vjerovatnosnih zadataka su dobijena korišćenjem metoda karakterističnih funkcija. U okvirima ove teorije Analiza i Vjerovatnoća se prožimaju i za neke od rezultata je teško presuditi da li po svom duhu pripadaju Analizi ili Vjerovatnoći.

Naša ideja je da čitaocu upoznamo sa minimumom teorije. Zbog toga što dokazi teorema traže poznavanje ozbiljnog analitičkog aparata, mi ih izostavljamo.

DEFINICIJA 1.11.1 Karakteristična funkcija $f_X(t)$ slučajne veličine X se zadaje sa

$$f_X(t) = Ee^{itX}, \quad t \in R.$$

Neka je X diskretna slučajna veličina sa raspodjelom $p(x_k)$, $k = 1, 2, \dots$. Pokažimo da red $\sum_k e^{itx_k} p(x_k)$ apsolutno konvergira. Kako je $|e^{i\theta}| = 1$ za svako $\theta \in R$, zaključujemo da je $|e^{itx_k} p(x_k)| = p(x_k)$ pa se zadatak ispitivanja apsolutne konvergencije gornjeg reda svodi na zadatak ispitivanja

konvergencije reda $\sum_k p(x_k)$. Ovaj red konvergira i njegova je suma jednaka

1. Na ovaj način je pokazano da postoji matematičko očekivanje kojim je zadana karakteristična funkcija diskretne slučajne veličine.

Ako je X apsolutno neprekidna slučajna veličina sa gustinom $\varphi(x)$, tada na osnovu teoreme ?? dobijamo $f_X(t) = \int_{-\infty}^{\infty} e^{itx} \varphi(x) dx$ (ovaj integral apsolutno konvergira, a dokaz je analogan dokazu apsolutne konvergencije reda iz diskretnog slučaja). U teoriji karakterističnih funkcija se u pisanju obično izostavljaju granice kod integrala, a takođe se, kada je jasno na koju se slučajnu veličinu misli, izostavlja donji indeks kod oznake za karakterističnu funkciju.

Karakterističnu funkciju imaju i slučajne veličine koje nisu ni diskretne ni apsolutno neprekidne.

Navešćemo svojstva karakterističnih funkcija.

a) $f(0) = 1$, $|f(t)| \leq 1$.

Dokazaćemo nejednakost za apsolutno neprekidnu slučajnu veličinu X . Neka je $\varphi(x)$ odgovarajuća gustina. Imamo

$$|f(t)| = \left| \int e^{itx} \varphi(x) dx \right| \leq \int \varphi(x) dx = 1.$$

b) Ako je $Y = aX + b$, a i b su konstante, tada je

$$f_Y(t) = e^{ibt} f_X(at).$$

Zaista,

$$f_Y(t) = E e^{it(aX+b)} = e^{itb} E e^{i(at)X} = e^{itb} f_X(at).$$

c) Ako su X i Y nezavisne slučajne veličine, tada je

$$f_{X+Y}(t) = f_X(t) f_Y(t).$$

Zaista,

$$f_{X+Y}(t) = E e^{i(X+Y)t} = E e^{itX} e^{itY} = E e^{itX} E e^{itY} = f_X(t) f_Y(t).$$

Prirodno, važi i uopštenje. Ako su $X_1, \dots, X_n$ nezavisne slučajne veličine, tada je

$$f_{X_1+\dots+X_n}(t) = f_{X_1}(t) \cdots f_{X_n}(t).$$

d) Ako postoji $E|X|^n$, tada za svako $k \in \{1, 2, \dots, n\}$ važi

$$f^{(k)}(0) = i^k E X^k.$$

Radićemo sa apsolutno neprekidnom slučajnom veličinom X čija je gustina $\varphi(x)$. Primijetimo,

$$|X|^k < |X|^n + 1, \quad k \in \{1, 2, \dots, n\},$$

te iz uslova slijedi da postoji $E|X|^k$. Takođe,

$$\left| \int x^k e^{itx} \varphi(x) dx \right| \leq \int |x|^k \varphi(x) dx = E|X|^k < \infty,$$

tako da funkciju $f(t) = \int e^{itx} \varphi(x) dx$ možemo diferencirati k puta. Pri tome je

$$f^{(k)}(t) = i^k \int x^k e^{itx} \varphi(x) dx,$$

a nakon zamjene $t = 0$ dobijamo našu jednakost.

e) Neposredno iz teoreme o Maklorenovom razvoju funkcije slijedi: Ako $E|X|^n$ postoji, tada za karakterističnu funkciju slučajne veličine X važi asimptotska jednakost

$$f(t) = \sum_{k=1}^n EX^k \frac{(it)^k}{k!} + o(t^n), \quad t \rightarrow 0.$$

Kao što postoji karakterizacija klase funkcija raspodjele (teorema ??), tako postoji i teorema (Bohner–Hinčinova teorema) koja daje potrebne i dovoljne uslove da bi neka funkcija bila karakteristična funkcija neke slučajne veličine. Već smo konstatovali da svaka slučajna veličina ima karakterističnu funkciju. Na pitanje da li karakteristična funkcija jednoznačno određuje raspodjelu slučajne veličine, pozitivni odgovor daje sljedeća teorema.

Teorema 1.11.1 (Teorema jedinstvenosti.) *Funkcija raspodjele slučajne veličine jednoznačno je određena njenom karakterističnom funkcijom.*

Sljedeći odjeljak će biti posvećen graničnim teoremama. Pripremićemo odgovarajući analitički aparat.

DEFINICIJA 1.11.2 *Niz funkcija raspodjele $F_n(x)$ kompletno konvergira ka funkciji raspodjele $F(x)$ ako važi*

1^o *U svim tačkama neprekidnosti funkcije $F(x)$ važi $F_n(x) \rightarrow F(x)$ kada $n \rightarrow \infty$.*

2^o *$F_n(\pm\infty) \rightarrow F(\pm\infty)$ kada $n \rightarrow \infty$.*

Teorema 1.11.2 *Ako niz karakterističnih funkcija $f_n(t)$ konvergira ka funkciji $f(t)$ koja je neprekidna za $t = 0$, tada niz odgovarajućih funkcija raspodjele $F_n(x)$ kompletno konvergira ka funkciji raspodjele $F(x)$ čija je karakteristična funkcija $f(t)$.*

DEFINICIJA 1.11.3 *Za niz slučajnih veličina X_n kažemo da konvergira u raspodjeli ka slučajnoj veličini X ako odgovarajući niz funkcija raspodjele $F_n(x)$ kompletno konvergira ka funkciji raspodjele $F(x)$ slučajne veličine X . Koristimo zapis $X_n \xrightarrow{R} X$.*

Slobodnije rečeno, sa uvećavanjem broja n , funkcije raspodjela slučajnih veličina X_n se približavaju funkciji raspodjele slučajne veličine X .

DEFINICIJA 1.11.4 *Za niz slučajnih veličina X_n kažemo da konvergira u vjerovatnoći ka slučajnoj veličini X ako za svako $\varepsilon > 0$ važi*

$$P\{|X_n - X| > \varepsilon\} \longrightarrow 0 \quad (\Leftrightarrow \quad P\{|X_n - X| \leq \varepsilon\} \longrightarrow 1), \quad n \rightarrow \infty.$$

Koristimo zapis $X_n \xrightarrow{P} X$.

Slobodnije rečeno, za veliko n je sa vjerovatnoćom koja je bliska 1 (veoma je vjerovatno) moduo razlike X_n i X manji od ε (ova interpretacija je najefektnija kada je ε malo).

Pokazuje se da ako niz $X_n \xrightarrow{P} X$, tada $X_n \xrightarrow{R} X$. Obrnuto u opštem slučaju ne važi. Takođe se pokazuje da ako $X_n \xrightarrow{R} a$, a je konstanta, tada $X_n \xrightarrow{P} a$.

Primjer 1. $f_{I_A}(t) = qe^{it0} + pe^{it1} = q + pe^{it}$. ◀

Primjer 2. $S_n : \mathcal{B}(n, p)$. Iz svojstva c) slijedi

$$f_{S_n}(t) = \prod_{k=1}^n f_{I_{A_k}}(t) = (q + pe^{it})^n. \quad \blacktriangleleft$$

Primjer 3. $X : \mathcal{P}(\lambda)$, $f(t) = \sum_{j=0}^{\infty} e^{itj} \frac{\lambda^j}{j!} e^{-\lambda} = e^{\lambda(e^{it}-1)}$. Na osnovu svojstva

d) imamo

$$f'(0) = iEX = i\lambda \implies EX = \lambda, \quad f''(0) = -EX^2 = -(\lambda^2 + \lambda) \implies DX = \lambda. \quad \blacktriangleleft$$

Primjer 4. Neka su $X_1 : \mathcal{P}(\lambda_1), \dots, X_n : \mathcal{P}(\lambda_n)$ nezavisne slučajne veličine. Naći raspodjelu slučajne veličine $Y = X_1 + \dots + X_n$.

$$\blacktriangleright f_Y(t) = \prod_{i=1}^n f_{X_i}(t) = e^{(\lambda_1 + \dots + \lambda_n)(e^{it}-1)},$$

pa na osnovu teoreme jedinstvenosti zaključujemo $Y : \mathcal{P}(\lambda_1 + \dots + \lambda_n)$. ◀

Primjer 5. Slučajna veličina X ima $\mathcal{N}(0, 1)$ raspodjelu.

$$f(t) = \int \frac{e^{itx}}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int \cos tx e^{-\frac{x^2}{2}} dx.$$

Nakon diferenciranja dobijamo $f'(t) = -tf(t)$. Rješenje ove diferencijalne jednačine je $f(t) = Ce^{-\frac{t^2}{2}}$. Kako je $f(0) = 1$, dobijamo $C = 1$. Dakle, $f(t) = e^{-\frac{t^2}{2}}$. ◀

Primjer 6. $X : \mathcal{U}(-1, 1)$, $f(t) = \int_{-1}^1 \frac{e^{itx}}{2} dx = \frac{\sin t}{t}$, $t \neq 0$, $f(0) = 1$. ◀

Primjer 7. Slučajne veličine X_n , $n = 1, 2, \dots$ su nezavisne i jednako-raspodijeljene,

$$X_n : \begin{matrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{matrix}.$$

Neka je $Y_n = \sum_{k=1}^n \frac{X_k}{2^k}$, $n = 1, 2, \dots$. Dokazati da $Y_n \xrightarrow{R} Y$, gdje je $Y : \mathcal{U}(-1, 1)$.

► Na početku izračunajmo

$$\begin{aligned} \lim_{n \rightarrow \infty} \cos \frac{t}{2} \cos \frac{t}{2^2} \cdots \cos \frac{t}{2^n} &= \lim_{n \rightarrow \infty} \frac{\cos \frac{t}{2} \cdots \cos \frac{t}{2^n} 2 \sin \frac{t}{2^n}}{2 \sin \frac{t}{2^n}} \\ &= \lim_{n \rightarrow \infty} \frac{\cos \frac{t}{2} \cdots \cos \frac{t}{2^{n-1}} \sin \frac{t}{2^{n-1}}}{2 \sin \frac{t}{2^n}} = \cdots = \lim_{n \rightarrow \infty} \frac{\sin t}{2^n \sin \frac{t}{2^n}} = \frac{\sin t}{t}, \quad t \neq 0 \text{ i} \end{aligned}$$

$f_{X_n}(t) = \frac{e^{-it} + e^{it}}{2} = \cos t$. Koristeći nezavisnost članova niza i svojstvo b) imamo

$$f_{Y_n}(t) = \prod_{k=1}^n \cos \frac{t}{2^k} \xrightarrow{n \rightarrow \infty} \frac{\sin t}{t}, \quad t \neq 0, \quad f_{Y_n}(0) \rightarrow 1.$$

Naš rezultat slijedi iz teoreme ?? i rezultata primjera 6.

Primjer 8. Slučajne veličine X i Y su nezavisne i svaka ima karakterističnu funkciju $f(t) = (2 - e^{it})^{-1}$. Naći raspodjelu slučajne veličine $Z = X + Y$.

$$\blacktriangleright f(t) = \frac{1}{2 - e^{it}} = \frac{1}{2} \cdot \frac{1}{1 - e^{it}/2} = \sum_{n=0}^{\infty} \frac{e^{int}}{2^{n+1}} \implies$$

$$P\{X = k\} = \frac{1}{2^{k+1}}, \quad k = 0, 1, 2, \dots$$

Koristeći razvoj $(1 - z)^{-2} = \sum_{n=1}^{\infty} nz^{n-1}$, dobijamo

$$f_Z(t) = \frac{1}{4} \cdot \left(1 - \frac{e^{it}}{2}\right)^{-2} = \sum_{n=0}^{\infty} \frac{n+1}{2^{n+2}} e^{int}.$$

Iz teoreme jedinstvenosti, dobija se

$$P\{Z = k\} = \frac{k+1}{2^{k+2}}, \quad k = 0, 1, 2, \dots \blacktriangleleft$$

Zadaci

1. Slučajna veličina X ima $\mathcal{N}(m, \sigma^2)$ raspodjelu. Metodom karakterističnih funkcija dokazati da slučajna veličina $X^* = \frac{X-m}{\sigma}$ ima $\mathcal{N}(0, 1)$ raspodjelu.

2. Neka su $X_1 : \mathcal{N}(m_1, \sigma_1^2), \dots, X_n : \mathcal{N}(m_n, \sigma_n^2)$ nezavisne slučajne veličine. Dokazati da slučajna veličina $S_n = \sum_{k=1}^n X_k$ ima $\mathcal{N}\left(\sum_{k=1}^n m_k, \sum_{k=1}^n \sigma_k^2\right)$ raspodjelu.

3. Slučajne veličine iz niza $X_n, n = 1, 2, \dots$ imaju gustine

$$\varphi_n(x) = n(1-x)^{n-1}, \quad x \in (0, 1), \quad n = 1, 2, \dots$$

Pokazati da niz slučajnih veličina $Y_n = nX_n$ u raspodjeli konvergira ka slučajnoj veličini Y koja ima $\mathcal{E}(1)$ raspodjelu.

1.12 Granične teoreme Teorije vjerovatnoća

Motiv za uvođenje pojma slučajne veličine je bio da se elementarnom ishodu pridruži realni broj. Pridruživanje realnog broja ishodu ostvaruje slučajna veličina i to pridruživanje realnog broja ishodu možemo tretirati kao mjerenje ishoda. U ovom tretmanu slučajna veličina predstavlja rezultat mjerenja. Ako se aktivira slučajna veličina X , tada je shema mjerenja sljedeća:

$$\omega \longrightarrow X(\omega) = \alpha.$$

Znači, slučajna veličina X je izmjerila ishod ω i rezultat mjerenja je broj α . Neka slučajne veličine $X_1, X_2, \dots$ predstavljaju redom rezultate prvog, drugog, ... mjerenja u seriji mjerenja. Davno je primijećeno da aritmetička

sredina mjerenja $\frac{X_1 + \dots + X_n}{n}$ pokazuje visoki stepen stabilnosti, iako se pojedinačna mjerenja mogu bitno razlikovati.

Svojstvo stabilnosti aritmetičkih sredina rezultata mjerenja, tj. realizacija slučajne veličine, se u vjerovatnoći izražava u obliku zakona velikih brojeva.

DEFINICIJA 1.12.1 *Kažemo da za niz slučajnih veličina $X_n, n = 1, 2, \dots$, važi zakon velikih brojeva ako niz $Y_n = \frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n EX_k$ u vjerovatnoći konvergira ka 0, tj. ako za svako $\varepsilon > 0$ vrijedi*

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n EX_k \right| > \varepsilon \right\} = 0.$$

Teorema 1.12.1 (Bernulijev zakon velikih brojeva.) *Neka je S_n slučajna veličina koja je jednaka broju uspjeha u Bernulijevoj shemi sa n ponavljanja polaznog opita i sa vjerovatnoćom uspjeha p . Tada $\frac{S_n}{n} \xrightarrow{P} p, n \rightarrow \infty$, tj. za svako $\varepsilon > 0$ vrijedi*

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{S_n}{n} - p \right| > \varepsilon \right\} = 0 \left(\iff \lim_{n \rightarrow \infty} P \left\{ \left| \frac{S_n}{n} - p \right| \leq \varepsilon \right\} = 1. \right) \quad (1)$$

Dokaz. Primjenom Čebišovljeve nejednakosti i korišćenjem činjenica $E \frac{S_n}{n} = \frac{np}{n} = p, D \frac{S_n}{n} = \frac{npq}{n^2} = \frac{pq}{n}$, dobijamo

$$P \left\{ \left| \frac{S_n}{n} - p \right| > \varepsilon \right\} \leq \frac{D \frac{S_n}{n}}{\varepsilon^2} = \frac{pq}{n\varepsilon^2} \rightarrow 0, n \rightarrow \infty. \blacklozenge$$

Primijetimo da slučajna veličina $\frac{S_n}{n}$ predstavlja relativnu frekvenciju uspjeha u seriji od n opita.

Asimptotskoj jednakosti (1) sada možemo dati sljedeću interpretaciju: Kada je n veliko, u seriji od n ponavljanja opita, sa velikom je vjerovatnoćom odstupanje relativne frekvencije uspjeha od vjerovatnoće uspjeha manje od ε ($\varepsilon > 0$ je proizvoljan broj). Malo drugačije rečeno, relativna frekvencija uspjeha dobro aproksimira vjerovatnoću uspjeha. Bernulijev zakon velikih brojeva opravdava klasičnu definiciju vjerovatnoće a posteriori.

Kako je $S_n = \sum_{k=1}^n I_{A_K}$, to je $\frac{S_n}{n} - p = \frac{1}{n} \sum_{k=1}^n I_{A_K} - \frac{1}{n} \sum_{k=1}^n EI_{A_K}$. Dakle, nalazimo se u modelu iz definicije ?? pa se time opravdava upotreba fraze "zakon velikih brojeva".

Teorema 1.12.2 (Čebišovljev zakon velikih brojeva.) *Neka su $X_1, X_2, \dots$ nezavisne slučajne veličine i neka su njihove disperzije ravnomjerno ograničene tj. $DX_n \leq C, n = 1, 2, \dots$, C je konstanta. Tada je za svako $\varepsilon > 0$*

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n EX_k \right| > \varepsilon \right\} = 0.$$

Dokaz. Primjenom nejednakosti Čebišova i svojstva b) disperzije, dobijamo

$$\begin{aligned} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n EX_k \right| > \varepsilon \right\} &\leq \frac{1}{\varepsilon^2} E \left(\frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n EX_k \right)^2 \\ &= \frac{1}{n^2 \varepsilon^2} E \left(\sum_{k=1}^n (X_k - EX_k) \right)^2 = \frac{1}{n^2 \varepsilon^2} D \sum_{k=1}^n X_k = \frac{1}{n^2 \varepsilon^2} \sum_{k=1}^n DX_k \\ &\leq \frac{C}{n \varepsilon^2} \rightarrow 0, n \rightarrow \infty. \end{aligned}$$

Naš rezultat slijedi na osnovu teoreme o policajcima. ♦

Teorema 1.12.3 (Hinčinov zakon velikih brojeva.) *Neka je $X_n, n = 1, 2, \dots$, niz nezavisnih jednakoraspodijeljenih slučajnih veličina, $EX_n = a, n = 1, 2, \dots$. Za niz X_n važi zakon velikih brojeva tj. $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} a$.*

Dokaz. Prvo ćemo naći karakterističnu funkciju konstante a (a je "slučajna veličina" koja svakom ishodu dodjeljuje vrijednost a). Očigledno, $f_a(t) = Ee^{iat} = e^{iat}$.

Neka je $f(t)$ karakteristična funkcija slučajne veličine $X_n, n = 1, 2, \dots$. Iz svojstva e) karakterističnih funkcija imamo

$$f(t) = 1 + ait + o(t), t \rightarrow 0,$$

a na osnovu svojstava b) i c) dobijamo da je karakteristična funkcija slučajne veličine $\frac{1}{n} \sum_{k=1}^n X_k$ jednaka $\left(f\left(\frac{t}{n}\right)\right)^n$. Sada je

$$\left(f\left(\frac{t}{n}\right)\right)^n = \left(1 + \frac{ait}{n} + o\left(\frac{t}{n}\right)\right)^n \rightarrow e^{iat}, n \rightarrow \infty.$$

Iz teoreme ?? slijedi $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{R} a$, no kako je a konstanta, na ovom mjestu važi i konvergencija u vjerovatnoći. ♦

Da bismo prokomentarisali Hinčinov zakon velikih brojeva, vratimo se interpretaciji u kojoj je slučajna veličina X tretirana kao rezultat mjerenja. Neka u seriji nezavisnih opita X_1 predstavlja rezultat prvog mjerenja, X_2 – drugog, ... itd. Zbog nezavisnosti opita, slučajne veličine $X_1, X_2, \dots$ su nezavisne. One su jednakoraspodijeljene i imaju istu raspodjelu kao i X (sve ove veličine jednako djeistvuju – mjere ishod). Ako je $EX = a$, tada je zbog jednake raspodijeljenosti $EX_n = a$, $n = 1, 2, \dots$. Nakon ovog uvoda možemo dati sljedeću interpretaciju Hinčinovog zakona velikih brojeva: Za veliko n , sa velikom je vjerovatnoćom odstupanje aritmetičke sredine $\frac{X_1 + \dots + X_n}{n}$ od a , manje od ε . Rečeno malo drugačije, aritmetička sredina rezultata mjerenja dobro aproksimira matematičko očekivanje slučajne veličine koja je identifikovana sa rezultatima mjerenja. Ovim je udahnut smisao pojmu matematičkog očekivanja.

Sa dva primjera ilustrirajmo naša teorijska razmatranja.

Očekivana suma brojeva koji se izvlače u igri LOTO je 140 (primjer 7 iz odjeljka 1.10). Pratićemo, recimo, 100 igara i nakon svake igre bilježićemo sumu izvučenih brojeva (u terminologiji mjerenja, mjerimo ishod). Potražićemo aritmetičku sredinu zabilježenih brojeva. Hinčinov zakon velikih brojeva tvrdi da će sa velikom vjerovatnoćom dobijena aritmetička sredina biti bliska broju 140. Iako ne znamo kolika će biti suma brojeva u nekoj konkretnoj igri (jer postoji veliki stepen neodređenosti), izvjesno je da će se u duгой seriji igara, aritmetička sredina suma stabilizovati oko matematičkog očekivanja slučajne veličine koja predstavlja sumu izvučenih brojeva.

Pretpostavimo da mjerimo neku fizičku veličinu čija je stvarna vrijednost a . Greška mjerenja je slučajna veličina ε . Naime, na grešku utiču faktori koji fluktuiraju – stalno se mijenjaju te ih smatramo slučajnim. Faktori su meteorološki, seizmički, subjektivni (vezani za psihofizičko stanje onoga ili onih koji mjere) itd. Ako ne postoji sistematska greška, tada je $E\varepsilon = 0$. Slučajna veličina koja predstavlja rezultat mjerenja je $X = a + \varepsilon$. Očigledno $EX = a$. Naše teorijsko razmatranje nam sugerise da rezultat mjerenja treba prihvatiti kao realizaciju slučajne veličine. Praksa potvrđuje pravilnost ovakvog pristupa. Rezultati mjerenja variraju od mjerenja do mjerenja i cilj mjerenja je da se što bliže odredi teorijsko a . Hinčinov zakon velikih brojeva savjetuje da se nepoznato a ocijeni aritmetičkom sredinom rezultata serije mjerenja.

Rasprostranjenost normalne raspodjele u opisivanju slučajnih pojava objašnjava centralna granična teorema (koristićemo skraćenicu CGT).

Teorema 1.12.4 (Centralna granična teorema.) *Neka su $X_1, X_2, \dots$ nezavisne, jednakoraspodijeljene slučajne veličine, $EX_n = w$, $DX_n = \sigma^2$,*

$S_n = \sum_{k=1}^n X_k$, $n = 1, 2, \dots$ Tada važi

$$P \left\{ \frac{S_n - wn}{\sigma\sqrt{n}} < x \right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad n \rightarrow \infty. \quad (2)$$

Dokaz. Neka je $X_n^* = \frac{X_n - w}{\sigma}$, $n = 1, 2, \dots$ Kako je $EX_n^* = 0$, $DX_n^* = 1$, odgovarajuća karakteristična funkcija je

$$f(t) = 1 - \frac{t^2}{2} + o(t^2), \quad t \rightarrow 0.$$

Na lijevoj strani formule (??) figuriše funkcija raspodjele vjerovatnoća slučajne veličine $\frac{S_n - wn}{\sigma\sqrt{n}}$, a na desnoj strani funkcija raspodjele vjerovatnoća slučajne veličine $X^* : \mathcal{N}(0, 1)$. Zbog toga, tvrđenje teoreme je ekvivalentno sa $\frac{S_n - wn}{\sigma\sqrt{n}} \xrightarrow{R} X^*$. Poslužimo se metodom karakterističnih funkcija.

$$\frac{S_n - wn}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k^* = W_n,$$

pa je $\left(f\left(\frac{t}{\sqrt{n}}\right)\right)^n$ odgovarajuća karakteristična funkcija slučajne veličine W_n . Na kraju imamo

$$\left(f\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \rightarrow e^{-\frac{t^2}{2}}, \quad n \rightarrow \infty.$$

Tvrđenje slijedi iz teoreme ?? i činjenice da je $e^{-\frac{t^2}{2}}$ karakteristična funkcija slučajne veličine $X^* : \mathcal{N}(0, 1)$. ♦

Posljedica CGT je sljedeća teorema.

Teorema 1.12.5 (Muavr–Laplasova teorema.) U Bernulijevoj shemi u kojoj je vjerovatnoća uspjeha p , važi

$$P \left\{ \frac{S_n - np}{\sqrt{npq}} < x \right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad n \rightarrow \infty.$$

Da bismo dokazali Muavr–Laplasovu teoremu, dovoljno se pozvati na reprezentaciju $S_n = \sum_{k=1}^n I_{A_k}$ i činjenice $EI_{A_k} = p$, $DI_{A_k} = pq$. Tvrđenje sada slijedi iz CGT.

Označimo sa $X_{n,k}^* = \frac{X_k}{\sqrt{n}}$. Pošto je $EX_{n,k}^* = 0$ i $DX_{n,k}^* = \frac{1}{n}$, možemo konstatovati da su slučajne veličine $X_{n,k}^*$ za velike brojeve n , "koncentrisane" oko 0. Slobodnije rečeno, $X_{n,k}^*$ je "mala" slučajna veličina. Zadržimo se na jednakosti $\frac{S_n - wn}{\sigma\sqrt{n}} = \sum_{k=1}^n X_{n,k}^*$. CGT se tvrdi da se ova suma asimptotski u raspodjeli ponaša po normalnom zakonu. Ovo tvrđenje važi i pri znatno slabijim uslovima. Imamo u vidu situacije kada su članovi sume zavisne i nejednako raspodijeljene slučajne veličine. Poruka CGT glasi: Suma "malih" slučajnih veličina (faktora) se asimptotski, pri veoma opštim uslovima, ponaša po normalnom zakonu.

U praksi često posmatramo slučajne veličine koje su dobijene u rezultatu sumiranja velikog broja nezavisnih ili slabo zavisnih "malih" slučajnih faktora. Nas ne zanimaju komponente, već efekat njihovog udruživanja. Model se uklapa u CGT pa je sumarna tj. posmatrana slučajna veličina raspodijeljena po normalnom zakonu.

Već smo konstatovali da rezultat mjerenja fizičke veličine čija je stvarna vrijednost a , opisuje slučajna veličina $X = a + \varepsilon$, $E\varepsilon = 0$. Greška ε je jednaka sumi grešaka koje su rezultat uticaja temperature, vlažnosti, pritiska, vibracije tla i mnogih drugih faktora. CGT tvrdi da je sumarna greška ε raspodijeljena po normalnom zakonu te je $X : \mathcal{N}(a, \sigma^2)$. Broj σ^2 je disperzija greške i on karakteriše preciznost mjerenja.

Slika 16: Gausova funkcija.

Podaci o konkretnim mjerenjima potvrđuju izvanrednu saglasnost teorije i prakse.

Prije nego što uradimo nekoliko konkretnih primjera, skrećemo pažnju na neke detalje. Polazeći od činjenice da je

$$P\{a \leq X < b\} = P\{X < b\} - P\{X < a\},$$

CGT se može zapisati kao

$$P\left\{a < \frac{S_n - wn}{\sqrt{n\sigma^2}} < b\right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt, n \rightarrow \infty.$$

Koristićemo zapise $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$ i $\Phi^*(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$, $x >$

0. U tablicama normalne raspodjele su date vrijednosti za $\Phi^*(x)$. Sa z_α se označava broj za koji je $\Phi^*(z_\alpha) = \alpha$. Broj z_α nalazimo u tablicama.

Primjer 1. Kocka se baca 100 puta. Kolika je vjerovatnoća da je zbir palih brojeva između 340 i 370?

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

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

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

Niz vjerovatnoća u CGT brzo konvergira, 100 je dovoljno veliki 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 koristiti simbol za približnu vrijednost već simbol jednakosti.

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

Primjer 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. Koliko treba da je n pa da vjerovatnoća pravilne odluke bude 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

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

Primjer 3. Strijelac pogađa metu sa vjerovatnoćom 0,4 i gađa u nju 150 puta. Naći bar jedan interval u kojem će se, sa vjerovatnoćom koja iznosi bar 0,8, nalaziti broj pogodaka.

► Slučajna veličina $S_{150} : \mathcal{B}(150; 0,4)$ predstavlja broj pogodaka. Tražiće-mo interval oblika (A, B) .

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

te je zbog cjelobrojnosti broja pogodaka traženi interval $(52,68)$. ◀

Primjer 4. Koliko puta treba baciti novčić pa da sa vjerovatnoćom 0,95 odstupanje relativne učestalosti pojave pisma od $\frac{1}{2}$ bude manje od 0,02.

► Bernulijevim zakonom velikih brojeva se tvrdi da relativna učestalost pojave pisma u vjerovatnoći konvergira ka $\frac{1}{2}$. Dakle, kada opit tj. bacanje novčića ponovimo n puta, n je veliki broj, sa vjerovatnoćom koja je bliska 1 relativna učestalost pisma će biti bliska $\frac{1}{2}$.

Muavr-Laplasova teorema nam daje mogućnost da nađemo n za koje je sa vjerovatnoćom α (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.

$$\begin{aligned} P\left\{\left|\frac{S_n}{n} - 0,5\right| < 0,02\right\} &= P\left\{\left|\frac{S_n - 0,5n}{\sqrt{0,5 \cdot 0,5n}}\right| < 0,04\sqrt{n}\right\} = 2\Phi^*(0,04\sqrt{n}) \\ &= 0,95 \implies 0,04\sqrt{n} = 1,96 \implies n = 2401. \end{aligned}$$

Korišćenjem nejednakosti Čebišova možemo naći grubu aproksimaciju broja n . Tražiće-mo n za koje je vjerovatnoća $\geq 0,95$.

$$\begin{aligned} 0,95 \leq P\left\{\left|\frac{S_n}{n} - 0,5\right| < 0,02\right\} &\implies P\left\{\left|\frac{S_n}{n} - 0,5\right| \geq 0,02\right\} \leq 0,05 \\ \implies \frac{1}{4n(0,02)^2} < 0,05 &\implies n > 12500. \quad \blacktriangleleft \end{aligned}$$

Primjer 5. Vratimo se na primjer 3 iz odjeljka 1.5. Vjerovatnoća traženog događaja je 0,24, a njegova relativna frekvencija zabilježena nakon 100 konkretnih dijeljenja je 0,26. Izračunajmo kolika je vjerovatnoća da odstupanje relativne frekvencije od vjerovatnoće, nakon 100 ponavljanja opita–dijeljenja, bude bar 0,02.

$$\begin{aligned} \blacktriangleright P \left\{ \left| \frac{S_{100}}{100} - p \right| \geq 0,02 \right\} &= 1 - P \left\{ \left| \frac{S_{100} - 100p}{\sqrt{100pq}} \right| < 0,02 \sqrt{\frac{100}{pq}} \right\} \\ &= 1 - 2\Phi^*(0,455) \approx 0,65, \quad p = 0,24. \end{aligned}$$

Ako mnogo puta ponovimo seriju od 100 dijeljenja, u približno $\frac{2}{3}$ serija, nakon stotog dijeljenja relativna frekvencija će se razlikovati od vjerovatnoće bar za 0,02.

Znači, odstupanje od 0,02 koje smo registrovali nije neočekivano. ◀

Zadaci

1. Novčić se baca 400 puta. Kolika je vjerovatnoća da će grb pasti više od 196 i manje od 206 puta?

2. Bazen zapremine $200 m^3$ se puni preko dva izvora, od kojih prvi daje $1 m^3$ na sat, a drugi $2 m^3$ na sat. Uvijek radi samo jedan izvor u toku sata, prvi sa vjerovatnoćom 0,6, a drugi sa vjerovatnoćom 0,4. Izvori se po satima nezavisno uključuju. Kolika je vjerovatnoća da punjenje bude završeno za manje od 165 sati?

3. Osiguravajuće društvo osigurava 10000 lica od povrede. Vjerovatnoća povrede u toku godine je 0,006. Premija osiguranja je 700 dinara, a u slučaju povrede, odšteta je 100000 dinara. Kolika je vjerovatnoća da će društvo poslovati sa dobitkom?

4. Broj automobila koji prođu kroz raskrsluicu ima $\mathcal{P}(6)$ raspodjelu. Kolika je vjerovatnoća da će u toku dva sata kroz raskrsluicu proći bar 700 automobila?

5. Kolika je vjerovatnoća da među 10000 novorođenčadi bude bar 50 dječaka više nego djevojčica?

6. Aparat za igru izbacuje broj k , $k \in \{0, 1, 2, \dots\}$, sa vjerovatnoćom $p_k = \frac{1}{e k!}$. Ako se pojavi parni broj, igrač dobija jedan dinar, a ako se pojavi neparni broj, gubi jedan dinar. Naći vjerovatnoću da je nakon 1000 uključivanja aparata dobitak igrača između 10 i 50 dinara.

7. Prosječno, svaki treći prolaznik koji prođe pored kioska kupi novine. Neka je S_{100} broj lica koja prođu pored kioska dok se ne proda prvih 100 primjeraka novina. Naći približnu raspodjelu slučajne veličine S_{100} .

8. Računar sabira N brojeva koje prethodno zaokruži na n -toj decimali. Naći vjerovatnoću da je sumarna greška po modulu manja od δ .

9. Vijek trajanja sijalice (u satima) je slučajna veličina sa $\mathcal{E}(0,005)$ raspodjelom. Koliko sijalica treba imati u rezervi da bi sa pouzdanošću 0,98 sumarno vrijeme trajanja sijalica bilo bar 10000 sati?

10. Fabrika u toku dana proizvede 1000 automobila. Vjerovatnoća da automobil zahtijeva doradu iznosi 0,05. Koliko treba da bude kapacitet parkinga pa da sa vjerovatnoćom 0,9 bude dovoljan za automobile koji čekaju doradu?

11. U pozorište koje ima 1000 mjesta, posjetioци ulaze tako što svako za sebe slučajno odabere jedan od dva ulaza. Pored svakog ulaza postoji garderoba. Koliko mjesta treba da bude u svakoj garderobi pa da sa vjerovatnoćom 0,99 svi posjetioци ostave stvari na bližoj garderobi?

12. Akumulacija se prazni svakog dana u toku slučajnog vremena i to sa raspodjelom $\mathcal{U}(50 \text{ min}, 70 \text{ min})$. Brzina pražnjenja je 1 l/min . Koliko najmanje vode treba da sadrži akumulacija da bi sa vjerovatnoćom 0,95 nakon 400 dana pražnjenja u njoj još bilo vode?

13. Strijelac pogađa cilj sa vjerovatnoćom 0,4. Koliko najmanje gađanja treba da planira pa da šansa da će imati bar 80 pogodaka bude bar 0,9?

Dio 2

Statistika

2.1 Statistički model

U vjerovatnoći smo slučajni opit modelirali vjerovatnosnim prostorom tj. trojkom $(\Omega, \mathcal{F}, P)$, gdje je Ω prostor ishoda, $\mathcal{F}$ je σ -algebra događaja i P je vjerovatnoća na $\mathcal{F}$. U vjerovatnosnim zadacima je trebalo naći vjerovatnoću nekog složenog događaja A iz σ -algebre $\mathcal{F}$.

Prilikom proučavanja slučajnih pojava iz prakse, model je uglavnom takav da sugerise prostor ishoda Ω slučajnog opita kao i klasu $\mathcal{F}$ podskupova od Ω koji predstavljaju događaje. Međutim, vjerovatnoća P nije unaprijed data. Na osnovu nama dostupnih informacija o karakteru slučajne pojave, obično imamo dovoljno razloga da pretpostavimo da vjerovatnoća P pripada nekoj klasi $\mathcal{P}$ vjerovatnoća zadatih na $\mathcal{F}$. Klasu $\mathcal{P}$ nazivamo familija dopustivih vjerovatnoća.

DEFINICIJA 2.1.1 *Uređena trojka $(\Omega, \mathcal{F}, \mathcal{P})$ se naziva statistički model razmatranog opita.*

Primjer 1. Nesimetrični novčić se baca n puta. Ako sa 1 označimo pojavu pisma, a sa 0 pojavu grba, onda je prostor ishoda

$$\Omega = \{(a_1, \dots, a_n) : a_k = 0 \text{ ili } a_k = 1, k = 1, 2, \dots, n\},$$

dok je $\mathcal{F} = \mathcal{P}(\Omega)$. Familija dopustivih vjerovatnoća je

$$\mathcal{P} = \{P_\theta : 0 \leq \theta \leq 1, \theta \neq 0, 5\},$$

gdje je vjerovatnoća P_θ određena sa

$$P_\theta((a_1, \dots, a_n)) = \theta^{a_1 + \dots + a_n} (1 - \theta)^{n - (a_1 + \dots + a_n)},$$

a θ je vjerovatnoća pojavljivanja pisma.

Primjer 2. Pretpostavimo da se mjeri neka fizička veličina čija je stvarna-teorijska vrijednost a nepoznata. Za skup Ω možemo uzeti skup R kojim su obuhvaćeni svi mogući rezultati mjerenja. Na ovom mjestu je udobnije obuhvatiti i brojeve koji čak ni teorijski ne mogu biti rezultati mjerenja (recimo, mjerimo dužinu koja je uvijek nenegativna), nego restriktivnošću pokvariti statistički model. Za σ -algebru događaja ćemo uzeti kolekciju

Borelovih skupova na R . Zbog toga što rezultatima mjerenja upravlja normalna raspodjela, za familiju vjerovatnoća na $\mathcal{B}$ ćemo uzeti

$$\mathcal{P} = \{P_{a, \sigma^2} : a \in R, \sigma^2 > 0\},$$

gdje je vjerovatnoća P_{a, σ^2} određena gustinom raspodjele

$$\varphi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left(\frac{x-a}{\sigma} \right)^2 \right\}, \quad x \in R. \blacktriangleleft$$

Proučavanje slučajnih pojava podrazumijeva mogućnost registrovanja tj. mjerenja raznih podataka. U primjeru 1 registruje se broj palih pisama u n bacanja novčića. U primjeru 2, nepoznata veličina se mjeri više puta, a registruju se sve izmjerene vrijednosti. Registrovani podaci se zovu statistički podaci.

Zadatak Matematičke statistike je da se na osnovu registrovanih podataka, odabere ona vjerovatnoća iz klase $\mathcal{P}$ koja će najbolje opisivati razmatranu slučajnu pojavu ili da se što je moguće više suzi klasa dopustivih vjerovatnoća. Postupak donošenja odluke o izboru vjerovatnoće P na osnovu statističkih podataka, naziva se statističko zaključivanje.

U vjerovatnoći smo u okvirima precizno zadatog modela tražili vjerovatnoću ostvarivanja nekog događaja. U slučaju kada su ishodi brojevi, računamo vjerovatnoću da se dobiju neki konkretni brojevi – podaci. Statistika rješava inverzni problem: Na osnovu realizovanih – registrovanih podataka treba nešto zaključiti o vjerovatnosnom modelu koji je generisao te podatke.

U mnogim istraživanjima registruju se podaci koji predstavljaju numeričke karakteristike nekog skupa Ω . Skup Ω koji se razmatra u statistici zove se **populacija**, a numerička karakteristika elemenata populacije zove se **obilježje**. Termin populacija je preuzet iz demografije još u vrijeme kada je statistika bila u povoju. Prva statistička istraživanja su rađena sa ciljem da se bolje upozna stanje u društvu (bila su to uglavnom istraživanja u demografiji i ekonomiji) pa latinska riječ status=stanje ulazi u korijen imena nauke koju želimo da upoznamo.

Primjer 1. Populaciju čine svi stanovnici Crne Gore, a obilježje elementa populacije tj. stanovnika je njegov nivo šećera u krvi. $\blacktriangleleft$

Primjer 2. Populaciju čine sve linije koje održava neka avio kompanija, a obilježje je vrijeme trajanja leta na liniji. $\blacktriangleleft$

Primjer 3. Populaciju čine svi dani u godini, a obilježje je vodostaj Morače u Podgorici izmjeren u podne. $\blacktriangleleft$

Neka se opit sastoji u izboru elemenata populacije i registrovanju vrijednosti obilježja na izabranom elementu. Ovaj opit generiše slučajnu veličinu

X koja je zadata na $(\Omega, \mathcal{F})$ i koja elementu ω iz populacije pridružuje njegovo obilježje. Nas obično zanima raspodjela slučajne veličine X te je pogodno razmatrati njen prostor vrijednosti $(R, \mathcal{B})$. Ako se opit ponovi n puta, ta serija opita generiše slučajni vektor $(X_1, \dots, X_n)$, a njegov prostor vrijednosti je $(R^n, \mathcal{B}^n)$. Slučajni vektor $(X_1, \dots, X_n)$ se zove **slučajni uzorak** ili kratko **uzorak**, a slučajna veličina X – **obilježje**.

Kako je na $(\Omega, \mathcal{F})$ zadata familija dopustivih vjerovatnoća $\mathcal{P}$, to ona generiše familiju $\mathcal{P}_X$ dopustivih raspodjela vjerovatnoća obilježja X , a isto tako generiše i familiju $\mathcal{F}_X$ dopustivih funkcija raspodjele vjerovatnoća obilježja X . Uobičajeni ulazni podaci u statističkim istraživanjima su: obilježje X i familija $\mathcal{P}_X$ njegovih dopustivih raspodjela vjerovatnoća. Često se umjesto familije dopustivih raspodjela vjerovatnoća zadaje familija $\mathcal{F}_X$ dopustivih funkcija raspodjele vjerovatnoća obilježja X .

Primjer 4. Obilježje X je broj kosmičkih čestica koje padnu na neki lokalitet u toku jednog sata, a familija dopustivih raspodjela obilježja je $\{\mathcal{P}(\lambda), \lambda > 0\}$. Obuhvaćene su sve Puasonove raspodjele, a za njih smo se odlučili zbog toga što vjerovatnosna teorija tvrdi da brojem palih čestica upravlja Puasonov zakon. ◀

Pretpostavimo da je na statističkom modelu $(\Omega, \mathcal{F}, \mathcal{P})$ zadato obilježje X . Pretpostavimo da n puta po modelu sa vraćanjem biramo element populacije i svaki put registrujemo obilježje na izabranom elementu populacije. Ovakav postupak generiše n -dimenzionalnu slučajnu veličinu $(X_1, \dots, X_n)$ kod koje su komponente nezavisne i jednako raspodijeljene kao i obilježje X .

DEFINICIJA 2.1.2 Prost slučajni uzorak je slučajni vektor $(X_1, \dots, X_n)$ kod koga su komponente nezavisne i svaka je raspodijeljena kao obilježje X .

Uбудuće ćemo pod terminom uzorak podrazumijevati 1^o elemente populacije na kojima je registrovano obilježje, 2^o prost slučajni uzorak. Koji od ova dva pojma imamo u vidu, biće jasno iz konteksta. **Realizovani uzorak** $(x_1, \dots, x_n)$ su registrovane vrijednosti obilježja u sprovedenoj seriji biranja elemenata populacije i registrovanja njihovih obilježaja. Broj n nazivamo **obim uzorka**.

Za uzorak važi $(X_1, \dots, X_n) : \Omega^n \longrightarrow R^n$ i

$$(X_1, \dots, X_n)(\omega) = (X(\omega_1), \dots, X(\omega_n)), \quad \omega = (\omega_1, \dots, \omega_n).$$

Osnovni zadatak Statistike, nakon uvođenja pojma obilježja, možemo nešto drugačije formulisati. Obilježje X ima svoju teorijsku i nama nepoznatu funkciju raspodjele vjerovatnoća $F(x)$. Zadatak Statistike je da ukaže na postupke kojima ćemo na osnovu registrovanih vrijednosti obilježja nešto

bliže zaključiti o nepoznatoj funkciji $F(x)$. To bliže zaključivanje je ekvivalentno sa sužavanjem familije $\mathcal{F}_X$.

DEFINICIJA 2.1.3 *Neka je $(X_1, \dots, X_n)$ uzorak obilježja X . Funkcija $F_n : \Omega^n \times R \rightarrow R$ data sa*

$$F_n(x, \omega) = \frac{1}{n} \sum_{k=1}^n I\{X_k(\omega) < x\},$$

*zove se **empirijska funkcija raspodjele obilježja X** .*

Argument ω se obično ne zapisuje. Za svaki realni broj x važi: $F_n(x)$ je slučajna veličina koja ima $\mathcal{B}(n, p)$ raspodjelu gdje je $p = F(x)$.

Neka je npr. $n = 5$, a $(x_1, \dots, x_5)$ realizovani uzorak. Realizacija slučajne funkcije $F_5(x)$ data je na slici 17.

Slika 17: Empirijska funkcija raspodjele.

Primijetimo da su skokovi veličine $\frac{1}{5}$.

Navešćemo bez dokaza Glivenko–Kantelijevu teoremu. U literaturi se ova teorema pominje i kao centralna teorema matematičke statistike. Njome se tvrdi da empirijska funkcija raspodjele asimptotski, i to ravnomjerno na R , aproksimira teorijsku i nama nepoznatu funkciju raspodjele vjerovatnoća obilježja X . Ovaj rezultat koristimo tako što na osnovu realizovanog uzorka nacrtamo realizaciju empirijske funkcije raspodjele, a ova funkcija nam sada daje prvu bližu predstavu o nepoznatoj funkciji $F(x)$. Što je obim uzorka veći to je aproksimacija bolja.

Teorema 2.1.1 (Glivenko–Kantelijeva teorema.) *Važi asimptotska jednakost*

$$P \left\{ \lim_{n \rightarrow \infty} \sup_{x \in R} |F_n(x) - F(x)| = 0 \right\} = 1.$$

DEFINICIJA 2.1.4 *Neka je $(X_1, \dots, X_n)$ uzorak obilježja X i neka je $f : R^n \rightarrow R$ Borelova funkcija. Slučajna veličina $V = f(X_1, \dots, X_n)$ naziva se **statistika**.*

Primjer 5. Statistika $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ se naziva **uzoračka sredina**. ◀

Primjer 6. Statistika $\bar{S}_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2$ se naziva **uzoračka disperzija**. Elementarno se dokazuje da je $\bar{S}_n^2 = \frac{1}{n} \sum_{k=1}^n X_k^2 - \bar{X}_n^2$. ◀

Primjer 7. Statistika $Y_1 = \min_{1 \leq k \leq n} X_k$ se naziva **uzorački minimum**, a statistika $Y_n = \max_{1 \leq k \leq n} X_k$ se naziva **uzorački maksimum**. ◀

Primjer 8. Neka za obilježje X važi $EX = m$, $DX = \sigma^2$. Imamo $E\bar{X}_n = m$, $D\bar{X}_n = \frac{\sigma^2}{n}$, kao i

$$E\bar{S}_n^2 = \frac{1}{n} \sum_{k=1}^n EX_k^2 - E\bar{X}_n^2 = \sigma^2 + m^2 - \left(\frac{\sigma^2}{n} + m^2 \right) = \frac{n-1}{n} \sigma^2. \blacktriangleleft$$

Sljedeće dvije raspodjele imaju veliki značaj u Statistici.

Neka su $X_1, \dots, X_n$ nezavisne slučajne veličine sa $\mathcal{N}(0, 1)$ raspodjelom. Raspodjela slučajne veličine

$$\chi_n^2 = X_1^2 + \dots + X_n^2,$$

naziva se χ kvadrat raspodjela sa n stepeni slobode i označava sa χ_n^2 . Odgovarajuću gustina je

$$h_n(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

Slika 18: Gustina χ_n^2 raspodjele.

U tablicama χ_n^2 raspodjele nalaze se vrijednosti $h_{n,\alpha}$ za koje je

$$\int_0^{h_{n,\alpha}} h_n(x) dx = \alpha.$$

Neka su $X_0, X_1, \dots, X_n$ nezavisne slučajne veličine sa $\mathcal{N}(0, 1)$ raspodjelom. Raspodjela slučajne veličine

$$t_n = \frac{X_0}{\sqrt{\frac{1}{n} (X_1^2 + \dots + X_n^2)}} = \frac{X_0}{\sqrt{\frac{1}{n} \chi_n^2}},$$

naziva se Studentova raspodjela sa n stepeni slobode i označava sa t_n . Odgovarajuća gustina je određena sa

$$v_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \quad x \in \mathbb{R}.$$

Slika 19: Gustina studentove t_n raspodjele.

U tablicama t_n raspodjele nalaze se vrijednosti $v_{n,\alpha}$ za koje je

$$\int_0^{v_{n,\alpha}} v_n(x) dx = \alpha.$$

Sljedeća teorema je posvećena svojstvima uzoračke sredine i uzoračke disperzije obilježja koje ima normalnu raspodjelu.

Teorema 2.1.2 (Fišerova teorema.) *Neka je $(X_1, \dots, X_n)$ uzorak obilježja X koje ima $\mathcal{N}(m, \sigma^2)$ raspodjelu. Tada važi:*

- Slučajna veličina $\frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}} \sqrt{n}$ ima $\mathcal{N}(0, 1)$ raspodjelu.*
- Slučajna veličina $\frac{n\bar{S}_n^2}{\sigma^2}$ ima χ_{n-1}^2 raspodjelu.*
- Slučajne veličine $\bar{X}_n$ i $\bar{S}_n^2$ su nezavisne.*
- Slučajna veličina $\frac{\bar{X}_n - m}{\frac{\bar{S}_n}{\sqrt{n}}} \sqrt{n-1}$ ima t_{n-1} raspodjelu.*

2.2 Tačkasta ocjena nepoznatog parametra

U uvodnom dijelu smo konstatovali da je zadatak onih koji rade statistiku da na osnovu statističkih podataka nešto zaključuje o nepoznatoj raspodjeli

obilježja X . Vjerovatnosni model koji generiše nama dostupne podatke nije nam poznat i prirodna je potreba da se model što bolje ocijeni – aproksimira. Obično imamo situaciju kada su dati obilježje X i familija dopustivih funkcija raspodjele

$$\mathcal{F}_X = \{F(x, \theta), \theta \in \Theta\}$$

kojoj pripada nepoznata teorijska funkcija raspodjele vjerovatnoća $F(x)$ obilježja X . Zbog korespondencije na relaciji raspodjela vjerovatnoća slučajne veličine – funkcija raspodjele vjerovatnoća slučajne veličine, familija $\mathcal{F}_X$ jednoznačno određuje familiju $\mathcal{P}_X$ dopustivih raspodjela obilježja X . Jasno, važi i obrnuto. U slučaju kada je obilježje X apsolutno neprekidna slučajna veličina, dopustiva familija raspodjela se često zadaje navođenjem odgovarajuće familije gustina $\varphi(x, \theta)$, $\theta \in \Theta$. U slučaju kada je obilježje X diskretna slučajna veličina, dopustiva familija raspodjela se uobičajeno zadaje navođenjem vjerovatnoća $p(x_\nu, \theta) = P_\theta\{X = x_\nu\}$, $\theta \in \Theta$, sa kojima obilježje X uzima odgovarajuće vrijednosti x_ν .

Primjer 1. Navešćemo neke familije raspodjela.

$$\mathcal{P}_1 = \{\mathcal{N}(m, \sigma^2), m \in R, \sigma^2 > 0\}.$$

$$\mathcal{P}_2 = \{\mathcal{N}(m, 1), m \in R\}.$$

$$\mathcal{P}_3 = \{\mathcal{N}(0, \sigma^2), \sigma^2 > 0\}.$$

$$\mathcal{P}_4 = \{\mathcal{U}(0, \theta), \theta \in (1, 2)\}.$$

$$\mathcal{P}_5 = \{\mathcal{P}(\lambda), \lambda > 0\}.$$

$$\mathcal{P}_6 = \{\mathcal{B}(n, p), p \in (0, 1)\}.$$

Vidimo da familije $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3, \mathcal{P}_4$ obuhvataju raspodjele slučajnih veličina koje su apsolutno neprekidne, a familije $\mathcal{P}_5$ i $\mathcal{P}_6$ obuhvataju raspodjele slučajnih veličina koje su diskretne.

Problem ocjenjivanja nepoznate funkcije $F(x)$ svodi se na problem ocjenjivanja nepoznatog parametra θ . Ako je θ_0 ocjena nepoznatog parametra θ , tada se za ocjenu nepoznate funkcije $F(x)$ uzima funkcija $F(x, \theta_0)$.

Neka je $(X_1, \dots, X_n)$ uzorak obilježja X . Pod tačkastim ocjenjivanjem parametra θ podrazumijeva se sljedeća procedura: Bira se statistika $V_n = V_n(X_1, \dots, X_n)$ koja se zove ocjena parametra θ . Ako je $(x_1, \dots, x_n)$ realizovani uzorak, onda se za ocjenu nepoznatog parametra θ uzima broj $V_n(x_1, \dots, x_n)$ tj. realizacija statistike V_n .

Pretpostavimo da obilježje X ima funkciju raspodjele vjerovatnoća $F(x, \theta)$. Za ocjenu $V_n = V(X_1, \dots, X_n)$ kažemo da je **centrirana** (ili **nepristrasna**) ako je $EV_n = \theta$. Za ocjenu V_n kažemo da je **postojana** ako $V_n \xrightarrow{P} \theta$.

Primjer 2. Neka je $(X_1, \dots, X_n)$ uzorak obilježja X , pri čemu je $EX = m$, $DX = \sigma^2$. Elementarno se provjerava

- a) statistika $\bar{X}_n$ je centrirana ocjena parametra m ,
- b) statistika $\bar{S}_n^2$ nije centrirana ocjena parametra σ^2 ,
- c) statistika $\bar{S}_n'^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$ je centrirana ocjena parametra

σ^2 i naziva se **popravljen uzoračka disperzija**. ◀

Ako su V_n i T_n dvije centrirane ocjene nepoznatog parametra θ i ako je $DV_n < DT_n$, tada kažemo da je ocjena V_n bolja od ocjene T_n . Prirodno je za bolju ocjenu proglasiti onu kod koje je stepen koncentracije oko nepoznatog parametra veći.

Primjer 3. Obilježje X ima raspodjelu koja pripada familiji dopustivih raspodjela $\mathcal{P} = \{\mathcal{U}(0, \theta), \theta > 0\}$. Dokazati da su statistike $T_n = 2\bar{X}_n$ i $U_n = \frac{n+1}{n}Y_n$ centrirane i postoje ocjene parametra θ . Koja je od ove dvije ocjene bolja?

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

$$ET_n = \theta, \quad DT_n = \frac{4}{n} \cdot \frac{\theta^2}{12} = \frac{\theta^2}{3n}.$$

Za svako $\varepsilon > 0$, na osnovu Čebišovljeve nejednakosti, dobijamo

$$P\{|T_n - \theta| > \varepsilon\} \leq \frac{DT_n}{\varepsilon^2} = \frac{\theta^2}{3n\varepsilon^2} \longrightarrow 0, \quad n \rightarrow \infty.$$

Znači, ocjena T_n je centrirana i postojana.

Istovjetno kao u primjeru 8 odjeljka 1.9 dobijamo da je $\varphi_n(x) = nx^{n-1}/\theta^n$, $x \in (0, \theta)$, gustina slučajne veličine Y_n . Sada imamo

$$EU_n = (n+1) \int_0^\theta \frac{x^n}{\theta^n} dx = \theta,$$

$$DU_n = \left(\frac{n+1}{n}\right)^2 \int_0^\theta n \frac{x^{n+1}}{\theta^n} dx - \theta^2 = \frac{\theta^2}{n(n+2)}.$$

Za svako $\varepsilon > 0$, na osnovu Čebišovljeve nejednakosti, dobijamo

$$P\{|U_n - \theta| > \varepsilon\} \leq \frac{DU_n}{\varepsilon^2} = \frac{\theta^2}{n(n+2)\varepsilon^2} \longrightarrow 0, \quad n \rightarrow \infty.$$

Dakle, i ocjena U_n je centrirana i postojana.

Primijetimo da je $DT_1 = DU_1$, dok je $DU_n < DT_n$, $n = 2, 3, \dots$. Dakle, za $n \geq 2$ ocjena U_n je bolja od ocjene T_n .

Naše obilježje X ima neku konkretnu $\mathcal{U}(0, \theta)$, $\theta > 0$, raspodjelu, θ je nama nepoznat parametar. Dokazali smo da je statistika U_n koju generiše obilježje X , centrirana ocjena nepoznatog parametra θ . Ukazaćemo na praktično značenje ovog rezultata.

Pretpostavimo da n puta registrujemo obilježje. Neka je $(x_1^{(1)}, \dots, x_n^{(1)})$ dobijena n -torka. Neka je $(x_1^{(2)}, \dots, x_n^{(2)})$ n -torka dobijena nakon što smo još jednom n puta registrovali obilježje. Neka je $(x_1^{(N)}, \dots, x_n^{(N)})$ n -torka koju dobijemo kada N -ti put (N je veliki broj) obavimo istu proceduru. Potražimo $U_n(x_1^{(1)}, \dots, x_n^{(1)}) = u_n^{(1)}, \dots, U_n(x_1^{(N)}, \dots, x_n^{(N)}) = u_n^{(N)}$. Podatke tj. vrijednosti obilježja generiše slučajna veličina koja ima $\mathcal{U}(0, \theta)$ raspodjelu. Pošto ne znamo θ , vjerovatnosni model koji upravlja podacima nam nije poznat. Zbog centriranosti ocjene U_n , a na osnovu zakona velikih brojeva, možemo konstatovati da se aritmetička sredina brojeva $u_n^{(1)}, \dots, u_n^{(N)}$ nalazi "veoma blizu" nepoznatog θ . Uz pomoć statističkih podataka stvorena je dobra predstava o vjerovatnosnom modelu koji nam "šalje" baš te podatke. ◀

Otvora se problem traženja dobre ocjene, tj. ocjene koja će valjano aproksimirati nepoznati parametar. Metod maksimalne vjerodostojnosti, kome će biti posvećeni redovi koji slijede, daje efektivni postupak za traženje kvalitetne ocjene. Ocjena dobijena ovim metodom je u izvjesnom smislu optimalna.

Pretpostavimo da je obilježje X diskretno i da je familija dopustivih raspodjela $\{p(x_\nu, \theta), \theta \in \Theta\}$. Funkcija

$$L(\theta) = L(\theta, x_1, \dots, x_n) = p(x_1, \theta) \cdots p(x_n, \theta), \quad \theta \in \Theta,$$

gdje je θ argument, a $x_1, \dots, x_n$ parametri, naziva se **funkcija vjerodostojnosti**. Zbog nezavisnosti komponenti $X_1, \dots, X_n$ imamo

$$L(\theta) = L(\theta, x_1, \dots, x_n) = P_\theta\{(X_1, \dots, X_n) = (x_1, \dots, x_n)\},$$

pa je $L(\theta)$ vjerovatnoća da n -torka $(x_1, \dots, x_n)$ bude realizacija uzorka $(X_1, \dots, X_n)$ u slučaju kada je obilježje raspodijeljeno po zakonu koji određuje parametar θ .

Pretpostavimo da je obilježje X apsolutno neprekidno i da je dopustiva familija raspodjela određena sa $\{\varphi(x, \theta), \theta \in \Theta\}$. Funkcija

$$L(\theta) = L(\theta, x_1, \dots, x_n) = \varphi(x_1, \theta) \cdots \varphi(x_n, \theta), \quad \theta \in \Theta,$$

gdje je θ argument, a $x_1, \dots, x_n$ su parametri, naziva se **funkcija vjerodostojnosti**. Slijedeći vezu na relaciji vjerovatnoća – gustina, možemo konstatovati da je vrijednost $L(\theta, x_1, \dots, x_n)$ proporcionalna vjerovatnoći da

realizacija uzorka $(X_1, \dots, X_n)$ pripadne malom paralelopipedu koji sadrži tačku $(x_1, \dots, x_n)$, u slučaju kada je raspodjela obilježja X određena parametrom θ .

Pretpostavimo da smo dobili podatke koji formiraju n -torku $x_1, \dots, x_n$. Te podatke je generisao vjerovatnosni model u kojem je raspodjela obilježja X određena nekim fiksiranim $\theta \in \Theta$. Najrealnije je da su podaci stigli iz modela u kome je vjerovatnoća realizacije nama zabilježene n -torke najveća. Ovu tvrdnju baziramo na rasuđivanju čija će suština biti izložena u sljedećem primjeru.

Primjer 4. U kutiji se nalaze kuglice numerisane brojevima $1, \dots, N$; $N \in \{3, 4\}$. Iz kutije se vade četiri kuglice po modelu sa vraćanjem. Pretpostavimo da je registrovana četvorka $(1, 3, 1, 2)$. Na osnovu dobijenih brojeva tj. podataka, treba da ocijenimo nepoznato N . Praktično, treba da se odlučimo između dvije mogućnosti

1⁰ $(1, 3, 1, 2)$ potiče iz modela u kome je $N = 3$.

2⁰ $(1, 3, 1, 2)$ potiče iz modela u kome je $N = 4$.

Neka je A događaj da se izvuče četvorka $(1, 3, 1, 2)$. U slučaju kada je $N = 3$, imamo $P_3(A) = 1/3^4$. U slučaju kada je $N = 4$, imamo $P_4(A) = 1/4^4$. Kako je $P_3(A) > P_4(A)$, ocjenjujemo da četvorka $(1, 3, 1, 2)$ potiče iz modela u kome je $N = 3$, jer je u tom modelu vjerovatnoća registrovanja nama registrovane četvorke, veća. Znači, nepoznato N je ocijenjeno brojem 3. Mi nismo izričiti u tome da je $N = 3$, već se samo na osnovu zabilježenih podataka opredjeljujemo za model sa tri kuglice. Da je zabilježena, recimo četvorka $(2, 1, 4, 3)$, bilo bi jasno da se u kutiji nalaze četiri kuglice. ◀

Vjerovatnoća realizacije n -torke $(x_1, \dots, x_n)$ je ili jednaka (diskretni model) ili proporcionalna (apsolutno neprekidni model) sa $L(\theta, x_1, \dots, x_n)$ te za ocjenu nepoznatog parametra uzimamo ono θ iz Θ za koje je vrijednost funkcije $L(\theta, x_1, \dots, x_n)$ maksimalna. Na ovaj način iz skupa kandidata biramo ono θ za koje je vjerovatnoća nama zabilježene realizacije $(x_1, \dots, x_n)$ najveća.

Postupak traženja ocjene je sljedeći. Potražimo ono θ za koje funkcija $L(\theta, x_1, \dots, x_n)$ ($\theta \in \Theta$ je argument, $x_1, \dots, x_n$ su parametri) ima maksimum. Neka je dobijeno $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$. Za statistiku $\hat{\theta}(X_1, \dots, X_n)$ kažemo da je ocjena parametra θ dobijena **metodom maksimalne vjerodostojnosti**.

Pošto je funkcija $\ln x$, $x > 0$, monotono rastuća, problem traženja maksimuma funkcije $L(\theta, x_1, \dots, x_n)$ je ekvivalentan sa problemom traženja maksimuma funkcije $\ln L(\theta, x_1, \dots, x_n)$. Ova činjenica se često koristi u situacijama kada metodom maksimalne vjerodostojnosti treba naći ocjenu nepoznatog parametra.

Metod maksimalne vjerodostojnosti se primjenjuje i u slučaju kada raspodjela obilježja zavisi od dva i više parametara. Formira se funkcija vjerodostojnosti, a zatim se metodima višedimenzionalne analize nađe tačka u kojoj ta funkcija dostiže maksimum. Navešćemo jedan rezultat koji se dobija u okviru te teorije. Ako obilježje X ima raspodjelu koja pripada familiji $\{\mathcal{N}(m, \sigma^2), m \in R, \sigma^2 > 0\}$, m i σ^2 su nepoznati parametri, metod maksimalne vjerodostojnosti za ocjenu nepoznatog m predlaže statistiku $\bar{X}_n$, a za ocjenu nepoznatog σ^2 statistiku $\bar{S}_n^2$.

Primjer 5. Obilježje X ima raspodjelu koja pripada familiji dopustivih raspodjela $\mathcal{P} = \{\mathcal{U}(0, \theta), \theta > 0\}$. Koristeći uzorak obima n , metodom maksimalne vjerodostojnosti naći ocjenu nepoznatog parametra θ .

► Imamo da je $\varphi(x, \theta) = \begin{cases} 1/\theta, & x \in (0, \theta) \\ 0, & x \notin (0, \theta) \end{cases}$, te je

$$L(\theta, x_1, \dots, x_n) = \begin{cases} \frac{1}{\theta^n}, & x_1, \dots, x_n \in (0, \theta) \\ 0, & \text{inače} \end{cases} = \theta^{-n} I\{\theta > \max_{1 \leq k \leq n} x_k\}.$$

Koristimo uobičajenu oznaku za karakterističnu funkciju skupa. Funkcija $L(\theta)$ ima maksimum za $\hat{\theta} = \max_{1 \leq k \leq n} x_k$. Tražena ocjena je statistika $Y_n = \max_{1 \leq k \leq n} X_k$. Ovim smo zadatak riješili na opštem nivou. Kada u konkretnoj situaciji registrujemo n podataka – brojeva, treba da potražimo najveći među njima i da ga proglasimo za ocjenu nepoznatog parametra θ . ◀

Primjer 6. Obilježje X ima raspodjelu koja pripada familiji dopustivih raspodjela $\mathcal{P} = \{\mathcal{U}(\theta, 1), \theta \in (0, 1)\}$. Koristeći uzorak obima n , metodom maksimalne vjerodostojnosti naći ocjenu nepoznatog parametra θ .

► Imamo da je $\varphi(x, \theta) = \begin{cases} 1/(1 - \theta), & x \in (\theta, 1) \\ 0, & x \notin (\theta, 1) \end{cases}$, te je

$$L(\theta, x_1, \dots, x_n) = \frac{1}{(1 - \theta)^n} I\{0 < \theta < \min_{1 \leq k \leq n} x_k\}.$$

Znači, tražena ocjena je statistika $Y_1 = \min_{1 \leq k \leq n} X_k$. ◀

Primjer 7. Obilježje X ima raspodjelu koja pripada familiji dopustivih raspodjela $\mathcal{P} = \{\mathcal{P}(\lambda), \lambda > 0\}$. Koristeći uzorak obima n , metodom maksimalne vjerodostojnosti naći ocjenu nepoznatog parametra λ .

► Kod Puasonove raspodjele imamo $p(x_\nu, \lambda) = \frac{\lambda^{x_\nu}}{x_\nu!} e^{-\lambda}$, $x_\nu \in \{0, 1, \dots\}$. Potražimo logaritam funkcije vjerodostojnosti. Imamo

$$\ln L(\lambda, x_1, \dots, x_n) = \ln \frac{\lambda^{\sum_{k=1}^n x_k} e^{-n\lambda}}{\prod_{k=1}^n x_k!}.$$

Nakon rješavanja jednačine $\frac{\partial}{\partial \lambda} \ln L(\lambda, x_1, \dots, x_n) = 0$, dobijamo da je $\bar{X}_n$ tražena statistika. ◀

Primjer 8. U kutiji se nalazi N kuglica numerisanih brojevima od 1 do N , broj N je nepoznat. Iz kutije se po modelu sa vraćanjem vadi n kuglica. Neka su $x_1, \dots, x_n$ brojevi na izvađenim kuglicama. Metodom maksimalne vjerodostojnosti ocijeniti nepoznati broj N .

► Neka je obilježje X broj na izvađenoj kuglici. Imamo $p(x_\nu, N) = \frac{1}{N}$, $x_\nu \in \{1, \dots, N\}$ i

$$L(N, x_1, \dots, x_n) = \frac{1}{N^n} I\{N \geq \max_{1 \leq k \leq n} x_k\},$$

te je tražena ocjena $\max_{1 \leq k \leq n} x_k$. Statistika – ocjena koju ovaj rezultat određuje je $Y_n = \max_{1 \leq k \leq n} X_k$. Dakle, kada ne znamo koliko kuglica postoji u kutiji, metod maksimalne vjerodostojnosti sugerise da nepoznati broj kuglica ocijenimo sa najvećim od izvučenih brojeva. ◀

Zadaci

1. Obilježje X ima gustinu $\varphi(x) = e^{c-x}$, $x > c$, c je nepoznati parametar. Ispitati centriranost i postojanost ocjene $c^* = Y_1 - \frac{1}{n}$.

2. Obilježje X je ravnomjerno raspodijeljeno na nepoznatom intervalu. Za ocjenu sredine intervala predlažu se statistike $a_1 = \bar{X}_n$ i $a_2 = \frac{Y_1 + Y_n}{2}$. Pokazati da su obje ocjene centrirane. Koja je ocjena bolja?

3. Strijelac pogađa cilj sa nepoznatom vjerovatnoćom p , $0 < p < 1$. Da bi se odredilo p , strijelac dobija 20 metaka i ispaljuje ih u cilj. Izvedeno je 5 serija (po 20 metaka u seriji) i pri tom su dobijeni sljedeći rezultati:

serija	1	2	3	4	5
broj pogodaka	14	10	12	9	15

Metodom maksimalne vjerodostojnosti ocijeniti nepoznato p .

4. Obilježje X ima raspodjelu

$$X : \begin{matrix} -2 & 0 & 7 \\ \theta/5 & \theta/5 & 1 - 2\theta/5 \end{matrix}, \quad 0 < \theta < \frac{5}{2}.$$

Na osnovu uzorka obima n , metodom maksimalne vjerodostojnosti, naći ocjenu za θ . Naći ocjenu na osnovu realizovanog uzorka $(0, -2, 7, -2)$.

5. U ribnjaku se nalazi N riba, N je nepoznato. Mrežom se izvadi n riba, one se markiraju i vrata u ribnjak. Nakon izvjesnog vremena se ponovo vadi n riba i konstatuje da među njima ima r markiranih. Koristeći ideju iz metoda maksimalne vjerodostojnosti, ocijeniti nepoznati broj N .

6. Obilježje X ima raspodjelu koja pripada familiji

$$\mathcal{P} = \{\mathcal{N}(\theta, 2\theta), \theta > 0\}.$$

Primjenom metoda maksimalne vjerodostojnosti ocijeniti nepoznati parametar θ . Dokazati da je dobijena ocjena postojana.

7. Obilježje X ima raspodjelu koja pripada familiji dopustivih raspodjela

- a) $\mathcal{P} = \{\mathcal{E}(\theta), \theta > 0\}$.
- b) $\mathcal{P} = \{\mathcal{N}(\theta, 2\theta), \theta > 0\}$.
- c) $\mathcal{P} = \{\mathcal{U}(-\theta, \theta), \theta > 0\}$.
- d) $\mathcal{P} = \{\mathcal{U}(\theta, \theta + 2), \theta \in \mathbb{R}\}$.
- e) $\mathcal{P} = \{\mathcal{U}(\theta, 2\theta), \theta > 0\}$.
- f) $\mathcal{P} = \{\mathcal{U}(1/\theta, \theta), \theta > 1\}$.
- g) $\mathcal{P} = \{\varphi(x, \theta) = \exp\{-(x - \theta)\}, x > \theta\}$.

Koristeći uzorak obima n , metodom maksimalne vjerodostojnosti naći ocjenu nepoznatog parametra θ .

2.3 Intervali povjerenja

Neka obilježje X ima funkciju raspodjele vjerovatnoća koja pripada familiji $\{F(x, \theta), \theta \in \Theta\}$. Kod tačkastog ocjenjivanja, formirana je statistika–ocjena $W_n(X_1, \dots, X_n)$ i njena realizacija $W_n(x_1, \dots, x_n)$ je uzimana za ocjenu nepoznatog parametra θ . Kod intervalnog ocjenjivanja nepoznatog parametra, formira se interval sa slučajnim krajevima. Taj interval sa velikom vjerovatnoćom (mi je biramo) sadrži nepoznati parametar.

DEFINICIJA 2.3.1 *Neka su $V_n = V_n(X_1, \dots, X_n)$ i $U_n = U_n(X_1, \dots, X_n)$ dvije statistike takve da za svako $\theta \in \Theta$ važi:*

- a) $P_\theta\{V_n \leq U_n\} = 1$,
- b) $P_\theta\{V_n < \theta < U_n\} = \beta$.

Tada se interval (V_n, U_n) naziva interval povjerenja za nepoznati parametar θ , a broj β se naziva nivo povjerenja.

Sa P_θ označavamo vjerovatnoću u modelu u kojem je stvarna vrijednost parametra jednaka θ .

Prirodno je težiti ka tome da interval povjerenja bude što kraći, a nivo povjerenja što veći. U opštem slučaju ovi zahtjevi su oprečni. Za nivo povjerenja se obično uzimaju brojevi $\beta = 0,9$, $\beta = 0,95$, $\beta = 0,99$. Interval povjerenja je slučajni interval koji može i da ne sadrži nepoznati parametar.

Intervalnoj ocjeni možemo dati sljedeće tumačenje. Pretpostavimo da smo N puta (N je veliko) sproveli proceduru registrovanja vrijednosti obilježja kod n elemenata populacije. Neka su $(v_n^{(1)}, u_n^{(1)}), \dots, (v_n^{(N)}, u_n^{(N)})$ intervali dobijeni nakon traženja realizacija statistika V_n i U_n na osnovu dobijenih realizacija uzorka. Polazeći od interpretacije vjerovatnoće koja je data kroz zakon velikih brojeva, jednakosti b) iz definicije možemo dati sljedeći smisao. Približno $100\beta\%$ intervala će sadržati nepoznati parametar θ , a $100(1 - \beta)\%$ intervala neće sadržati nepoznati parametar θ . Da bi ovo tumačenje imalo valjani smisao, serija mora da bude velika. Bilo bi pogrešno nakon n registrovanja podataka i dobijenih realizacija v_n i u_n statistika V_n i U_n , reći da je $P\{v_n < \theta < u_n\} = \beta$. Posljednja jednakost je besmislena jer se u velikoj zagradi ne nalazi događaj (svi parametri su konstante).

Primjer 1. Obilježje X ima raspodjelu koja pripada familiji zadatoj gustinama

$$\Phi = \{\varphi(x, \theta) = \exp\{-(x - \theta)\}, x > \theta, \theta > 0\}.$$

Dokazati da je $(Y_1 + n^{-1} \ln(1 - \beta), Y_1)$, β interval povjerenja za nepoznati parametar θ .

►Potražimo,

$$\begin{aligned} F_{Y_1}(x) &= P\{\min_{1 \leq k \leq n} X_k < x\} = 1 - P\{\min_{1 \leq k \leq n} X_k \geq x\} \\ &= 1 - \left(\int_x^\infty \exp\{-(t - \theta)\} dt \right)^n = 1 - \exp\{-n(x - \theta)\}, x > \theta, \end{aligned}$$

pa je $\varphi_{Y_1}(x) = n \exp\{-n(x - \theta)\}$, $x > \theta$. Provjerimo,

$$\begin{aligned} P_\theta\{Y_1 + n^{-1} \ln(1 - \beta) < \theta < Y_1\} &= P_\theta\{\theta < Y_1 < \theta - n^{-1} \ln(1 - \beta)\} \\ &= \int_{\theta}^{\theta - n^{-1} \ln(1 - \beta)} n e^{-n(x - \theta)} dx = \beta. \blacktriangleleft \end{aligned}$$

Primjer 2. Obilježje X ima raspodjelu koja pripada familiji $\mathcal{P} = \{\mathcal{U}(0, \theta), \theta > 0\}$. Za nepoznati parametar θ , naći β interval povjerenja oblika

$$(aY_n, 2aY_n), a > 1.$$

►Gustina statistike Y_n je $\varphi(x, \theta) = nx^{n-1}/\theta^n$, $x \in (0, \theta)$ (primjer 8 iz odjeljka 1.9). Imamo,

$$\begin{aligned}
P_\theta \{aY_n < \theta < 2aY_n\} &= P_\theta \left\{ \frac{\theta}{2a} < Y_n < \frac{\theta}{a} \right\} = \int_{\frac{\theta}{2a}}^{\frac{\theta}{a}} n \frac{x^{n-1}}{\theta^n} dx \\
&= \frac{2^n - 1}{2^n a^n} = \beta \implies a = \frac{1}{2} \left(\frac{2^n - 1}{\beta} \right)^{\frac{1}{n}}. \blacktriangleleft
\end{aligned}$$

Primjer 3. Obilježje X ima raspodjelu koja pripada familiji

$$\mathcal{P} = \{\mathcal{E}(\theta), \theta > 0\}.$$

Naći 90% ($\beta = 0,9$) interval povjerenja za nepoznati parametar θ . Koristiti uzorak velikog obima n .

► Zbog eksponencijalnosti raspodjele obilježja X imamo: $EX = 1/\theta$ i $DX = 1/\theta^2$. Neka je $(X_1, \dots, X_n)$ uzorak obilježja X i neka je $S_n = \sum_{k=1}^n X_k$. Pošto je n veliki broj, na osnovu CGT, možemo konstatovati da statistika

$$\frac{S_n - n\theta}{\sqrt{\frac{n}{\theta^2}}} = \sqrt{n}(\theta\bar{X}_n - 1),$$

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

$$\begin{aligned}
P_\theta \{-z_{0,45} < (\theta\bar{X}_n - 1)\sqrt{n} < z_{0,45}\} &= 0,9 \iff \\
P_\theta \left\{ \left(1 - \frac{z_{0,45}}{\sqrt{n}}\right) \bar{X}_n^{-1} < \theta < \left(1 + \frac{z_{0,45}}{\sqrt{n}}\right) \bar{X}_n^{-1} \right\}.
\end{aligned}$$

Iz tablica čitamo da je $z_{0,45} = 1,65$. Dakle, traženi interval je

$$\left(\left(1 - \frac{1,65}{\sqrt{n}}\right) \bar{X}_n^{-1}, \left(1 + \frac{1,65}{\sqrt{n}}\right) \bar{X}_n^{-1} \right). \blacktriangleleft$$

Zbog velike rasprostranjenosti normalne raspodjele, važno je imati intervale povjerenja za njene parametre.

Primjer 4. Obilježje X ima raspodjelu koja pripada familiji

$$\mathcal{P} = \{\mathcal{N}(m, 1), m \in \mathbb{R}\}.$$

Naći β interval povjerenja za nepoznati parametar m .

► Na osnovu Fišerove teoreme imamo $(\bar{X}_n - m)\sqrt{n} : \mathcal{N}(0, 1)$, pa je

$$\begin{aligned}
P_m \{-z_{\beta/2} < (\bar{X}_n - m)\sqrt{n} < z_{\beta/2}\} &= \beta \iff \\
P_m \left\{ \bar{X}_n - \frac{z_{\beta/2}}{\sqrt{n}} < m < \bar{X}_n + \frac{z_{\beta/2}}{\sqrt{n}} \right\} &= \beta
\end{aligned}$$

Dakle, $\left(\bar{X}_n - \frac{z_{\beta/2}}{\sqrt{n}}, \bar{X}_n + \frac{z_{\beta/2}}{\sqrt{n}}\right)$ je traženi interval. ◀

Primjer 5. Obilježje X ima raspodjelu koja pripada familiji

$$\mathcal{P}\{\mathcal{N}(m, \sigma^2), m \in R, \sigma^2 > 0\}$$

(oba parametra su nepoznata). Naći β interval povjerenja za parametar m . Takođe, naći β interval povjerenja za parametar σ^2 .

► Statistika $\frac{\bar{X}_n - m}{\bar{S}_n} \sqrt{n-1}$, na osnovu Fišerove teoreme, ima t_{n-1} raspodjelu, pa je

$$P_m \left\{ -v_{n-1, \beta/2} < \frac{\bar{X}_n - m}{\bar{S}_n} \sqrt{n-1} < v_{n-1, \beta/2} \right\} = \beta \iff$$

$$P_m \left\{ \bar{X}_n - \frac{v_{n-1, \beta/2}}{\sqrt{n-1}} \bar{S}_n < m < \bar{X}_n + \frac{v_{n-1, \beta/2}}{\sqrt{n-1}} \bar{S}_n \right\} = \beta.$$

Dakle, $\left(\bar{X}_n - \frac{v_{n-1, \beta/2}}{\sqrt{n-1}} \bar{S}_n, \bar{X}_n + \frac{v_{n-1, \beta/2}}{\sqrt{n-1}} \bar{S}_n\right)$ je traženi interval povjerenja za nepoznati parametar m .

Statistika $\frac{n\bar{S}_n^2}{\sigma^2}$, na osnovu Fišerove teoreme, ima χ_{n-1}^2 raspodjelu, pa je

$$P_{\sigma^2} \left\{ \frac{n\bar{S}_n^2}{\sigma^2} > h_{n-1, 1-\beta} \right\} = \beta \iff P_{\sigma^2} \left\{ 0 < \sigma^2 < nh_{n-1, 1-\beta}^{-1} \bar{S}_n^2 \right\} = \beta,$$

te je $\left(0, nh_{n-1, 1-\beta}^{-1} \bar{S}_n^2\right)$ traženi interval. Primijetimo da je samo jedan kraj intervala, statistika – slučajna veličina. Ovakvi intervali se zovu jednostrani. Intervali kod kojih su oba kraja statistike zovu se dvostrani.

Da smo pošli od jednakosti

$$P_{\sigma^2} \left\{ h_{n-1, \frac{1-\beta}{2}} < \frac{n\bar{S}_n^2}{\sigma^2} < h_{n-1, \frac{1+\beta}{2}} \right\} = \beta,$$

nakon rješavanja po σ^2 dvostruke nejednakosti iz velike zagrade, dobili bismo neki drugi interval povjerenja. Savjetujemo čitaocu da naznačeno i uradi. Dobijeni interval će biti dvostrani. Interval povjerenja za nepoznati parametar, u opštem slučaju nije jedinstven. U slučajevima kada postoji više intervala povjerenja, prirodno je tražiti onaj koji ima najmanju dužinu. Pos-toje razrađeni metodi kojima se rješava ovaj zadatak. ◀

Zadaci

1. Obilježje X ima raspodjelu koja pripada familiji

$$\mathcal{P} = \{\mathcal{U}(0, 2\theta), \theta > 0\}.$$

Koristeći uzorak velikog obima n , primjenom CGT, naći 90% interval povjerenja za nepoznati parametar θ .

2. Obilježje X ima raspodjelu koja pripada familiji

$$\mathcal{P} = \{\mathcal{N}(\theta, \theta^2), \theta > 0\}.$$

Naći β interval povjerenja za nepoznati parametar θ . Koristiti uzorak obima n .

3. Obilježje X ima raspodjelu koja pripada familiji

$$\mathcal{P} = \{\mathcal{N}(m; (0, 1)^2), m \in R\}.$$

Odrediti najmanji obim uzorka, tako da dužina 95% intervala povjerenja za nepoznato m ne bude veća od 0,05.

4. Uzorak obima 10, slučajno uzet iz populacije čije obilježje ima normalnu raspodjelu, daje podatke za koje važi: $\sum_{k=1}^{10} x_k = 34$, $\sum_{k=1}^{10} x_k^2 = 138$.

Nakon izvjesnog vremena je uzet dopunski uzorak obima 5 i tom su prilikom registrovani podaci: 2, 8; 3, 5; 3, 8; 3, 2; 3, 4. Za koliko se promijenila dužina 90% intervala povjerenja za nepoznato matematičko očekivanje.

5. Prilikom 7 kvarova jedne mašine, izmjereni su sljedeći brojevi časova rada do kvara: 53; 48; 50; 54; 51; 50; 51. Naći 99% interval povjerenja za srednji broj časova rada mašine do kvara (prirodno je smatrati da je vrijeme rada mašine raspodijeljeno po normalnom zakonu).

6. Dokazati da je interval koji smo dobili u primjeru 4, najkraći među intervalima za nepoznati parametar m .

7. Prilikom određivanja naelektrisanja elektrona $e_0 10^{-10}$, Miliken je dobio $n = 58$ nezavisnih rezultata mjerenja X_i veličine e_0 (mjerenje je obavljeno u cgs jedinicama). Za uzoračku sredinu i uzoračku disperziju je dobijeno

$$\overline{X}_{58} = 4,7808 \text{ i } \overline{S}_{58}^2 = 22981 \cdot 10^{-8}.$$

Polazeći od toga da je greška normalno raspodijeljena i da ne postoji sistematska greška u mjerenju, naći $\beta = 0,9$ interval povjerenja za nepoznato e_0 .

2.4 Testiranje statističkih hipoteza

Pojam statističke hipoteze i ideju testiranja približićemo kroz jedan primjer.

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

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

B) 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):

1⁰ Kutija je napunjena po strategiji A – govorimo o hipotezi H_A .

2⁰ Kutija je napunjena po strategiji B – govorimo o hipotezi H_B .

Postupak presuđivanja u korist jedne od hipoteza zvaćemo **testom**. Ako je kutija napunjena po strategiji A, tada je zbog velike vjerovatnoće pojave broja 1, vjerovatnoća da se registruje mala suma brojeva, velika. Ako je kutija napunjena po strategiji B, 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 ≥ 7 prihvaćemo H_B , a H_A odbaciti. U suprotnom tj. ako je zbir izvučenih brojeva < 7 prihvaćemo H_A , a H_B odbaciti.

Označimo sa X obilježje koje predstavlja broj na izvučenoj kuglici. U slučaju kada je važeća strategija A, 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 B, 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.

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

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

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_B , a ako "upadne" u C^c , tada prihvatamo H_A . Očigledno, oblast C određuje postupak – pravilo odlučivanja tj. test. "Upadanje" uzorka u oblast C ne odgovara hipotezi H_A (tada imamo onu situaciju kada je zbir izvučenih brojeva veliki tj. među brojevima dominira 2). Zbog toga oblast C nazivamo kritična oblast za H_A .

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_A koja je tačna i samim tim prihvatimo H_B koja je netačna. Jasno, moguća je i situacija u kojoj H_A i H_B imaju zamijenjene uloge.

Izračunajmo vjerovatnoću α da odbacimo tačnu hipotezu H_A .

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

Izračunajmo vjerovatnoću β da odbacimo tačnu hipotezu H_B .

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

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

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

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

Motivisani prethodnim primjerom, možemo preći na izlaganje teorije.

Neka je X obilježje čija funkcija raspodjele vjerovatnoća pripada familiji

$$\mathcal{F}_X = \{F(x, \theta), \theta \in \Theta\}.$$

Pretpostavka oblika $H_o(\theta \in \Theta_o)$, $\Theta_o \subset \Theta$, zove se **statistička hipoteza**. Malo drugačije rečeno, pretpostavka se sastoji u tome da obilježje X ima raspodjelu koja pripada užoj familiji određenoj parametarskim skupom Θ_o .

Ako je Θ_o jednočlani skup, tada se hipoteza H_o naziva prosta. U protivnom govorimo o složenoj hipotezi. Hipotezi H_o ćemo suprotstaviti hipotezu H_1 ($\theta \in \Theta \setminus \Theta_o$). Hipoteza H_o se naziva nulta, a hipoteza H_1 alternativna. Postupak odlučivanja u korist jedne hipoteze se naziva **statistički test**.

Na početku testiranja se zada broj α koji se naziva **prag značajnosti testa**. Obično je $\alpha = 0,01$, $\alpha = 0,02$, $\alpha = 0,05$, $\alpha = 0,1$. Ako je $(X_1, \dots, X_n)$ uzorak obilježja X , tada se skup $C \subset R^n$, za koji važi

$$\alpha = P_{H_o}\{(X_1, \dots, X_n) \in C\}, \quad (1)$$

naziva **kritična oblast veličine α za testiranje hipoteze H_o** . Odlučuje se po sljedećem kriterijumu:

Ako realizovani uzorak $(x_1, \dots, x_n)$ pripada C ("upadne" u C) tada H_o odbacujemo u korist H_1 i to radimo sa pragom značajnosti α . Ako realizovani uzorak $(x_1, \dots, x_n)$ pripada C^c ("upadne" u C^c) tada H_1 odbacujemo u korist H_o i to radimo sa pragom značajnosti α .

Iz izloženog se vidi da je test (pravilo odlučivanja) određen kritičnom oblašću i zato se kaže da kritična oblast zadaje test. Kritična oblast veoma često ima neki od sljedeća tri oblika:

$$\begin{aligned} C_1 &= \{(x_1, \dots, x_n) : T_1(x_1, \dots, x_n) \geq k_1\}, \\ C_2 &= \{(x_1, \dots, x_n) : T_2(x_1, \dots, x_n) \leq k_2\}, \\ C_3 &= \{(x_1, \dots, x_n) : |T_3(x_1, \dots, x_n)| \geq k_3\}, \end{aligned}$$

gdje su $T_1, T_2, T_3 : R^n \rightarrow R$ Borelova preslikavanja. U slučaju kada je test zadat recimo sa C_1 , jednakost (??) se svodi na

$$\alpha = P_{H_o}\{T_1(X_1, \dots, X_n) \geq k_1\}.$$

Statistika T_1 se naziva test statistika.

Prilikom odlučivanja, postoji mogućnost da se napravi greška.

1° **Grešku prve vrste** pravimo kada odbacimo faktički tačnu hipotezu H_o .

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

Već pomenuto α predstavlja **vjerovatnoću greške prve vrste**. **Vjerovatnoća greške druge vrste** se označava sa β i za nju, na osnovu rečenog, važi

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

Prirodna je potreba da se pronađe test u kome su brojevi α i β mali. Kod rješavanja ovog zadatka poteškoće izviru iz činjenice da smanjivanje jednog od ova dva parametra povlači uvećavanje drugog.

Neka je C kritična oblast koja određuje neki test za testiranje $H_o(\theta \in \Theta_o)$ protiv $H_1(\theta \in \Theta \setminus \Theta_o)$. **Funkcija moći testa** se zadaje sa

$$M(\theta) = P_\theta\{(X_1, \dots, X_n) \in C\}, \theta \in \Theta.$$

Sa ciljem da čitalac potpuno usvoji ideju testiranja, navodimo nekoliko primjera.

Primjer 2. Na osnovu ranijih rezultata, vjerovatnoća da strijelac A pogodi cilj je procijenjena na 0,9. Nakon pripreme za takmičenje, izvršeno je probno gađanje u kome je strijelac imao 92 pogotka od 100 gađanja. Sa pragom značajnosti $\alpha = 0,05$, testirati hipotezu da je vjerovatnoća pogotka ostala ista protiv alternativne da je povećana.

► Neka je $(X_1, \dots, X_n)$ uzorak obilježja X koje ima raspodjelu

$$X : \begin{matrix} 0 & 1 \\ q & p \end{matrix}, p \text{ je vjerovatnoća pogotka.}$$

Statistika $S_{100} = \sum_{k=1}^{100} X_k$ predstavlja broj pogodaka u seriji od 100 gađanja i ona ima $\mathcal{B}(100, p)$ raspodjelu. Neka je $H_o(p = 0,9)$ hipoteza da je vjerovatnoća pogotka 0,9, a $H_1(p > 0,9)$ hipoteza da je vjerovatnoća pogotka veća od 0,9. Za test statistiku ćemo uzeti

$$W_{100} = \frac{S_{100} - 90}{\sqrt{100 \cdot 0,9 \cdot 0,1}}$$

i ova statistika, na osnovu CGT, u slučaju kada je tačna hipoteza H_o ima približno $\mathcal{N}(0,1)$ raspodjelu. U zoni hipoteze H_1 imamo $ES_{100} > 90$, te povoljni slučajevi za H_1 nastaju kada test statistika uzima pozitivne vrijednosti. Zbog rečenog ćemo za kritičnu oblast uzeti

$$C = \left\{ (x_1, \dots, x_{100}) : w_{100} = \frac{s_{100} - 90}{3} > k \right\}, k > 0.$$

Odredimo k .

$$0,05 = P_{H_o}\{W_{100} > k\} \implies k = 1,65.$$

Kod nas je $w_{100} = 0,66$. Ne nalazimo se u kritičnoj oblasti te sa pragom značajnosti 0,05 prihvatamo hipotezu H_o . Predlažemo čitaocu da hipotezi

H_o suprotstavi hipotezu $H_1(p \neq 0,9)$, tj. hipotezu da je došlo do promjene vjerovatnoće pogotka. I u ovom slučaju treba koristiti test statistiku W_{100} , a za kritičnu oblast uzeti $C^* = \{(x_1, \dots, x_{100}) : |w_{100}| > k^*\}$. ◀

Primjer 3. Ispitivanja koja su u prošlosti obavljena kod jedne sociološke grupe pokazuju da obilježje X – broj djece u porodici ima približno $\mathcal{P}(1,8)$ raspodjelu. Postoji pretpostavka da je u naše vrijeme došlo do smanjenja broja djece u porodici. Da bi se ta pretpostavka provjerila, uzet je uzorak od 100 porodica i na njemu je registrovano da je srednji broj djece $\bar{x}_{100} = 1,5$. Da li se sa pragom značajnosti $\alpha = 0,02$ može prihvatiti hipoteza da je očekivani broj djece u porodici manji u odnosu na raniji period.

► Za familiju dopustivih raspodjela obilježja X – broj djece u porodici ćemo uzeti $\mathcal{P} = \{\mathcal{P}(\lambda), \lambda > 0\}$. Dobro je poznato da je $EX = \lambda$, $DX = \lambda$. Treba testirati $H_o(\lambda = 1,8)$ (obilježje X ima $\mathcal{P}(1,8)$ raspodjelu) protiv $H_1(\lambda < 1,8)$ (obilježje X ima $\mathcal{P}(\lambda)$ raspodjelu u kojoj je $\lambda < 1,8$). Za test statistiku ćemo uzeti

$$W_{100} = \frac{S_{100} - 100 \cdot 1,8}{\sqrt{100 \cdot 1,8}} = \frac{\bar{X}_{100} - 1,8}{\sqrt{1,8}} \cdot 10.$$

Statistika W_{100} , u slučaju kada je H_o tačna, na osnovu CGT, ima približno $\mathcal{N}(0,1)$ raspodjelu. U zoni hipoteze H_1 očekivanje statistike W_{100} je negativno te je razumno za kritičnu oblast uzeti

$$C = \{(x_1, \dots, x_{100}) : w_{100} < k\}, k < 0.$$

Potražimo k .

$$0,02 = P_{H_o}\{W_{100} < k\} \implies k = -2,05.$$

Kod nas je $w_{100} = -2,24$. Realizovani uzorak se nalazi u kritičnoj oblasti te hipotezu H_1 o smanjenju očekivanog broja djece u porodici prihvatamo sa pragom značajnosti $\alpha = 0,02$ ◀

Primjer 4. U kutiji se nalaze kuglice bijele i crne boje. O sastavu imamo još samo informaciju da bijelih kuglica ima ili 40% ili 50%. Iz kutije je po modelu sa vraćanjem izvučeno 100 kuglica i konstatovano da među njima ima više bijelih kuglica. Na osnovu toga je izveden zaključak da bijelih kuglica ima 50% tj. pola od ukupnog broja. Kolika je vjerovatnoća da je izveden pogrešan zaključak?

► Neka obilježje X uzima vrijednost 0 ako je izvučena kuglica crne boje, a vrijednost 1 ako je izvučena kuglica bijele boje. U našem modelu raspodjela obilježja X pripada familiji raspodjela

$$X : \begin{matrix} 0 & 1 \\ q & p \end{matrix}, p \in \{0,4; 0,5\}, p + q = 1.$$

Neka je $(X_1, \dots, X_{100})$ uzorak kojim se opisuje 100 izvađenih kuglica i neka je $S_{100} = \sum_{k=1}^{100} X_k$.

Tražena vjerovatnoća je

$$p = P_{0,4}\{S_{100} > 50\} = P_{0,4}\left\{\frac{S_{100} - 100 \cdot 0,4}{\sqrt{100 \cdot 0,4 \cdot 0,6}} > \frac{50 - 40}{\sqrt{24}}\right\} \\ = 0,5 - \Phi^*(2,04) = 0,021. \blacktriangleleft$$

Primjer 5. Obilježje X ima raspodjelu koja pripada familiji

$$\mathcal{P} = \{\mathcal{N}(\theta, 1), \theta \in \{\theta_0, \theta_1\}\}, \theta_0 < \theta_1.$$

Testiraćemo hipotezu $H_0(\theta = \theta_0)$ protiv $H_1(\theta = \theta_1)$. Statistika $\bar{X}_n$, u slučaju kada je tačna hipoteza H_0 , ima $\mathcal{N}(\theta_0, \frac{1}{n})$ raspodjelu te se njene vrijednosti koncentrišu oko θ_0 . Statistika $\bar{X}_n$, u slučaju kada je tačna hipoteza H_1 , ima $\mathcal{N}(\theta_1, \frac{1}{n})$ raspodjelu te se njene vrijednosti koncentrišu oko θ_1 . Stoga je razumno za kritičnu oblast uzeti

$$C = \{(x_1, \dots, x_n) : \bar{x}_n > c\},$$

gdje je c neki fiksirani broj sa (θ_0, θ_1) .

Slika 20: Primjer 5.

Na slici ??, vjerovatnoća greške prve vrste α je jednaka površini tamnije osjenčenog dijela, a vjerovatnoća greške druge vrste β je jednaka površini blaže osjenčenog dijela. Pomjeranjem broja c , jedan od dijelova se uvećava, a drugi smanjuje. $\blacktriangleleft$

Primjer 6. Obilježje X ima raspodjelu koja pripada familiji

$$\mathcal{P} = \{\mathcal{N}(\theta, 1), \theta \in R\}.$$

Sa pragom značajnosti α testirati hipotezu $H_0(\theta = 0)$ protiv $H_1(\theta \neq 0)$.

► Test statistika $\sqrt{n} \bar{X}_n$, u slučaju kada je tačna hipoteza H_o , ima $\mathcal{N}(0, 1)$ raspodjelu. Zato je

$$\alpha = P_{H_o} \{ |\sqrt{n} \bar{X}_n| > z_{(1-\alpha)/2} \},$$

pa je kritična oblast $C = \{(x_1, \dots, x_n) : |\sqrt{n} \bar{x}_n| > z_{(1-\alpha)/2}\}$.

Preporučujemo čitaocu da formira testove za alternative $H_1(\theta > 0)$ i $H_1(\theta < 0)$. ◀

Primjer 7. Obilježje X ima raspodjelu koja pripada familiji

$$\mathcal{P} = \{\mathcal{N}(\theta, \sigma^2), \theta \in R, \sigma^2 > 0\},$$

oba parametra su nepoznata. Sa pragom značajnosti α , testirati hipotezu $H_o(\theta = \theta_0)$ protiv $H_1(\theta \neq \theta_0)$.

► Test statistika $\frac{\bar{X}_n - \theta_0}{\bar{S}_n} \sqrt{n-1}$, u slučaju kada je tačna hipoteza H_o , ima t_{n-1} raspodjelu. Zato je

$$\alpha = P_{H_o} \left\{ \left| \frac{\bar{X}_n - \theta_0}{\bar{S}_n} \sqrt{n-1} \right| > v_{n-1, (1-\alpha)/2} \right\},$$

pa je kritična oblast

$$C = \left\{ (x_1, \dots, x_n) : \left| \frac{\bar{x}_n - \theta_0}{\bar{s}_n} \sqrt{n-1} \right| > v_{n-1, (1-\alpha)/2} \right\}. \blacktriangleleft$$

Primjer 8. Obilježje X ima raspodjelu koja pripada familiji dopustivih raspodjela

$$\mathcal{P} = \{\mathcal{U}(0, \theta), \theta > 0\}.$$

Za testiranje hipoteze $H_o(\theta = \theta_0)$, $\theta_0 > 0$, θ_0 je fiksirani broj, protiv alternativne $H_1(\theta \neq \theta_0)$, koristi se kritična oblast

$$C = \left\{ (x_1, \dots, x_n) : \max_{1 \leq k \leq n} x_k > \theta_0 \text{ ili } \max_{1 \leq k \leq n} x_k < c \right\}, c < \theta_0.$$

Za dati prag značajnosti α , odrediti konstantu c . Naći funkciju moći testa $M(\theta)$, $\theta > 0$.

$$\begin{aligned} \blacktriangleright \alpha &= P_{\theta_0} \{(X_1, \dots, X_n) \in C\} = P_{\theta_0} \left\{ \max_{1 \leq k \leq n} X_k > \theta_0 \text{ ili } \max_{1 \leq k \leq n} X_k < c \right\} \\ &= P_{\theta_0} \left\{ \max_{1 \leq k \leq n} X_k < c \right\} = \left(\frac{c}{\theta_0} \right)^n = \alpha \implies c = \sqrt[n]{\alpha} \theta_0. \end{aligned}$$

Prilikom traženja funkcije moći testa, razlikovaćemo tri slučaja.

$$\begin{aligned}
 1^o \quad M(\theta) &= P_\theta \left\{ \max_{1 \leq k \leq n} X_k < c \right\} = 1, \quad 0 < \theta < c. \\
 2^o \quad M(\theta) &= P_\theta \left\{ \max_{1 \leq k \leq n} X_k < c \right\} = \left(\frac{c}{\theta} \right)^n = \alpha \left(\frac{\theta_0}{\theta} \right)^n, \quad c < \theta < \theta_0. \\
 3^o \quad M(\theta) &= P_\theta \{ (X_1, \dots, X_n) \in C \} = P_\theta \left\{ \max_{1 \leq k \leq n} X_k < c \right\} + 1 \\
 &\quad - P_\theta \left\{ \max_{1 \leq k \leq n} X_k \leq \theta_0 \right\} = 1 - (1 - \alpha) \left(\frac{\theta_0}{\theta} \right)^n, \quad \theta > \theta_0. \quad \blacktriangleleft
 \end{aligned}$$

Slika 21: Primjer 8.

Slika 22: Primjer 8.

Zadaci

1. Numerisana kocka se baca 100 puta. Testira se hipoteza H_0 – kocka je homogena protiv H_1 – kocka nije homogena. H_0 se odbacuje ako se u tih 100 bacanja paran broj pojavi manje od 40 puta ili više od 60 puta. Naći prag značajnosti testa.

2. Broj X stamparskih grešaka po stranici knjige, standardno ima $\mathcal{P}(0, 4)$ raspodjelu. Slučajni uzorak od 20 stranica dao je ukupno 14 grešaka. Sa pragom značajnosti $\alpha = 0,05$, testirati nultu hipotezu da je broj grešaka u okviru standarda, protiv alternativne da je van standarda.

3. Obilježje X ima raspodjelu iz familije $\mathcal{P} = \{\mathcal{U}(0, \theta), \theta > 1\}$. Za testiranje $H_0(\theta = 1)$ protiv $H_1(\theta > 1)$, sa pragom značajnosti α , uzima se uzorak obima $n = 48$ i predlažu se dva testa koja se zadaju kritičnim oblastima

$$\begin{aligned} a) C_1 &= \left\{ (x_1, \dots, x_n) : \sum_{k=1}^{48} x_k > c_1 \right\}, \\ b) C_2 &= \left\{ (x_1, \dots, x_n) : \max_{1 \leq k \leq 48} x_k > c_2 \right\}. \end{aligned}$$

Naći k_1 i k_2 . U slučaju b) naći funkciju moći testa.

4. Kocka je "fer" ako su strane numerisane od 1 do 6, a "nefer" ako umjesto 6 ima ponovljeno 5. Na osnovu rezultata 10 bacanja, testirati hipotezu H_0 da je kocka "fer" protiv hipoteze H_1 da je kocka "nefer". Koristiti test statistiku S – broj petica u 10 bacanja. Razumno definisati kritičnu oblast i odrediti α i β .

5. U primjeru 1 naći α i β , ako je kritična oblast za H_A :

$$E = \{(x_1, x_2, x_3, x_4) : x_1 + x_2 + x_3 + x_4 \geq 6\}, E \subset R^4.$$

Tablice

A) Normalna funkcija raspodjele: $\Phi^*(z_\alpha) = \frac{1}{\sqrt{2\pi}} \int_0^{z_\alpha} e^{-\frac{t^2}{2}} dt$.

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0159	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3380
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3718	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4083	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4430	.4441
1.6	.4452	.4463	.4474	.4485	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4758	.4762	.4767
2.0	.4773	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4865	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4980	.4980	.4981
2.9	.4981	.4982	.4983	.4984	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4986	.4987	.4987	.4988	.4988	.4988	.4989	.4989	.4989	.4990

B) χ_n^2 raspodjela: $\int_0^{h_n, \alpha} h_n(x) dx = \alpha$.

	α								
	0,10	0,20	0,30	0,50	0,70	0,80	0,90	0,95	0,99
1	0,016	0,064	0,148	0,455	1,074	1,642	2,706	3,841	6,635
2	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	9,210
3	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	11,345
4	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	12,277
5	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	15,086
6	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	16,812
7	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	18,475
8	3,490	4,594	5,557	7,344	9,524	11,030	13,362	15,507	20,090
9	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	21,666
10	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	23,209
11	5,579	6,989	8,148	10,341	12,899	14,631	17,276	19,675	24,725
12	6,304	7,807	9,034	11,340	14,011	15,812	18,549	21,026	26,217
13	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,362	27,688
14	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	29,141
15	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	30,578
16	9,312	11,152	12,624	15,338	18,418	20,465	23,542	26,296	32,000
17	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	33,409
18	10,865	12,857	14,440	17,338	20,601	22,860	25,989	29,869	34,805
19	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	36,191
20	12,443	14,578	16,266	19,338	22,775	25,038	28,412	31,410	37,564
21	13,240	15,445	17,182	20,337	23,958	26,171	29,615	32,671	38,932
22	14,041	16,314	18,101	21,337	24,839	27,301	30,813	33,924	40,289
23	14,848	17,187	19,021	22,337	26,018	28,429	32,007	35,172	41,638
24	15,659	18,062	19,943	23,337	27,096	29,553	33,196	36,415	42,980
25	16,473	18,940	20,867	24,337	28,172	30,675	34,382	37,652	44,314
27	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,113	46,963
28	18,739	21,588	23,647	27,336	31,391	34,027	37,916	41,337	48,278
29	18,768	22,475	24,577	28,336	32,461	35,139	39,087	42,557	49,588
30	20,599	23,364	25,508	29,336	33,530	36,250	40,256	43,773	50,892

C) Studentova raspodjela: $\int_0^{v_{n,\alpha}} v_n(x) dx = \alpha$.

n	α							
	0,10	0,20	0,30	0,40	0,45	0,475	0,49	0,495
1	0,325	0,727	1,376	3,078	6,314	12,706	31,823	63,657
2	0,289	0,617	1,061	1,886	2,920	4,403	6,965	9,925
3	0,277	0,584	0,978	1,638	2,353	3,182	4,531	5,841
4	0,271	0,569	0,941	1,533	2,132	2,776	3,747	4,604
5	0,267	0,559	0,920	1,476	2,015	2,571	3,365	4,032
6	0,265	0,553	0,906	1,440	1,943	2,447	3,143	3,707
7	0,263	0,549	0,896	1,415	1,895	2,365	2,998	3,499
8	0,262	0,546	0,889	1,397	1,860	2,306	2,896	3,355
9	0,261	0,543	0,883	1,383	1,833	2,262	2,821	3,250
10	0,260	0,542	0,879	1,372	1,812	2,228	2,764	3,169
11	0,260	0,540	0,876	1,363	1,796	2,201	2,718	3,106
12	0,259	0,539	0,873	1,356	1,782	2,179	2,681	3,055
13	0,259	0,538	0,870	1,350	1,771	2,160	2,650	3,012
14	0,258	0,537	0,868	1,345	1,761	2,145	2,624	2,977
15	0,258	0,536	0,866	1,341	1,753	2,131	2,602	2,947
16	0,258	0,535	0,865	1,337	1,746	2,120	2,583	2,921
17	0,257	0,534	0,863	1,133	1,740	2,110	2,567	2,898
18	0,257	0,534	0,862	1,330	1,734	2,101	2,552	2,878
19	0,257	0,533	0,861	1,328	1,729	2,093	2,539	2,861
20	0,257	0,533	0,860	1,325	1,725	2,086	2,528	2,845
21	0,257	0,532	0,859	1,323	1,721	2,080	2,518	2,831
22	0,256	0,532	0,858	1,321	1,717	2,074	2,508	2,819
23	0,256	0,532	0,858	1,319	1,714	2,069	2,500	2,807
24	0,256	0,531	0,857	1,318	1,711	2,064	2,492	2,797
25	0,256	0,531	0,856	1,316	1,708	2,060	2,485	2,787
26	0,256	0,531	0,856	1,315	1,706	2,056	2,479	2,779
27	0,256	0,531	0,855	1,314	1,703	2,052	2,473	2,771
28	0,256	0,530	0,855	1,313	1,701	2,048	2,467	2,763
29	0,256	0,530	0,854	1,311	1,699	2,045	2,462	2,756
30	0,256	0,530	0,854	1,310	1,697	2,042	2,457	2,750
40	0,255	0,529	0,851	1,303	1,684	2,021	2,423	2,704
60	0,254	0,527	0,848	1,296	1,671	2,000	2,390	2,660
120	0,254	0,526	0,845	1,289	1,658	1,980	2,358	2,617
∞	0,253	0,524	0,842	1,282	1,645	1,960	2,326	2,576

Sadržaj