VISOKA ŠKOLA ZA TURISTIČKI MENADŽMENT U ŠIBENIKU

STATISTIKA

predavanja i vježbe

Šibenik, 2004.

I. KOMBINATORIKA

1. UVOD

Prisjetimo se nekih osnovnih pojmova i oznaka iz teorije skupova koje ćemo često koristiti.

Skup S je **konačan skup** ako ne postoji bijekcija između skupa S i nekog njegovog pravog podskupa. Ili, skup S je konačan skup ako postoji bijekcija između skupa S i prvih n prirodnih brojeva $\{1,2,3,\ldots,n\}$. U tom slučaju se kaže da skup S ima n elemenata ili da je **kardinalni broj** kS jednak n: $kS = n$. Posebno se uzima da je kardinalni broj praznog skupa nula: $k\varnothing = 0$.

Primjer. Skup S = {Ivan, Andrija, Niko} je konačan skup jer postoji bijekcija između tog skupa i skupa $\{1,2,3\} \subset \mathbf{N}$. Vrijedi $kS = 3$.

Umjesto navedenog skupa S čiji su elementi imena (osoba ili hotela), u matematici se radi sa skupom $\{1,2,3\}$ čiji su elementi početni prirodni brojevi. To ne umanjuje općenitost razmatranja jer se uvijek može izvršiti identifikacija: 1 = Ivan, 2 = Andrija, 3 = Niko. Iz navedenog slijedi da pod konačnim skupom možemo uvijek, bez smanjenja općenitosti, smatrati skup $S = \{1,2,\ldots,n\}$. Vrijedi $kS = n$.

Po dogovoru se smatra da su svi **elementi** nekog skupa **međusobno različiti**. Npr. $\{1,1,1,2,2\}$ nije skup. Osim toga, **redoslijed** ispisivanja elemenata skupa **nije važan**. Npr. skupovi $\{2,3,1\}$ i $\{1,2,3\}$ su jednaki: $\{2,3,1\} = \{1,2,3\}$. Isto tako vrijedi $\{a,b\} = \{b,a\}$. To slijedi iz definicije jednakosti dvaju skupova.

Ako u dvočlanom skupu $\{a,b\}$ uvedemo uređaj tako da se zna koji je od elemenata prvi, a koji drugi, onda takav skup zovemo **uređeni par** ili **uređena dvojka** i označavamo sa (a,b). Element a se zove **prva komponenta** ili **koordinata**, a element b **druga komponenta** ili **koordinata** u uređenom paru (a,b). Stoga vrijedi $(a,b) \neq (b,a)$, odnosno:

$$(a,b) = (c,d) \Leftrightarrow a = c, b = d\,.$$

Neka su A i B neprazni skupovi. Skup čiji su elementi uređeni parovi (a,b), gdje je $a \in A$ i $b \in B$, zove se **direktni** ili **kartezijev umnožak skupova A i B** i označava sa $A \times B$. Dakle,

$$A \times B = \{(a,b) \mid a \in A, b \in B\}.$$

Primjer. A = {1,2,3}, B = {x, y} $\Rightarrow$ $A \times B = \{(1,x),(1,y),(2,x),(2,y),(3,x),(3,y)\}$

$$B \times A = \{(x,1),(x,2),(x,3),(y,1),(y,2),(y,3)\}$$

Iz navedene definicije slijedi da množenje skupova nije komutativno: $A \times B \neq B \times A$. Analogno se definira i kartezijev umnožak tri skupa: $A \times B \times C = \{(a,b,c) \mid a \in A, b \in B, c \in C\}$ (elementi ovog skupa zovu se uređene trojke), odnosno kartezijev umnožak n skupova:

$$A_1 \times A_2 \times \cdots \times A_n = \{(a_1, a_2, \ldots, a_n) \mid a_1 \in A_1, a_2 \in A_2, \ldots, a_n \in A_n\}.$$

Elementi ovog skupa zovu se **uređene *n*–torke** (entorke) koje imaju n komponenti.

Primjer. A = {1,2,3} $\Rightarrow$ $A \times A = A^2 = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)\}$

Skup A^2 zove se kvadrat skupa A.

$$\begin{aligned} A \times A \times A = \{&(1,1,1),(1,1,2),(1,1,3),(1,2,1),(1,2,2),(1,2,3),(1,3,1),(1,3,2),(1,3,3),\\ &(2,1,1),(2,1,2),(2,1,3),(2,2,1),(2,2,2),(2,2,3),(2,3,1),(2,3,2),(2,3,3),\\ &(3,1,1),(3,1,2),(3,1,3),(3,2,1),(3,2,2),(3,2,3),(3,3,1),(3,3,2),(3,3,3)\} \end{aligned}$$

Vidimo da vrijedi: $kA^2 = 9 = kA \cdot kA$, $kA^3 = 27 = 3^3 = kA \cdot kA \cdot kA$. Osim toga, vidimo da komponente u uređenoj n-torki mogu biti jednake.

Navedeni primjeri ilustriraju slijedeće teoreme:

Teorem 1. Ako su A i B konačni i disjunktni skupovi (tj. $A \cap B = \emptyset$), onda je

$$k(A \cup B) = k(A) + k(B).$$

Teorem 2. Ako su A i B konačni skupovi, onda je

$$k(A \times B) = k(A) \cdot k(B).$$

Teorem 3. Ako su $A_1, A_2, \ldots, A_n$ konačni skupovi, onda je

$$k(A_1 \times A_2 \times \cdots \times A_n) = kA_1 \cdot kA_2 \cdot \ldots \cdot kA_n.$$

Često se javlja problem određivanja broja elemenata konačnog skupa ili prebrojavanja elemenata konačnog skupa. Moguća su dva slučaja. Prvi slučaj je kada je skup S tako zadan da "vidimo" sve njegove elemente, pa ih je lako prebrojiti (npr. prste na ruci, klupe u razredu, itd.). Drugi slučaj je kada je skup S zadan na neki drugi način, npr. nekim svojim svojstvom. Tada prebrojavanje elemenata skupa S, tj. određivanje broja *k*S, može biti dosta složeno.

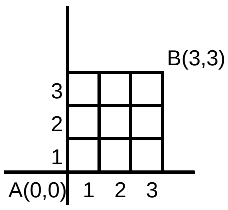
Primjer. S = $\{n \in \mathbf{N} \mid 100 < n < 300$ i n je prost broj$\}$, *k*S = ?

Primjer. T = {1,2,3,4,5,6,7,8}

a) Koliko ima troznamenkastih brojeva koji se mogu zapisati pomoću elemenata skupa T?
b) Koliko ima troznamenkastih brojeva koji se mogu zapisati od elemenata skupa T, a čije su znamenke različite?
c) Koliko ima tročlanih podskupova skupa T?

Uskoro ćemo ovaj zadatak moći lako riješiti.

Primjer.



Na slici u ravnini prikazane su točke A(0,0) i B(3,3). Neka je skup S staza iz A u B čija je duljina jednaka 6. Koliki je *k*S?

Iz brojeva unutar kružnica, zaključujemo da je *k*S = 20.

Matematička disciplina u kojoj se proučavaju metode za određivanje broja elemenata konačnog skupa zove se **kombinatorika**. Nama je kombinatorika potrebna zbog primjene u teoriji vjerojatnosti i statistici.

Osnovni teorem kombinatorike je ***Teorem o uzastopnom prebrojavanju:***

Neka je $S \subset A_1 \times A_2 \times \cdots \times A_n$ skup uređenih *n*-torki $(a_1, a_2, \ldots, a_n)$ definiran ovako: prva komponenta a_1 može se izabrati na p_1 načina, za svaki izbor od a_1 druga komponenta a_2 može se izabrati na p_2 načina,…, za svaki izbor a_1, $a_2, \ldots, a_{n-1}$ *n*-tu komponentu a_n može se izabrati na p_n načina. Tada je *k*S = $p_1 \cdot p_2 \cdot \ldots \cdot p_n$.

Primjer. Neka je $A = \{1,2,3,4\}$, a $S = \{(a, b) \mid a, b \in A$ i $a \neq b\} \subset A^2$. Izračunajmo kS. Budući da se prva komponenta a može izabrati na $p_1 = 4$ načina, a druga komponenta b na $p_2 = 3$ načina (jer mora biti $a \neq b$), vrijedi $kS = 4 \cdot 3 = 12$, a u što se možemo uvjeriti i ako skup S ispišemo eksplicite: $S = \{(1,2),(1,3),(1,4),\ldots,(4,3)\}$.

Primjer. Izračunajmo koliko ima troznamenkastih brojeva.

To je skup $S = \{(a_1, a_2, a_3) \mid a_i \in \{0,1,2,3,4,5,6,7,8,9\}\}$. Vrijedi $p_1 = 9$ (jer prva komponenta ne može biti nula), $p_2 = 10$, $p_3 = 10$, pa je kS $= 9 \cdot 10 \cdot 10 = 900$.

2. PERMUTACIJE

Definicija. Ako je $n \in \mathbf{N}$, tada umnožak prvih n prirodnih brojeva $1 \cdot 2 \cdot \ldots \cdot n$ zovemo ***n*** **faktorijela** i označavamo s ***n!***.

Dakle, $n! = 1 \cdot 2 \cdot \ldots \cdot n$. Posebno se definira $0! = 1$.

Primjer. $1! = 1$, $2! = 2$, $3! = 6$, $4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$, $10! = 1 \cdot 2 \cdot \ldots \cdot 10 = 3\,628\,800$, $20!$ je broj koji ima 19 znamenaka. Uočavamo da faktorijele vrlo brzo rastu porastom broja n.

$$\frac{5!}{3!} = 4 \cdot 5 = 20, \quad \frac{10!}{7!} = 8 \cdot 9 \cdot 10 = 720$$

Definicija. Neka je $n, r \in \mathbf{N} \cup \{0\}$ i $r \leq n$. Tada se broj $\binom{n}{r} = \dfrac{n!}{(n-r)!r!}$ zove **binomni koeficijent** (čita se: n povrh r ili n iznad r).

Kraćenjem brojnika i nazivnika dobiva se:

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

Lako se provjeri da je $\binom{n}{0} = \binom{n}{n} = 1$ i $\binom{n}{r} = \binom{n}{n-r}$.

Primjer. $\binom{6}{5} = \frac{6!}{(6-5)!5!} = \frac{6!}{5!} = 6$, $\binom{11}{5} = \frac{11!}{6!5!} = \frac{7 \cdot 8 \cdot 9 \cdot 10 \cdot 11}{5!} = 462$

Definicija. Neka je S = {1,2,…,*n*} konačan skup. Uređena *n*-torka $(a_1, a_2, \ldots, a_n)$ različitih elemenata skupa S zove se **permutacija skupa** S.
Drugim riječima, bijekcija $f: S \to S$ zove se permutacija skupa S.

Primjer. Nađimo sve permutacije skupa S = {1,2,3}.
To su (1,2,3), (1,3,2), (2,1,3), (2,3,1), (3,1,2), (3,2,1), ili kraće ispisano 123, 132, 213, 231, 312, 321. Ima ih $6 = 1 \cdot 2 \cdot 3 = 3!$
Uočite da su to zaista bijekcije skupa S. Npr. (3,1,2) je kraća oznaka za preslikavanje $1 \to 3, 2 \to 1, 3 \to 2$ koje je bijekcija.
Permutacija (1,2,3) zove se prva permutacija (ona je identično preslikavanje), a (3,2,1) zadnja permutacija. Ovaj poredak permutacija zove se leksikografski.

Teorem. Ako sa P(*n*) označimo broj permutacija skupa od *n* elemenata, onda vrijedi $P(n) = n!$

Dokaz: Prebrojimo koliko ima uređenih *n*-torki $(a_1, a_2, \ldots, a_n)$ različitih elemenata skupa S. a_1 možemo odabrati na *n* načina, a_2 na (*n* – 1) načina (jer je a_1 već "potrošen"), a_3 na (*n* –2) načina, …, a_n na 1 način.

Po teoremu o uzastopnom prebrojavanju broj tih *n*-torki je $n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1 = n!$

Primjer. Na koliko načina šest osoba može stajati u redu pred blagajnom?
Označimo li osobe sa 1, 2, 3, 4, 5, 6, mogući su 123456, 123465, …, 654321. Vidimo da su te šestorke permutacije, pa je traženi broj 6! = 720.

Primjer. Na koliko načina može 5 studenata i 4 studentice sjesti na klupu tako da svaka studentica sjedi između dva studenta?
Zaključujemo da studentice moraju sjediti na parnim mjestima. One mogu sjesti na 4! načina. Budući da studenti mogu sjesti na 5! načina i da uz svaki razmještaj studenata, studentice mogu sjesti na 4! načina, to je traženi broj prema teoremu o uzastopnom prebrojavanju jednak $4! \cdot 5! = 24 \cdot 120 = 2880$.

Definicija. Ako u nekoj n-torki $(a_1, a_2, \ldots, a_n)$ dopustimo da ima jednakih elemenata i to r_1 jednakih, r_2 jednakih, …, r_k jednakih, onda se ta uređena n-torka zove **permutacija s ponavljanjem**.

Može se pokazati da je broj tih permutacija s ponavljanjem $P_{r_1, r_2, \ldots, r_k}(n) = \dfrac{n!}{r_1!\, r_2! \cdot \ldots \cdot r_k!}$.

Jasno , tu je $r_1 + r_2 + \ldots + r_k \leq n$.

Primjer. Koliko se šesteroznamenkastih brojeva može napisati pomoću znamenki 3,3,4,4,4,9? Ovdje se radi o permutacijama od 6 elemenata među kojima su 2 trojke, 3 četvorke, tj. r_1=2, r_2=3, pa je $P_{2,3}(6) = \dfrac{6!}{2!3!} = \dfrac{4 \cdot 5 \cdot 6}{2} = 60$.

Primjer. Koliko različitih nizova možemo napraviti od 3 bijele, 4 plave i 5 crvenih perli?

Ovdje je $n = 3 + 4 + 5 = 12$, r_1=3, $r_2 = 4$, r_3=5, pa je $P_{3,4,5}(12) = \dfrac{12!}{3!4!5!} = 27720$.

3. VARIJACIJE

Definicija. Neka je S skup od n elemenata i $r \leq n$. Uređena r-torka $(a_1, a_2, \ldots, a_r)$ različitih elemenata skupa S zove se **varijacija *r*-tog razreda skupa od *n* elemenata**.

Teorem. Broj varijacija r-tog razreda skupa od n elemenata je $V_r(n) = \dfrac{n!}{(n-r)!}$.

Dokaz: Slijedi iz teorema o uzastopnom prebrojavanju:

$$V_r(n) = p_1 \cdot p_2 \cdot \ldots \cdot p_r = n \cdot (n-1) \cdot \ldots \cdot (n-r+1) = \frac{n!}{(n-r)!}$$

Iz formule za $V_r(n)$ slijedi: $r = 1 \Rightarrow V_1(n) = \dfrac{n!}{(n-1)!} = n$ (što smo i trebali dobiti jer varijacija 1-og razreda ima onoliko koliko skup ima elemenata),

$r = n \Rightarrow V_n(n) = \dfrac{n!}{(n-n)!} = \dfrac{n!}{0!} = n! = P(n)$ (što smo i trebali dobiti jer su varijacije n-tog razreda skupa od n elemenata permutacije tog skupa).

Primjer. $n=4, r=2 \Rightarrow$ S = {1,2,3,4}, a varijacije 2-og razreda tog skupa su:

12	21	31	41
13	23	32	42
14	24	34	43

Ima ih dakle 12, a to i slijedi iz formule $V_2(4)=\frac{4!}{(4-2)!}=12$.

Primjer. Riješimo sada zadatak iz Uvoda: Koliko ima troznamenkastih brojeva čije su znamenke različite i elementi su skupa T = {1,2,3,4,5,6,7,8}?

123
124
125
…

Rješenje je $V_3(8)=\frac{8!}{(8-3)!}=\frac{8!}{5!}=6\cdot 7\cdot 8=336$. Ili, po teoremu o uzastopnom prebrojavanju:

$p_1\cdot p_2\cdot p_3=8\cdot 7\cdot 6=336$.

Definicija. Neka je zadan skup S od n elemenata i $r\le n$ ili $r>n$. Uređena r-torka $(a_1, a_2, \dots, a_r)$ elemenata skupa S čije komponente mogu biti jednake, zove se **varijacija s ponavljanjem *r*-tog razreda skupa od *n* elemenata**.

Teorem. Broj varijacija s ponavljanjem r-tog razreda skupa od n elemenata je $\overline{V}_r(n)=n^r$.

Dokaz:

Slijedi iz teorema o uzastopnom prebrojavanju:

$$\overline{V}_r(n)=p_1\cdot p_2\cdot\ldots\cdot p_r=n\cdot n\cdot\ldots\cdot n=n^r .$$

Primjer. S = {1,2,3,4}, $r=2$. Ispišimo sve varijacije s ponavljanjem:

11	21	31	41
12	22	32	42
13	23	33	43
14	24	34	44

Ima ih dakle 16, odnosno $\overline{V}_2(4)=4^2=16$

Primjer. Koliko ima troznamenkastih brojeva čije su znamenke elementi su skupa T = {1,2,3,4,5,6,7,8}?

Ima ih $\overline{V}_3(8)=8^3=512$.

Primjer. Bacaju se dvije kocke.

a) Koliki je broj mogućih ishoda?

b) Koliki je broj mogućih ishoda kod kojih su kocke pale na različite brojeve?

a)

11	…	61
12	…	62
13	…	63
14	…	64
15	…	65
16	…	66

$$\overline{V}_2(6) = 6^2 = 36$$

b) $V_2(6) = \dfrac{6!}{(6-2)!} = \dfrac{6!}{4!} = 6 \cdot 5 = 30$

4. KOMBINACIJE

Definicija. Neka je zadan skup S od n elemenata i $r \le n$. Podskup od r (različitih) elemenata skupa S zove se **kombinacija *r*-tog razreda skupa od *n* elemenata**.

Teorem. Broj kombinacija r-tog razreda skupa od n elemenata je $K_r(n) = \binom{n}{r}$.

Dokaz:

Slijedi iz broja varijacija r-tog razreda skupa od n elemenata, odnosno iz veze između varijacija i kombinacija. Naime, ako u svakoj kombinaciji načinimo sve permutacije, a njih je $r!$, dobivamo sve varijacije (zato se može reći da su varijacije "permutirane" kombinacije). Stoga vrijedi: $V_r(n) = n!K_r(n) \;\Rightarrow\; K_r(n) = \dfrac{V_r(n)}{n!} = \dfrac{n!}{(n-r)!r!} = \binom{n}{r}$.

Primjer. $n = 4, r = 2 \;\Rightarrow\;$ S = $\{1,2,3,4\}$, a kombinacije 2-og razreda tog skupa su:

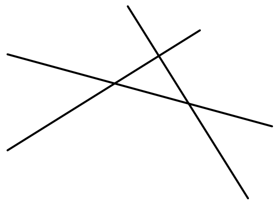
12		
13	23	
14	24	34

Ima ih dakle 6, a to slijedi i iz formule: $K_2(4) = \binom{4}{2} = \dfrac{4!}{2!2!} = 6$.

Primjer. Skup se sastoji od 20 proizvoda. Na koliko načina se može formirati uzorak od 5 proizvoda?

Kad se govori o **uzorku** uvijek se misli na njegov sadržaj, a ne na poredak elemenata u uzorku. Stoga je svaki uzorak jedna kombinacija, odnosno traženi broj je $\binom{20}{5}=\frac{20!}{15!5!}=\frac{16\cdot 17\cdot 18\cdot 19\cdot 20}{120}=15504$.

Primjer. U ravnini je zadano 6 točaka od kojih nijedna trojka ne leži na istom pravcu. Koliko je pravaca određeno tim točkama?



Budući je svaki pravac određen s dvije točke traženi broj je: $K_2(6)=\binom{6}{2}=15$.

Definicija. Ako se u kombinaciji r-tog razreda od n elemenata dopusti da elementi mogu biti i jednaki, onda se takva kombinacija zove **kombinacija s ponavljanjem *r*-tog razreda skupa od *n* elemenata** (ovdje može biti i $r>n$).

Teorem. Broj tih kombinacija je $\overline{K}_r(n)=\binom{n+r-1}{r}$.

Primjer. Ispišimo sve kombinacije s ponavljanjem za $n=4, r=2$ (S = {1,2,3,4}):

11
12 22
13 23 33
14 24 34 44 $\Rightarrow$ Ima ih 10, tj. $\overline{K}_2(4)=\binom{4+2-1}{2}=\binom{5}{2}=10$.

Primjer. Riješi jednadžbu $\overline{K}_2(n)=78$.

$$\binom{n+2-1}{2}=\binom{n+1}{2}=\frac{(n+1)!}{(n-1)!2!}=\frac{n(n+1)}{2}=78 \Rightarrow n^2+n-156=0 \Rightarrow n=12.$$

5. ZADACI

1. Koliko permutacija skupa S = {1,2,3,4,5,6,7} počinje sa 2 3 4?

Rješenje: $\underline{234}$ _ _ _ _ Imamo 4 "prazna mjesta" na koje dolaze permutacije skupa T = {1,5,6,7}, a njih ima $P(4) = 4! = 24$.

2. Nađite 83. permutaciju skupa S = {1,2,3,4,5}.

Rješenje: Budući da s jednim fiksiranim brojem na prvom mjestu ima 4! = 24 permutacije, s dva fiksirana broja na prva dva mjesta ima 3! = 6 permutacija, lako se zaključi da 73. permutacija 4 1 2 3 5, 79.permutacija je 4 2 1 3 5, odnosno tražena 83. permutacija je 4 2 5 13

3. Koliko se osmeroznamenkastih brojeva može napisati od znamenki 0,1,2,2,2,3,3,4.

Rješenje:

$$P_{3,2}(8) - P_{3,2}(7) = \frac{8!}{3!2!} - \frac{7!}{3!2!} = 3360 - 420 = 2940$$

4. Koliko ima četveroznamenkastih brojeva koji se mogu zapisati znamenkama 1,2,3,4,5 i kod kojih se znamenke ne ponavljaju? Koliko tih brojeva završava sa 13?

Rješenje:

$$V_4(5) = \frac{5!}{(5-4)!} = 120,$$

$$\_\,\_\ \underline{13} \quad V_2(3) = \frac{3!}{(3-2)!} = 6.$$

5. Katanac na šifru ima tri koluta s 10 brojeva. Koliko ima različitih šifri?

Rješenje:

$$n = 10, r = 3 \;\Rightarrow\; \overline{V}_3(10) = 10^3 = 1000.$$

6. Sportska prognoza sastoji se od 12 redaka u koje treba upisati 0, 1 ili 2. Na koliko se načina to može popuniti?

Rješenje:

$$n = 3, r = 12 \;\Rightarrow\; \overline{V}_{12}(3) = 3^{12} = 531441.$$

7. Na koliko se načina u razredu od 30 učenika može izabrati 2 redara?

Rješenje:

$$K_2(30) = \binom{30}{2} = 435\,.$$

8. Skup se sastoji od 15 dobrih i 5 loših proizvoda.

a) Na koliko načina se može formirati uzorak od 4 dobra proizvoda?

b) Na koliko načina se može formirati uzorak koji sadrži 4 dobra i 3 loša proizvoda?

Rješenje:

a) $K_4(15) = \binom{15}{4} = 1365$

b) $\binom{15}{4} \cdot \binom{5}{3} = 13650\,.$

9. U sobi ima 6 sijalica i sve su ispravne. Na koliko načina se soba može osvijetliti?

Rješenje:

$$\binom{6}{1} + \binom{6}{2} + \binom{6}{3} + \binom{6}{4} + \binom{6}{5} + \binom{6}{6} = 63\,.$$

10. U firmi ima 20 muškaraca i 10 žena. Na koliko je načina moguće izabrati peteročlanu delegaciju u kojoj su barem 2 žene?

Rješenje:

Delegacija može imati: 2 žene (Ž) i 3 muškarca (M) ili 3Ž i 2M ili 4Ž i 1M ili 5Ž, dakle tih delegacija ima:

$$\binom{10}{2} \cdot \binom{20}{3} + \binom{10}{3} \cdot \binom{20}{2} + \binom{10}{4} \cdot \binom{20}{1} + \binom{10}{5} = 78552\,.$$

11. Riješi jednadžbu: $\dfrac{1}{K_4(n)} - \dfrac{1}{K_5(n)} = \dfrac{2}{K_6(n)}$

Rješenje:

$$\frac{1}{K_4(n)} - \frac{1}{K_5(n)} = \frac{2}{K_6(n)} \quad \Rightarrow \quad \frac{1}{\frac{n!}{(n-4)!4!}} = \frac{1}{\frac{n!}{(n-5)!5!}} + \frac{2}{\frac{n!}{(n-6)!6!}} \quad / \cdot n! \quad \Rightarrow$$

$$\Rightarrow \quad (n-4)!4! = (n-5)!5! + 2(n-6)!6! \quad / : \ (n-6)!4!$$

$$(n-5)(n-4) = (n-5)\cdot 5 + 2\cdot 5\cdot 6$$

$$n^2 - 9n + 20 - 5n + 25 - 60 = 0$$

$$n^2 - 14n - 15 = 0 \Rightarrow n = \frac{14 \pm \sqrt{196+60}}{2} = \begin{cases} n = 15 \\ n = -1 \end{cases} \Rightarrow n = 15$$

12. Koliko različitih riječi možemo sastaviti od 8 suglasnika i 5 samoglasnika, ali tako da svaka riječ sadrži 5 različitih suglasnika i 4 različitih samoglasnika?

Rješenje:

Pet različitih suglasnika može biti izabrano na $\binom{8}{5}$ načina,

Četiri različita samoglasnika može biti izabrano na $\binom{5}{4}$ načina.

Od 9 slova (5 suglasnika i 4 samoglasnika) može se napraviti 9! poredaka. Stoga je prema teoremu o uzastopnom prebrojavanju traženi broj:

$$\binom{8}{5}\cdot\binom{5}{4}\cdot 9! = 101\ 606\ 400.$$

13. a) Koliko ima različitih putnih karata na željezničkoj pruzi od 27 stanica?

b) Koliko ima stanica na željezničkoj pruzi ako postoji 5852 putnih karata?

Rješenje:

a)

$$\begin{array}{lll} 1,2 & 2,1 & \dots \\ 1,3 & 2,3 & \dots \\ \vdots & \vdots & \\ 1,27 & 2,27 & \dots \end{array} \qquad \Rightarrow \ V_2(27) = \frac{27!}{25!} = 26\cdot 27 = 702$$

b)

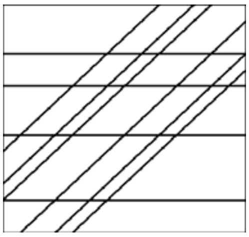
$$V_2(n) = \frac{n!}{(n-2)!} = 5852$$

$$(n-1)\cdot n = 5852 \quad \Rightarrow \quad n = 77$$

14. Koliko paralelograma određuje skup od 4 paralelna pravca u presjeku sa skupom od 7 paralelnih pravaca?

Rješenje:

$$\binom{7}{2}\cdot\binom{4}{2}=126$$



15. Na koliko načina možemo 5 novčića od 1 kune podijeliti između 3 djeteta?

Rješenje:

$n=3, r=5$

11111 (1. dijete je dobilo svih 5 kuna)
11112 (1. dijete je dobilo 4, a 2. dijete 1 kunu)
11113
11122
…

Ovo su dakle kombinacije s ponavljanjem, pa je njihov broj: $\overline{K}_5(3)=\binom{3+5-1}{5}=\binom{7}{5}=21$.

16. Na šahovskom turniru odigrano je 78 partija pri čemu je svaki igrač odigrao po jednu partiju sa svakim sudionikom turnira. Koliko je igrača bilo?

Rješenje:

$$K_2(n)=\binom{n}{2}=78 \quad\Rightarrow\quad n=13.$$

17. Na koliko načina 6 osoba može sjesti za okrugli stol, ako nije važno gdje tko sjedi, nego je samo važan njihov međusobni raspored?

Rješenje:

Prvo jedna osoba sjedne gdje hoće, a zatim se ostalih 5 osoba može smjestiti na $5! = 120$ načina. Dakle, odgovor je 120 načina.

II. VJEROJATNOST

1. PROSTOR DOGAĐAJA

Osnovni zadatak znanosti je da otkriva zakonitosti u prirodi i društvu. Te zakonitosti mogu biti dvojake prirode: determinističke i stohastičke.

Deterministička zakonitost je kada se pri svakoj realizaciji određenih uvjeta U ostvaruje događaj A. Ubuduće ćemo realizaciju određenih uvjeta U zvati **pokusom** ili **eksperimentom**.

Primjer 1. Ako kovani novčić ispustimo iz ruke (to je skup uvjeta U, tj. Pokus ili eksperiment), on će pasti na tlo (događaj A).

Primjer 2. Ako se u vakuumu različiti predmeti puste da slobodno padaju (pokus), oni će u svakom trenutku imati jednake brzine (događaj A).

Primjer 3. Kod proizvoljne kemijske reakcije izolirane od okoline (pokus), ukupna količina materije ostaje nepromijenjena (događaj A).

Stohastička (ili nedeterministčka ili vjerojatnosna)**zakonitost** je kada se pri eksperimentu neki događaj A može, ali i ne mora ostvariti. Dakle, takav događaj je mogući ishod ili rezultat nekog pokusa, on je slučajna pojava.

Primjer 4. Ako novčić ispustimo iz ruke, pasti će na «pismo» (događaj A).

Primjer 5. Spol djeteta koje se rodi je muški.

Teorija vjerojatnosti je grana matematike koja izučava moguće ishode nekog pokusa, odnosno slučajne pojave, i one ukazuju na skup svih mogućih ishoda. Početak svog razvoja duguje igrama na sreću čija je popularnost u 17. stoljeću utjecala da se ti problemi počnu matematički tretirati uvođenjem metoda kombinatorike. Danas se teorija vjerojatnosti primjenjuje u raznim područjima fizike, tehnike, ekonomije, psihologije, genetike itd.

Jedan takav pokus je igra **bacanje kocke**. Kod opisa pokusa, pored preciziranja uvjeta pod kojima se pokus odvija (idealno napravljena kocka od homogenog materijala), treba reći što se traži ili očekuje kao rezultat pokusa. Kod bacanja kocke, kao rezultat bilježi se broj koji je na gornjoj strani kocke. Tada je skup svih mogućih ishoda tog pokusa skup $\{1,2,3,4,5,6\}$. Kod bacanja novčića taj skup je $\{pismo, glava\} = \{P, G\}$.

Definicija. Svaki rezultat ili ishod pokusa zove se **elementarni događaj** i označava se E_i. Skup svih elementarnih događaja označava se sa Ω: $\Omega = \{ E_i \,|\, E_i$ je elementarni događaj$\}$. Partitivni skup $\mathrm{P}(\Omega) = \{\mathrm{A} \,|\, \mathrm{A} \subset \Omega\}$ skupa Ω zove se **prostor događaja**, a elementi skupa $\mathrm{P}(\Omega)$ zovu se **slučajni događaji** ili **događaji**. Posebno, događaj Ω zove se **siguran događaj**, a događaj $\emptyset$ **nemoguć događaj** (jer je $\Omega, \emptyset \in \mathrm{P}(\Omega)$).
Ponovimo, događaj A je element skupa $\mathrm{P}(\Omega)$ odnosno podskup skupa Ω:

$$\mathrm{A} \in \mathrm{P}(\Omega),\ \mathrm{A} \subset \Omega.$$

Primjer. Kod bacanja kocke, skup elementarnih događaja je $\Omega = \{E_1, E_2, ..., E_6\}$, gdje E_i znači «pao je broj *i*». Tada je događaj D_1 «pao je neparan broj» skup $D_1 = \{E_1, E_3, E_5\}$, a događaj D_2 «pao je paran broj» skup $D_2 = \{E_2, E_4, E_6\}$.
Jasno, $D_1, D_2 \subset \Omega$, $D_1, D_2 \in \mathrm{P}(\Omega)$.

Na prostoru događaja $\mathrm{P}(\Omega)$ definirana je relacija inkluzije, te operacije unija, presjek i komplement.

Neka su A i B događaji, tj. A, B$\in \mathrm{P}(\Omega)$. Ako vrijedi $\mathrm{A} \subset \mathrm{B}$, tada kažemo da događaj A **implicira** ili **uvjetuje** događaj B. To znači da kad se realizira (ostvari) događaj A, onda se realizira i događaj B.
U prethodnom primjeru bacanja kocke, događaj E_1 implicira događaj D_1 (jer je $E_1 \subset D_1$), a isto vrijedi i za E_3 i E_5.

Unija A$\cup$B događaja A i B je događaj koji se realizira ako se realizira barem jedan od događaja A i B, tj. ili A ili B ili A i B. Zato se A$\cup$B čita «A ili B». Vrijedi:

$$\mathrm{A} \cup \emptyset = \mathrm{A},\ \mathrm{A} \cup \Omega = \Omega,\ \forall \mathrm{A} \in \mathrm{P}(\Omega).$$

Presjek A$\cap$B događaja A i B je događaj koji se realizira ako se istovremeno realizira i događaj A i događaj B. Zato se A$\cap$B čita «A i B». Vrijedi:

$$\mathrm{A} \cap \emptyset = \emptyset,\ \mathrm{A} \cap \Omega = \mathrm{A},\ \forall \mathrm{A} \in \mathrm{P}(\Omega).$$

Ako vrijedi $A \cap B = \emptyset$, tj. ako događaji A i B ne mogu istovremeno nastupiti, onda kažemo da su ti događaji **disjunktni** ili da se mađusobno **isključuju**.

Događaj A^C **komplementaran** je ili **suprotan događaj** događaja A ako je ishod pokusa točno jedan od ta dva događaja. Piše se i $A^C = \overline{A}$. Vrijedi:

$$A \cup \overline{A} = \Omega, \quad A \cap \overline{A} = \emptyset, \quad \overline{A} = \Omega / A.$$

2. DEFINICIJA VJEROJATNOSTI

<u>Definicija(klasična, a priori)</u>. Neka je rezultat pokusa element konačnog skupa elementarnih događaja $\Omega = \{E_1, E_2, ..., E_n\}$ koji su svi:

a) jednako mogući (vjerojatni) i

b) po parovima disjunktni.

Ako događaj A uvjetuje (implicira) *m* elementarnih događaja, onda je **vjerojatnost p(A) događaja A** broj $p(A) = \frac{m}{n}$.

Često se za elementarne događaje koji uvjetuju (impliciraju, realiziraju, ostvaruju) događaj A kaže da su **povoljni** za događaj A. U tom smislu se navedena formula za vjerojatnost p(A) događaja A obično pamti u obliku:

$$\frac{\text{broj povoljnih elementarnih događaja}}{\text{broj mogućih elementarnih događaja}}$$

Iz ove formule proizlazi da se računanje vjerojatnosti svodi na određivanje broja kA elemenata skupa A, a taj problem rješava kombinatorika.

Nedostatak ove klasične definicije vjerojatnosti je što se u njoj koristi pojam "jednako moguć ili vjerojatan" i što se odnosi na konačan broj elementarnih događaja.

Ovom definicijom zadana je **funkcija vjerojatnosti** $p : P(\Omega) \to \mathbf{R}$ koja ima svojstva:

1. $\forall A \in P(\Omega)$ vrijedi $0 \le p(A) \le 1$,
2. $p(\emptyset) = 0$ i $p(\Omega) = 1$,
3. $A \cap B = \emptyset \Rightarrow p(A \cup B) = p(A) + p(B)$ (svojstvo aditivnosti).

Iz ovih svojstava slijedi: $p(A) + p(\overline{A}) = 1$, jer je $A \cap \overline{A} = \emptyset$, $A \cup \overline{A} = \Omega$ i $p(\Omega) = 1$. Slično vrijedi i za uniju bilo kojeg broja događaja koji su po parovima disjunktni:

$$p(A_1 \cup A_2 \cup \ldots \cup A_n) = p(A_1) + p(A_2) + \ldots + p(A_n),$$

gdje su $A_i \in P(\Omega)$, $A_i \cap A_j = \emptyset$ za $i \ne j$.

Ako za događaje A_i vrijedi $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$ i $A_i \cap A_j = \emptyset$ za $i \ne j$, onda kažemo da oni čine **potpun sustav događaja**. Očito za potpun sustav događaja vrijedi:

$$p(A_1) + p(A_2) + \ldots + p(A_n) = 1.$$

Ako su $A, B \in P(\Omega)$ proizvoljni događaji, tj. koji jesu ili nisu disjunktni, onda je vjerojatnost događaja $A \cup B$ (A ili B):

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

(slijedi iz formule: $k(A \cup B) = k(A) + k(B) - k(A \cap B)$).

Primjer. Iz skupa $\{1,2,3,\ldots,12\}$ nasumce biramo jedan broj. Nađimo vjerojatnost da taj broj bude:

a) paran broj,
b) djeljiv s 3,
c) paran i djeljiv s 3,
d) paran ili djeljiv s 3.

Skup elementarnih događaja je $\Omega = \{E_1, E_2, \ldots, E_{12}\}$, gdje je E_i događaj da je izabran broj i, pa je $k\Omega = 12$. Vrijedi $p(E_i) = \frac{1}{12}, \forall i$, jer su svi ishodi pokusa jednako vjerojatni. Elementarni događaji su po parovima disjunktni: $E_i \cap E_j = \emptyset$ za $i \ne j$, pa možemo primijeniti navedenu definiciju vjerojatnosti.

a) Događaj A je "izabrani broj je paran" $\Rightarrow$ $A=\{E_2, E_4, E_6, E_8, E_{10}, E_{12}\}$ $\Rightarrow$

$kA = m = 6 \ \Rightarrow \ p(A)=\frac{6}{12}=\frac{1}{2}$.

b) Događaj B je "izabrani broj djeljiv je s 3" $\Rightarrow$ $B=\{E_3, E_6, E_9, E_{12}\}$ $\Rightarrow$

$kB = m = 4 \ \Rightarrow \ p(A)=\frac{4}{12}=\frac{1}{3}$.

c) Događaj C je "izabrani broj je paran i djeljiv s 3" $\Rightarrow$ $C=\{E_6, E_{12}\}$ $\Rightarrow$

$kC = m = 2 \ \Rightarrow \ p(C)=\frac{2}{12}=\frac{1}{6}$.

ILI

$C = A \cap B \ \Rightarrow \ p(C) = p(A\cap B) = p(A)\cdot p(B/A) = \frac{1}{2}\cdot\frac{2}{6}=\frac{1}{6}$, gdje je p(B/A) vjerojatnost da je B ako je bio A – ovo je tzv. uvjetna vjerojatnost koju ćemo definirati u slijedećoj točki.

d) Događaj D je "izabrani broj je paran ili djeljiv s 3" $\Rightarrow$

$D=\{E_2, E_3, E_4, E_6, E_8, E_9, E_{10}, E_{12}\} \ \Rightarrow \ kD = m = 8 \ \Rightarrow \ p(D)=\frac{8}{12}=\frac{2}{3}$.

ILI

$D = A\cup B \ \Rightarrow \ p(D) = p(A\cup B) = p(A) + p(B) - p(A\cap B) = \frac{1}{2}+\frac{1}{3}-\frac{1}{6}=\frac{4}{6}=\frac{2}{3}$

(vidi pod a), b), c)).

Pored klasične definicije postoji i statistička definicija vjerojatnosti koja se zasniva na izvođenju velikog broja pokusa.

Definicija **(statistička, a posteriori).** Ako se u istim uvjetima izvede serija od N pokusa i ako se u svakom pokusu događaj A može ostvariti ili ne ostvariti, onda se **relativna frekvencija događaja A** definira kao omjer $\frac{N(A)}{N}$, gdje je N(A) broj pokusa u kojima se događaj A ostvario, a N ukupan broj pokusa. Vjerojatnost događaja A je granična vrijednost njegove relativne frekvencije kada se N neizmjerno povećava: $p(A)=\lim_{N\to\infty}\frac{N(A)}{N}$.

U teoriji se može dokazati da taj limes postoji (tzv. zakon velikih brojeva), a u praksi se uzima

$p(A) \approx \frac{N(A)}{N}$ za dovoljno velike N.

Komentirajmo upotrebu danih dviju definicija vjerojatnosti. Pretpostavimo da imamo kutiju s crnim i bijelim kuglicama i da se pitamo kolika je vjerojatnost da ćemo izvući crnu kuglicu. Ako <u>znamo</u> sadržaj kutije, recimo da je u njoj 40 bijelih i 10 crnih kuglica, po klasičnoj definiciji vjerojatnosti, vjerojatnost da ćemo izvući crnu kuglicu je $\frac{10}{50} = \frac{1}{5} = 0.2$. Ako pak <u>ne znamo</u> sadržaj kutije, izvršit ćemo veći broj izvlačenja i bilježiti ishod koji će nam dati relativnu frekvenciju promatranog događaja. Napravit ćemo recimo 5000 izvlačenja. Ako smo pri tom 1065 puta izvukli crnu kuglicu onda je po formuli za vjerojatnost izvlačenja crne kuglice $\frac{N(A)}{N} = \frac{1065}{5000} = 0.213$. To je razlog zbog čega se klasična definicija zove "a priori", a stohastička "a posteriori".

3. UVJETNA VJEROJATNOST I TEOREM MNOŽENJA

Neka je $A \in P(\Omega)$ neki događaj i $p(A)>0$. Tada se u prostoru događaja $P(\Omega)$ može promatrati vjerojatnost da se pojavi neki događaj B pod uvjetom da se pojavio događaj A. Takva vjerojatnost zove se **uvjetna vjerojatnost**, a **p(B/A)** je **vjerojatnost događaja B pod uvjetom da se zbio događaj A**.

<u>Primjer</u>. Baca se kocka, te neka je A događaj "pao je neparan broj", a B događaj "pao je broj 3". Tada je vjerojatnost (bezuvjetna) događaja B: $p(B) = \frac{1}{6}$. Međutim uz pretpostavku da se zbio događaj A, tj. da je pao neparan broj, ta vjerojatnost je $p(B/A) = \frac{1}{3}$. Analogno vrijedi još: $p(B/\overline{A}) = 0$.

Teorem množenja. Ako su A,B∈P(Ω), onda vrijedi:

$$p(A \cap B) = p(A) \cdot p(B/A) = p(B) \cdot p(A/B).$$

Riječima, vjerojatnost da se dogode dva događaja jednaka je umnošku vjerojatnosti prvog i uvjetne vjerojatnosti drugog (to je vjerojatnost da se dogodi drugi, ako se dogodio prvi).

Dokaz:

Neka pokus ima *n* elementarnih događaja, od kojih *r* uvjetuje događaj A, a *t* događaj $A \cap B$. Tada je:

$$p(B/A) = \frac{t}{r} = \frac{\frac{t}{n}}{\frac{r}{n}} = \frac{p(A \cap B)}{p(A)}, \text{ tj. } p(A \cap B) = p(A) \cdot p(B/A).$$

Drugi dio tvrdnje slijedi iz prvog i komutativnosti presjeka.

Teorem množenja može se induktivno primijeniti i na višestruke presjeke:

$$p(A \cap B \cap C) = p((A \cap B) \cap C) = p(A \cap B) \cdot p(C/A \cap B) = p(A) \cdot p(B/A) \cdot p(C/A \cap B).$$

Ako pojava događaja A ne mijenja vjerojatnost događaja B, tj. ako je p(B) = p(B/A), onda kažemo da su A i B **nezavisni događaji**. Stoga za vjerojatnost dvaju nezavisnih događaja A i B vrijedi $p(A \cap B) = p(A) \cdot p(B)$. Ako ova formula ne vrijedi, događaji A i B su zavisni. Posebno, ako su događaji A i B iz različitih prostora događaja $P(\Omega_1)$, $P(\Omega_2)$ (npr. pri istovremenom bacanju kocke i novčića), oni su nezavisni i vrijedi $p(A \cap B) = p(A) \cdot p(B)$.

Analogna formula vrijedi i za više u parovima nezavisnih događaja:

$$p(A_1 \cap A_2 \cap \ldots \cap A_k) = p(A_1) \cap p(A_2) \cap \ldots \cap p(A_k).$$

Primjer. U kutiji se nalaze 4 bijele i 6 crnih kuglica. Kolika je vjerojatnost da prva i druga nasumce uzeta kuglica bude bijela, ako nakon prvog izvlačenja kuglicu ne vraćamo u kutiju?

$p(B_1) = \frac{4}{10} =$ vjerojatnost da prva uzeta kuglica bude bijela,

$p(B_2 / B_1) = \frac{3}{9} =$ vjerojatnost da druga kuglica bude bijela pod uvjetom da je i prva bijela

$$\Rightarrow \quad p(B_1 \cap B_2) = p(B_1) \cdot p(B_2 / B_1) = \frac{4}{10} \cdot \frac{3}{9} = \frac{2}{15} = 0.133...$$

Da je u ovom zadatku bila pretpostavka da nakon prvog izvlačenja kuglicu vratimo u kutiju, vjerojatnost bi bila $p(B_1 \cap B_2) = p(B_1) \cdot p(B_2 / B_1) = \frac{4}{10} \cdot \frac{4}{10} = 0.16.$

Primjer. U nekoj tvornici od 100 proizvedenih čaša prosječno je 6 neispravno, a od ispravnih čaša 75% je 1. klase. Kolika je vjerojatnost da čaša iz te tvornice bude 1. klase?

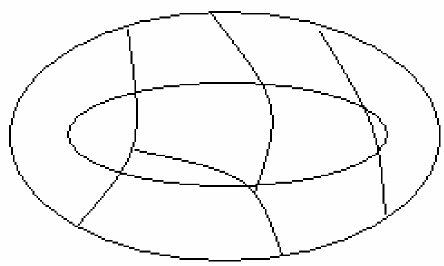
Neka je A događaj "čaša je ispravna", a B događaj "čaša je 1. klase". Tada je:

$$p(\overline{A}) = p\,(\text{čaša je neispravna}) = \frac{6}{100} = 0.06 \;\;\Rightarrow\;\; p(A) = 1 - p(\overline{A}) = 0.94\,,$$

$$p(B/A) = 75\% = 0.75\,,$$

$$p(A \cap B) = p(A)\cdot p(B/A) = 0.94 \cdot 0.75 = 0.705\,.$$

Neka događaji $A_1, A_2, \ldots, A_n \in \mathrm{P}(\Omega)$ čine potpun sustav događaja (dakle $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$, $A_i \cap A_j = \emptyset$) i neka je $\mathrm{B} \in \mathrm{P}(\Omega)$ proizvoljan događaj.



Tada je (vidi skicu) $B = (B \cap A_1) \cup (B \cap A_2) \cup \ldots (B \cap A_n)$, odnosno

$$p(B) = p(B \cap A_1) + p(B \cap A_2) + \ldots + p(B \cap A_n).$$

Ako na svaki sumand primijenimo teorem množenja, dobivamo

$$p(B) = p(A_1)\cdot p(B \cap A_1) + p(A_2)\cdot p(B \cap A_2) + \ldots + p(A_n)\cdot p(B \cap A_n)$$

i to je **formula totalne vjerojatnosti**.

Bayes-ov teorem (formula). Ako je događaj B nastupio kao posljedica jednaog od događaja $A_1, A_2, \ldots, A_n$ koji su u parovima disjunktni, onda vrijedi:

$$p(A_k / B) = \frac{p(A_k)\cdot p(B/A_k)}{\sum_{i=1}^{n} p(A_i)\cdot p(B/A_i)}\,, \quad k = 1,2,\ldots,n\,.$$

Dokaz:

Iz formule za uvjetnu vjerojatnost slijedi:

$p(A_k \cap B) = p(A_k)\cdot p(B/A_k) = p(B)\cdot p(A_k/B)$, odnosno $p(A_k/B) = \dfrac{p(A_k)\cdot p(B/A_k)}{p(B)}$, a

iz formule za totalnu vjerojatnost slijedi $p(B) = \sum_{i=1}^{n} p(A_i)\cdot p(B/A_i)$, a to je Bayes-ova formula.

<u>Primjer.</u> Imamo 3 jednake kutije: u prvoj ima 20 bijelih, u drugoj ima 10 bijelih i 10 crnih, a u trećoj kutiji ima 20 crnih kuglica. Iz slučajno odabrane kutije izvučemo jednu bijelu kuglicu. Kolika je vjerojatnost da je izvučena iz prve kutije?
Neka je A_1 događaj izbor 1. kutije, A_2 izbor 2. kutije, a A_3 izbor 3. kutije. Događaj B neka je izvlačenje bijele kuglice. Tada vrijedi:

$$p(A_1) = p(A_2) = p(A_3) = \frac{1}{3},$$

$$p(B/A_1) = 1,\quad p(B/A_2) = \frac{1}{2},\quad p(B/A_3) = 0$$

$$\Rightarrow\quad p(A_1/B) = \frac{p(A_1)\cdot p(B/A_1)}{\sum_{i=1}^{3} p(A_i)\cdot p(B/A_i)} = \frac{\frac{1}{3}\cdot 1}{\frac{1}{3}\cdot 1 + \frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot 0} = \frac{\frac{1}{3}}{\frac{1}{3}+\frac{1}{6}} = \frac{2}{3}$$

4. ZADACI

Zadatak 1. Zamislimo istovremeno bacanje dviju kocki. Neka je A događaj da obje kocke padnu na broj 1, B da obje pokažu isti broj, C da suma brojeva na kockama bude 5, D da razlika brojeva na kockama bude 3, E da suma brojeva na kockama bude 1, F sa suma brojeva na kockama bude manja od 20. Izračunajte pripadne vjerojatnosti.

Rješenje:

Skup elementarnih događaja je $\Omega = \{(1,1), \ldots, (1,6), (2,1), \ldots, (6,6)\}$, pa je broj elemenata tog skupa, tj. broj mogućih ishoda pokusa, $n = k\Omega = 6^2 = 36$ (varijacije s ponavljanjem drugog razreda skupa od 6 elemenata). Označimo sa $m(\mathrm{A})$ broj povoljnih ishoda za događaj A. Tada imamo:

$m(\mathrm{A}) = k\{(1,1)\} = 1 \quad \Rightarrow \quad p(A) = \dfrac{m(A)}{n} = \dfrac{1}{36}$

$m(\mathrm{B}) = k\{(1,1), (2,2), \ldots, (6,6)\} = 6 \quad \Rightarrow \quad p(B) = \dfrac{m(B)}{n} = \dfrac{6}{36} = \dfrac{1}{6}$

$m(\mathrm{C}) = k\{(1,4), (2,3), (3,2), (4,1)\} = 4 \quad \Rightarrow \quad p(C) = \dfrac{m(C)}{n} = \dfrac{4}{36} = \dfrac{1}{9}$

$m(\mathrm{D}) = k\{(1,4), (2,5), (3,6), (4,1), (5,2), (6,3)\} = 6 \quad \Rightarrow \quad p(D) = \dfrac{m(D)}{n} = \dfrac{6}{36} = \dfrac{1}{6}$

$m(\mathrm{E}) = k\emptyset \quad \Rightarrow \quad p(E) = 0$ (jasno, jer je E nemoguć događaj)

$m(\mathrm{F}) = k\Omega = 36 \quad \Rightarrow \quad p(F) = 1$ (jasno, jer je F siguran događaj).

Iz ovih rezultata možemo znati da li se više isplati kladiti na neki događaj A nego na neki događaj B, a to spada u hazardne igre.

Zadaci 2. Bacaju se tri kocke. Kolika je vjerojatnost da:

a) sve pokažu različite brojeve,
b) zbroj brojeva bude 9,
c) zbroj brojeva bude 4.

Rješenje:

Ovdje je $\Omega = \{(1,1,1), (1,1,2), \ldots, (6,6,6)\}$. Njegovi elementi su varijacije s ponavljanjem trećeg razreda skupa od 6 elemenata, pa je $k\Omega = \overline{V}_3(6) = 6^3 = 216$.

a) $m(A) = V_3(6) = \dfrac{6!}{(6-3)!} = 4 \cdot 5 \cdot 6 \quad \Rightarrow \quad p(A) = \dfrac{120}{216} = \dfrac{5}{9}$

b) Povoljni slučajevi su:

$1\ 2\ 6 \quad \xrightarrow{permutacije} \quad 3! = 6$

$1\ 3\ 5 \quad \xrightarrow{permutacije} \quad 6$

$1\ 4\ 4 \quad \xrightarrow{permutacije} \quad 3$

$2\ 3\ 4 \quad \xrightarrow{permutacije} \quad 6$

$2\ 5\ 2 \quad \xrightarrow{permutacije} \quad 3$

$3\ 3\ 3 \quad\quad\quad\quad\quad\quad 1$

Dakle, povoljnih slučajeva ima 25 $\Rightarrow \quad p(B) = \frac{25}{216}$

c) $p(C) = \frac{3}{216}$

Zadatak 3. U kutiji se nalazi 11 bijelih i 5 crnih kuglica. Ako izvučemo 2 kuglice kolika je vjerojatnost da će biti iste boje?

Rješenje:

A = izvučene obje bijele kuglice, B = izvučene obje crne kuglice

Vrijedi $A \cap B = \emptyset$, pa se ovi događaji međusobno isključuju. Stoga vrijedi:

$$p(A \cup B) = p(A) + p(B) = \frac{\binom{11}{2}}{\binom{16}{2}} + \frac{\binom{5}{2}}{\binom{16}{2}} = \frac{13}{24}$$

Zadatak se može riješiti i drugi način pomoću suprotnog događaja:

$(A \cup B)^C = \overline{A \cup B} =$ različitih su boja: $p(\overline{A \cup B}) = \frac{11 \cdot 5}{\binom{16}{2}} = \frac{11}{24} \qquad \Rightarrow$

$$p(A \cup B) = 1 - p(\overline{A \cup B}) = 1 - \frac{11}{24} = \frac{13}{24}.$$

Zadatak 4. U kutiji se nalazi 7 istovrsnih proizvoda, 4 dobra i 3 loša. Iz kutije se nasumce izvlače 4 proizvoda. Kolika je vjerojatnost da će:

a) sva četiri biti dobra,

b) 2 biti dobra, a 2 loša,

c) najviše 2 biti loša.

Rješenje:

$S=\{d_1,d_2,d_3,d_4,l_1,l_2,l_3\}$. Skup elementarnih događaja Ω sastoji se od svih mogućih izbora od 4 proizvoda iz skupa od 7 proizvoda. Tih izbora ima $K_4(7)=\binom{7}{4}=35$, pa vrijedi $n=$ kΩ=35.

a) $p=\dfrac{m}{n}=\dfrac{1}{35}$

b) $p=\dfrac{m}{n}=\dfrac{\binom{4}{2}\binom{3}{2}}{35}=\dfrac{6\cdot 3}{35}=\dfrac{18}{35}$

c) Elementarni događaji koji su nam u ovom slučaju povoljni su: 4 dobra ili 3 dobra,1 loš ili 2 dobra, 1 loš. Oni se isključuju, pa vrijedi:

$$p=\frac{1}{35}+\frac{\binom{4}{3}\binom{3}{1}}{35}+\frac{\binom{4}{2}\binom{3}{2}}{35}$$

Zadatak 5. Iz skupa S = $\{1,2,3,...,10\}$ nasumce izaberemo 6 brojeva. Kolika je vjerojatnost da bude izabrano 6 uzastopnih brojeva?

Rješenje:

$$n=\binom{10}{6},\ m=k\{\{1,2,3,4,5,6\},\{2,3,4,5,6,7\},\ldots,\{5,6,7,8,9,10\}\}=5 \Rightarrow p=\frac{5}{\binom{10}{6}}$$

Zadatak 6. Brojevi 1, 2, 3, 4, 5 i 6 poredaju se slučajnim poretkom u niz. Neka je A događaj da brojevi 1, 2 budu na svom prirodnom mjestu, a B događaj da brojevi 1, 2 budu jedan do drugoga. Nađi p(A) i p(B).

Rješenje:

$\underline{\underline{1,2}},3,4,5,6 \quad \Rightarrow \quad p(A)=\dfrac{4!}{6!}$

$p(B)=\dfrac{4!\cdot 2\cdot 5}{6!}=\dfrac{1}{3}$ (imamo 2 mogućnosti: 1,2 ili 2,1; imamo 5 mogućnosti: 1,2 mogu biti na 1. i 2. mjestu, ili na 2. i 3., ili na 3. i 4., ili na 4. i 5. ili na 5. i 6. mjestu)

Zadatak 7. Zadan je skup brojeva S = $\{1,2,3,...,8\}$. Izvlačimo 3 broja, pa neka je A događaj da je 3 najmanji od izvučenih brojeva. Izvlačimo 2 broja, pa neka je B događaj da je jedan od njih manji, a drugi veći od broja 5.

Rješenje:

$\underline{\underline{3}},4,5,6,7,8 \quad \Rightarrow \quad p(A) = \dfrac{\binom{5}{2}}{\binom{8}{3}}$

$1,2,3,4,\underline{\underline{5}},6,7,8 \quad \Rightarrow \quad p(B) = \dfrac{4 \cdot 3}{\binom{8}{2}}$

Zadatak 8. Bacamo novčić i kocku. Kolika je vjerojatnost da će novčić pasti na grb i kocka na broj 6?

Rješenje:

Tražimo da se dogode dva događaja A i B koji su nezavisni (jer nastupanje jednog ne utječe na nastupanje drugog). Stoga je

$$p(A \cap B) = p(A) \cdot p(B) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}$$

(ili $\Omega = \{P1, P2, \ldots, G6\} \quad \Rightarrow \quad p(A \cap B) = \dfrac{m}{n} = \dfrac{1}{12}$)

Zadatak 9. Kolika je vjerojatnost da ćemo bacajući kocku broj 6 dobiti tek pri trećem bacanju?

Rješenje:

$$p(A \cap B \cap C) = p(A) \cdot p(B) \cdot p(C) = \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6}$$

Zadatak 10. Iz skupa $\{1,2,3,...,100\}$ bira se jedan broj. Kolika je vjerojatnost da će taj slučajno izabrani broj biti oblika m^n ($m,n > 1$)?

Rješenje:

$p = \dfrac{m}{n}$, $n = 100$; Povoljni slučajevi su brojevi:

$2^2, 2^3, 2^4, 2^5, 2^6; 3^2, 3^3, 3^4, (4^2), 4^3, 5^2, 6^2, 7^2, (8^2, 9^2), 10^2$

Dakle, $m = 12$, pa je $p = \frac{12}{100}$.

Zadatak 11. Kolika je vjerojatnost da od 4 bačene kocke bar jedna pokaže broj 6?

Rješenje:

Neka je A događaj da bar jedna pokaže broj 6. Tada je $A^C = \overline{A}$ događaj da nijedna ne pokaže broj 6.

$$p(\overline{A}) = \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} = \left(\frac{5}{6}\right)^4 \qquad \Rightarrow \qquad p(A) = 1 - p(\overline{A}) = \frac{671}{1296} \approx 0.518$$

Iz ovog rješenja slijedi: <u>ako događaju A u nekom pokusu pripada vjerojatnost p i ako se napravi serija od k tih pokusa, onda vrijedi:</u>

a) vjerojatnost da se A dogodi k puta u toj seriji je $P = p^k$;

b) vjerojatnost da se A ne pojavi ni jednom u toj seriji je $P = (1-p)^k = q^k$;

c) vjerojatnost da se A pojavi bar jednom u toj seriji je $P = 1 - q^k$.

Zadatak 12. Imamo 2 špila igraćih karata, pa iz svakog izvlačimo jednu kartu. Kolika je vjerojatnost da će bar jedna od tih karata biti as špadi?

Rješenje:

Vjerojatnost da nećemo ni iz 1. ni iz 2. špila izvući tu kartu je $\frac{39}{40} \cdot \frac{39}{40}$ (jer su to nezavisni događaji), pa je tražena vjerojatnost $p = 1 - \left(\frac{39}{40}\right)^2 = \frac{79}{1600}$.

<u>2. način</u>: A je događaj da ćemo iz 1. špila izvući tu kartu $\Rightarrow$ $p(A) = \frac{1}{40}$

B je događaj da ćemo iz 2. špila izvući tu kartu $\Rightarrow$ $p(B) = \frac{1}{40}$

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) = p(A) + p(B) - p(A) \cdot p(B) = \frac{1}{40} + \frac{1}{40} - \frac{1}{40} \cdot \frac{1}{40} = \frac{79}{1600}$$

<u>3. način</u>: $p = \frac{m}{n} = \frac{39 + 39 + 1}{40^2} = \frac{79}{1600}$

Zadatak 13. Kocka se baca 12 puta. Nađi vjerojatnost da broj 6 padne:

a) samo 1. i 2. put,

b) točno 2 puta,

c) najviše 2 puta.

Rješenje:

a) $p(A)=\frac{1}{6}\cdot\frac{1}{6}\cdot\left(\frac{5}{6}\right)^{10}$

b) $p(B)=\frac{1}{6}\cdot\frac{1}{6}\cdot\left(\frac{5}{6}\right)^{10}+\cdots+\left(\frac{5}{6}\right)^{10}\cdot\frac{1}{6}\cdot\frac{1}{6}=\binom{12}{2}\cdot\left(\frac{1}{6}\right)^{2}\cdot\left(\frac{5}{6}\right)^{10}$

c) $p(C)=p(0puta)+p(1put)+p(2put)=\left(\frac{5}{6}\right)^{12}+\frac{1}{6}\cdot\left(\frac{5}{6}\right)^{11}\cdot\binom{12}{1}+\left(\frac{1}{6}\right)^{2}\cdot\left(\frac{5}{6}\right)^{10}\cdot\binom{12}{2}$.

Zadatak 14. U 1. kutiji nalazi se 10 bijelih i 3 crne kuglice, a u 2. kutiji 3 bijele i 5 crnih kuglica. Prenesemo 2 kuglice iz 1. u 2. kutiju, a zatim iz 2. kutije izvučemo jednu kuglicu. Kolika je vjerojatnost da ta kuglica bude bijela?

Rješenje:

$$p=\underbrace{\frac{\binom{10}{2}}{\binom{13}{2}}\cdot\frac{5}{10}}_{\text{ako su 2B prešle u 2. kutiju}}+\underbrace{\frac{\binom{10}{1}\binom{3}{1}}{\binom{13}{2}}\cdot\frac{4}{10}}_{\text{ako je 1B i 1C prešle u 2. kutiju}}+\underbrace{\frac{\binom{3}{2}}{\binom{13}{2}}\cdot\frac{3}{10}}_{\text{ako su 2C prešle u 2. kutiju}}=\frac{59}{130}$$

Zadatak 15. Istovremeno se baca 6 novčića. Kolika je vjerojatnost da:

a) točno 3 novčića padnu na pismo,

b) barem 3 novčića padnu na pismo?

Rješenje:

a) $p=\left(\frac{1}{2}\right)^{3}\cdot\left(\frac{1}{2}\right)^{3}\cdot\binom{6}{3}$

b)

$$p=\left(\frac{1}{2}\right)^{3}\cdot\left(\frac{1}{2}\right)^{3}\cdot\binom{6}{3}+\left(\frac{1}{2}\right)^{4}\cdot\left(\frac{1}{2}\right)^{3}\cdot\binom{6}{4}+\left(\frac{1}{2}\right)^{6}\cdot\binom{6}{5}+\left(\frac{1}{2}\right)^{6}\cdot\binom{6}{6}=\left(\frac{1}{2}\right)^{6}(20+15+6+1)=\frac{42}{64}=\frac{21}{32}$$

Zadatak 16. U kutiji se nalazi 5 dobrih i 3 loša proizvoda.

a) Ako odjednom izvučemo 3 proizvoda, kolika je vjerojatnost da će biti 2D i 1L?

b) Kolika je vjerojatnost da u 3 uzastopna izvlačenja bez vraćanja dobijemo rezultat LDD ili DLD?

Rješenje:

a) $p = \dfrac{\binom{5}{2} \cdot \binom{3}{1}}{\binom{8}{3}}$

b) $p = p_1 + p_2 = \overset{L}{\dfrac{3}{8}} \cdot \overset{D}{\dfrac{5}{7}} \cdot \overset{D}{\dfrac{4}{6}} + \overset{D}{\dfrac{5}{8}} \cdot \overset{L}{\dfrac{3}{7}} \cdot \overset{D}{\dfrac{4}{6}} = \dfrac{5}{14}$

$$p(L \cap D_1 \cap D_2) = p(L) \cdot p(D_1 / L) \cdot p(D_2 / L \cap D_1)$$

Zadatak 17. Student je došao na ispit znajući 84 od 102 pitanja. Na ispitu se izvlači papir s 3 pitanja. Kolika je vjerojatnost da će student znati sva 3 pitanja?

Rješenje:

A = student će znati 1. pitanje

B = student će znati 2. pitanje nakon što je znao 1. pitanje

C = student će znati 3. pitanje nakon što je znao 1. i 2. pitanje

$$p(A \cap B \cap C) = p(A) \cdot p(B / A) \cdot p(C / A \cap B) = \frac{84}{102} \cdot \frac{83}{101} \cdot \frac{82}{100} \approx 0.55$$

Zadatak 18. Iz skupa $\{1,2,3,...,100\}$ bira se jedan broj. Kolika je vjerojatnost da bude djeljiv sa 2 ili sa 3?

Rješenje:

Označimo sa $[x]$ najveći cijeli broj manji ili jednak od x. Npr. $\left[\frac{17}{3}\right] = 5, [11.2] = 11$. Tada je iz zadanog skupa $\left[\frac{100}{2}\right] = 50$ brojeva djeljivo sa 2, $\left[\frac{100}{3}\right] = 33$ broja djeljiva sa 3, a $\left[\frac{100}{6}\right] = 16$ brojeva djeljivo sa 2 i 3, tj. 6. Traženo rješenje je

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) = \frac{50}{100} + \frac{33}{100} - \frac{16}{100} = \frac{67}{100}$$

III. EMPIRIČKE DISTRIBUCIJE

1. DISTRIBUCIJA FREKVENCIJA

Statistika se bavi skupljanjem, organiziranjem i analiziranjem podataka vezanim za masovne pojave, odnosno pojave koje se javljaju na nekom skupu istovrsnih elemenata, tzv. statističkom skupu. Danas gotovo da nema znanosti koja ne koristi statistiku: ekonomija, medicina, psihologija, sociologija, fizika, meteorologija itd.

Statistički skup je skup čiji elementi imaju neko zajedničko obilježje čija vrijednost varira od elementa do elementa. Elementi tog skupa mogu biti osobe, predmeti, događaji, pojave, itd. Ako je na tom skupu definirano ili promatrano samo jedno obilježje, kažemo da je **skup jednodimenzionalan**. Analogno se definira dvodimenzionalan skup, … Prema karakteru obilježja elemenata, statističke skupove dijelimo u dvije klase:

a) **kontinuirane** – obilježje, nazovimo ga *x*, može poprimiti bilo koju vrijednost iz nekog segmenta $[a,b]$,
b) **diskontinuirane ili diskretne** – obilježje *x* može poprimiti vrijednost samo iz diskretnog niza vrijednosti: $x_1, x_2, x_3, ...$

Neka je zadan neki statistički skup i neka je *x* promatrano obilježje elemenata tog skupa. Pretpostavimo da smo mjerenjem izračunali vrijednost obilježje *x* na svakom elementu tog skupa. Tada kažemo da imamo **empiričke podatke**.

Prvi korak u sređivanju statističkog skupa, odnosno empiričkih podataka je grupiranje podataka u **razrede** i računanje **frekvencija razreda**. Podatke možemo grupirati na dva načina:

- **prvi način** je da sve elemente koji imaju istu vrijednost obilježja *x* stavimo u isti razred,
- **drugi način** je da u isti razred stavimo elemente koji imaju više uzastopnih vrijednosti obilježja *x*. Na taj način dobivamo manji broj razreda, a to činimo kad je područje rasipanja vrijednosti obilježja *x* veliko.

Broj elemenata statističkog skupa koji pripadaju istom razredu zove se **frekvencija tog razreda**. Ako se frekvencija podijeli ukupnim brojem elemenata statističkog skupa dobiva se **relativna frekvencija razreda**.

Ako smo izvršili podjelu statističkog skupa na razrede i izračunali pripadne frekvencije, kažemo da smo napravili **distribuciju (razdiobu) frekvencija**. Distribucija frekvencija prikazuje se obično u obliku: $x_1, f_1; x_2, f_2; \ldots, x_n, f_n$ ili u obliku tablice:

x_i	f_i

Primjer 1. Uspjeh razreda od 20 učenika je

3	4	1	2	5	3	1	4	2	3
4	2	3	2	3	2	5	2	1	2

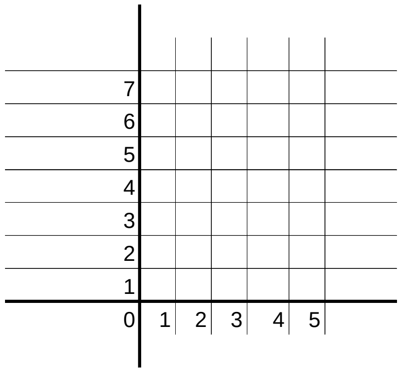
Statistički skup je dakle taj skup od 20 učenika, a promatrano obilježje *x* na tom skupu, uspjeh učenika. Dakle, *x* može poprimiti vrijednosti samo iz skupa {1,2,3,4,5}, pa je taj skup diskretnog (diskontinuiranog) karaktera. Napravimo distribuciju frekvencija (primijenit ćemo prvi način jer je broj razreda mali).

x_i	f_i	f_{ri}
1	3	$3/20$
2	7	$7/20$
3	5	$5/20$
4	3	$3/20$
5	2	$2/20$
	20 = N	

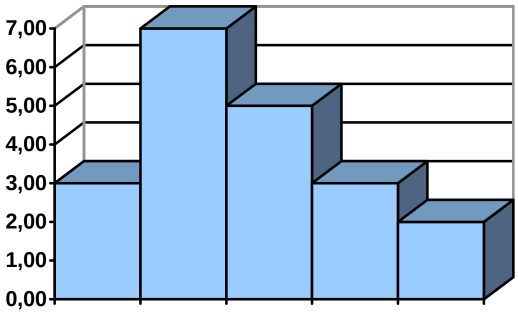
$N = \sum f_i =$ ukupni broj podataka

$\sum f_{ri} = 1$

Distribuciju frekvencija može se (zorno) prikazati i pomoću **poligona frekvencija**:



ili **histograma frekvencija**:



Iz ovog primjera vidimo značenje relativne frekvencije f_{ri}: to je vjerojatnost da slučajni uzorak pripada razredu x_i.

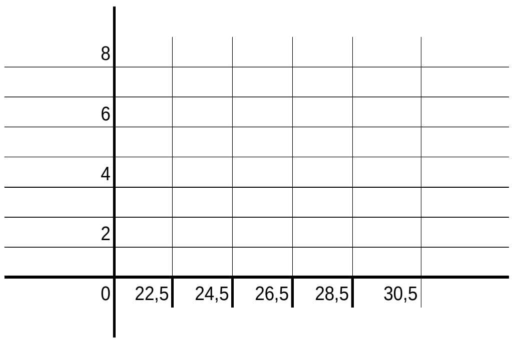
Primjer 2. Mjereći vrijeme izvođenja jedne radne operacije dobiveni su slijedeći podaci u sekundama:

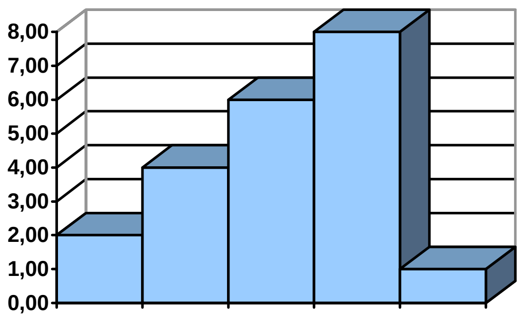
24	28	22	26	24
27	26	25	26	23
30	26	29	25	27
24	26	25	24	27

Najmanja vrijednost je 22, a najveća 30 sekundi. Međutim, ta operacija može trajati bilo koje vrijeme između 22 i 30 sekundi (pa i izvan tog intervala), pa je taj skup kontinuiranog karaktera. Napravimo distribuciju frekvencija na drugi način:

Razredi	Sredina razreda x_i	f_i
22-23	22.5	2
24-25	24.5	7
26-27	26.5	8
28-29	28.5	2
30-31	30.5	1
		20

Da bismo dobili zorniju sliku distribucije frekvencija nacrtajmo poligon frekvencije i histogram:





Ovdje smo za širinu razreda svakog razreda odabrali 2 sekunde. Stvarne granice između razreda se ne vide iz ove tablice. Npr. granica između prvog i drugog razreda je 23.5, između drugog i trećeg je 25.5, pa je tako drugi razred interval (23.5,25.5).

Svaki razred može se reprezentirati njegovom sredinom x_i (drugi stupac tablice).

2. ARITMETIČKA SREDINA I STANDARDNA DEVIJACIJA

Kod proučavanja velikog broja podataka važno je odrediti njihovu prosječnu vrijednost. Pod prosječnom vrijednošću obično se podrazumijeva aritmetička sredina. **Aritmetička sredina** *n* brojeva $x_1, x_2, \dots, x_n$ je broj

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

Ako je izrađena distribucija frekvencija $x_1, f_1; x_2, f_2; \dots, x_n, f_n$ nekog statističkog skupa, onda se aritmetička sredina tog skupa računa po formuli $\overline{x} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i$, gdje je $N = \sum_{i=1}^{n} f_i$ (tj. N je broj elemenata tog skupa). Aritmetičku sredinu možemo izraziti i pomoću relativne frekvencije f_{ri}. Budući da je po definiciji $f_{ri} = \frac{f_i}{N}$, vrijedi $\overline{x} = \sum_{i=1}^{n} f_{ri} x_i$.

Aritmetičku sredinu često računamo jednim drugim izrazom, posebno kada je x_i decimalan broj. Označimo sa x_0 (tzv. nulti razred) razred koji ima **najveću frekvenciju**, a sa *c* širinu razreda (*c* je konstantan za sve razrede). Uvedimo sada novu varijablu d_i relacijom $x_i = c \cdot d_i + x_0$ (tj. $d_i = \frac{x_i - x_0}{c}$). Tada imamo:

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{n} f_i (cd_i + x_0) = \frac{1}{N}\left(c\sum_{i=1}^{n} f_i d_i + x_0 \sum_{i=1}^{n} f_i \right) = x_0 + \frac{c}{N}\sum_{i=1}^{n} f_i d_i .$$

U obradi statističkih podataka pored aritmetičke sredine koriste se i neke druge veličine. Da bismo na neki način izmjerili rasipanje podataka x_i oko aritmetičke sredine $\overline{x}$, uvedimo razlike $x_1 - \overline{x}$, $x_2 - \overline{x}, \dots,$ $x_n - \overline{x}$ (to su dakle odstupanja podataka $x_1, x_2, \dots, x_n$ od aritmetičke sredine $\overline{x}$). Vrijedi: $\sum_{i=1}^{n} f_i (x_i - \overline{x}) = \sum_{i=1}^{n} f_i x_i - N\overline{x} = 0$, tj. zbroj svih odstupanja podataka jednak je nuli. Stoga se umjesto odstupanja $x_i - \overline{x}$, promatraju njihovi kvadrati $\left(x_i - \overline{x}\right)^2$ (kako bi se izbjeglo poništavanje negativnih sa pozitivnim odstupanjima), i definiraju slijedeće veličine.

Varijanca σ^2 statističkog skupa od N elemenata je broj $\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \overline{x})^2$ (*) , a **standardna devijacija** tog skupa je broj $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \overline{x})^2}$. Te veličine su važne jer karakteriziraju odstupanje, tj. rasipanje vrijednosti obilježja x_i oko aritmetičke sredine $\overline{x}$. Može se pokazati da je to rasipanje veće što je veća standardna devijacija σ (npr. gotovo sve vrijednosti od x_i za najvažnije, tzv. normalne distribucije, padaju u interval $(\overline{x} - 2\sigma, \overline{x} + 2\sigma)$.

Ako izraz na desnoj strani od σ^2 kvadriramo i sredimo, dobivamo prikladan izraz za računanje varijance:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i x_i^2 - \overline{x}^2 \quad (**)$$

Ako su pak u napravljenoj distribuciji frekvencija statističkog skupa podaci grupirani u razrede širine *c*, onda su x_i sredine pojedinih razreda, *c* širina razreda, x_0 nulti razred i $x_i = c \cdot d_i + x_0$. Tada je izraz za varijancu:

$$\sigma^2 = c^2 \left[\frac{1}{N}\sum_{i=1}^{n} f_i d_i^2 - \left(\frac{1}{N}\sum_{i=1}^{n} f_i d_i \right)^2 \right] \quad (***)$$

Izračunajmo aritmetičku sredinu i varijancu u prethodna dva primjera:

Primjer 1.

x_i	f_i	$f_i\ x_i$	d_i	$f_i\ d_i$	$f_i\ x_i^2$
1	3	3	-1	-3	3
2	7	14	0	0	28
3	5	15	1	5	45
4	3	12	2	6	48
5	2	10	3	6	50
	20 = N	54		14	174

$\overline{x} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i = \frac{1}{20} \cdot 54 = 2.7$ ILI $\overline{x} = x_0 + \frac{c}{N}\sum_{i=1}^{n} f_i d_i = 2 + \frac{1}{20} \cdot 14 = 2.7$

$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i x_i^2 - \overline{x}^2 = \frac{1}{20} \cdot 174 - 2.7^2 = 1.41$, $\sigma = 1.19$

Primjer 2.

Razredi	Sredina razreda x_i	f_i	d_i	$f_i\ d_i$	$f_i\ d_i^{\ 2}$	$x_i^{\ 2}$	$f_i\ x_i^{\ 2}$
22-23	22.5	2	-2	-4	8	506.25	1012.5
24-25	24.5	7	-1	-7	7	600.25	4201.75
26-27	26.5	8	0	0	0	702.25	5618
28-29	28.5	2	1	2	2	812.25	1624.5
30-31	30.5	1	2	2	4	930.25	930.25
		20=N		-7	21		13 387

$$\overline{x} = x_0 + \frac{c}{N}\sum_{i=1}^{n} f_i d_i = 26.5 + \frac{2}{20}(-7) = 25.8$$

$$\sigma^2 = c^2\left[\frac{1}{N}\sum_{i=1}^{n} f_i d_i^{\ 2} - \left(\frac{1}{N}\sum_{i=1}^{n} f_i d_i\right)^2\right] = 2^2\left[\frac{1}{20}\cdot 21 - \left(\frac{1}{20}(-7)\right)^2\right] = 4\left[1.05 - 0.35^2\right] = 3.71$$

$$\sigma = 1.926$$

NAPOMENA: Ako su podaci grupirani u intervale razreda, kao u ovom primjeru, tada dobivamo **aproksimativnu** vrijednost aritmetičke sredine i varijance (npr. umjesto sa 30.5 u zadnjem razredu trebali bi raditi sa 30 – jer imamo samo jedan podatak za 30). Međutim moguća pogreška je malena, pa preko nje prelazimo.

Primjer 3. Empirički podaci o broju šibica x_i u kutijama dani su u slijedećoj tablici:

x_i	43-44	45-46	47-48	49-50	51-52	53-54	55-56	57-58
f_i	4	6	16	22	24	18	6	4

Izračunajte $\overline{x}$, σ^2 i σ.

Razredi	x_i	f_i	d_i	$f_i\, d_i$	$f_i\, d_i^{\,2}$
43-44	43.5	4	-4	-16	64
45-46	45.5	6	-3	-18	54
47-48	47.5	16	-2	-32	64
49-50	49.5	22	-1	-22	22
51-52	51.5	24	0	0	0
53-54	53.5	18	1	18	18
55-56	55.5	6	2	12	24
57-58	57.5	4	3	12	36
		100		-46	282

$$N = 100,\quad x_0 = 51.5,\quad c = 2,\quad d_i = \frac{x_i - x_0}{c} = \frac{x_i - 51.5}{2} \quad \Rightarrow \quad \bar{x} = x_0 + \frac{c}{N}\sum_{i=1}^{n} f_i d_i = 50.58$$

$$\sigma^2 = c^2\left[\frac{1}{N}\sum_{i=1}^{n} f_i d_i^{\,2} - \left(\frac{1}{N}\sum_{i=1}^{n} f_i d_i\right)^2\right] = 2^2\left[\frac{1}{100}\cdot 282 - \left(\frac{1}{100}\cdot(-46)\right)^2\right] = 10.4336,\quad \sigma = 3.23$$

Zadatak. Izvjesna mjerenja dala su slijedeće podatke izražene u kilogramima:

x_i	4.92-4.94	4.95-4.97	4.98-5.00	5.01-5.03	5.04-5.06	5.07-5.08
f_i	2	7	14	10	6	3

Izračunajte $\bar{x}$ i σ. Nacrtajte poligon frekvencija i histogram.

(*Rješenje:* N = 42, $x_0 = \dfrac{4.98 + 5.00}{2} = 4.99$, $\bar{x} = 5.004$, $\sigma^2 = 0.0014238$, $\sigma = 0.0377$)

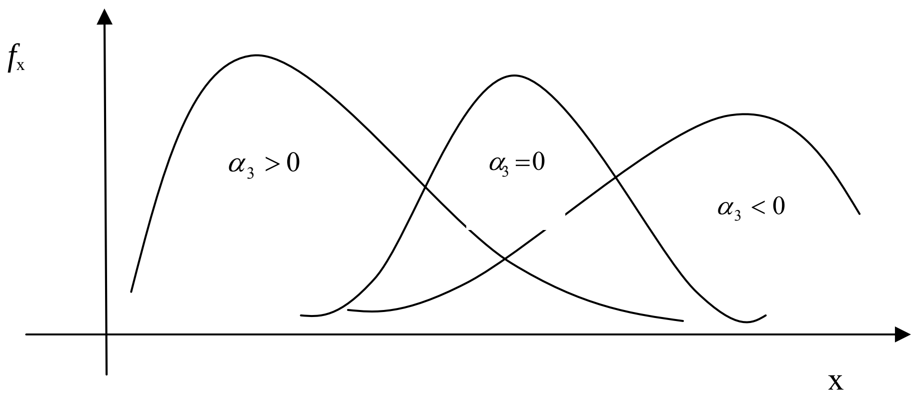
3. MOMENTI

Da bi se dobila potpunija slika o svojstvima distribucije frekvencija uvode se još neke veličine.

Centralni moment reda *r* je broj definiran relacijom $M_r = \frac{1}{N}\sum_i f_i (x_i - \overline{x})^r$.

Lako se vidi da uvijek vrijedi da je $M_0 = 1$ i $M_1 = 0$, pa ova dva momenta nisu važna, dok je $M_2 = \sigma^2$ varijanca o kojoj smo zbog njene važnosti posebno govorili. Posebno su važni momenti M_3 i M_4 pomoću kojih se definiraju **koeficijent asimetrije** α_3 i **koeficijent spljoštenosti** α_4: $\alpha_3 = \frac{M_3}{\sigma^3}$, $\alpha_4 = \frac{M_4}{\sigma^4}$.

Razdiobe najčešće u praksi imaju svojstvo da im je asimetrija oko točke $\overline{x}$ to veća što im je koeficijent α_3 veći, što se može vidjeti na slici:



Koeficijent α_4 služi kao mjera spljoštenosti distribucije frekvencija. Ako je $\alpha_4 = 3$ kaže se da je distribucija frekvencija normalno spljoštena.

Pomoćni moment reda *r* definiran je relacijom $m_r = \frac{1}{N}\sum_i f_i x_i^{\,2}$.

Jasno je da za svaku distribuciju vrijedi da je $m_0 = 1$ i $m_1 = \overline{x}$. Nađimo vezu između centralnih i pomoćnih momenata:

$$M_r = \frac{1}{N}\sum_i f_i (x_i - \overline{x})^r = \frac{1}{N}\sum_i f_i (x_i - m_1)^r = \frac{1}{N}\sum_i f_i \sum_{k=0}^{r} (-1)^k \binom{r}{k} x_i^{\,r-k} m_1^{\,k}$$

$$= \sum_{k=0}^{r} (-1)^k \binom{r}{k} m_1^{\,k} \left(\frac{1}{N}\sum_i f_i x_i^{\,r-k} \right) = \sum_{k=0}^{r} (-1)^k \binom{r}{k} m_{r-k} m_1^{\,k}\,.$$

$$\Rightarrow \quad M_2 = \sigma^2 = (-1)^0 \binom{2}{0} m_{2-0} m_1{}^0 + (-1)^1 \binom{2}{1} m_{2-1} m_1{}^1 + (-1)^2 \binom{2}{2} m_{2-2} m_1{}^2 =$$

$$= m_2 - 2m_1 m_1 + m_0 m_1{}^2 = m_2 - m_1{}^2 \quad \text{(ovo je već poznata formula za varijancu)}$$

$$M_2 = m_2 - m_1{}^2$$

$$M_3 = m_3 - 3m_1 m_2 + 2m_1{}^3$$

$$M_4 = m_4 - 4m_1 m_3 + 6m_1{}^2 m_2 - 3m_1{}^4$$

Pomoću ovih formula su centralni momenti izraženi pomoću pomoćnih momenata koji se lakše računaju.

IV. SLUČAJNE VARIJABLE

1. DISKRETNE SLUČAJNE VARIJABLE

U praksi je korisno slučajni ishod nekog eksperimenta okarakterizirati brojem. Tada se promjenjiva veličina koja poprima te brojčane vrijednosti s određenom vjerojatnošću zove **slučajna varijabla**.

Diskretna ili **diskontinuirana slučajna varijabla** je varijabla x koja poprima vrijednosti $x_1, x_2, x_3, \ldots$ s vjerojatnostima $p(x_1), p(x_2), p(x_3), \ldots$ s tim da vrijedi $\sum_i p(x_i) = 1$.

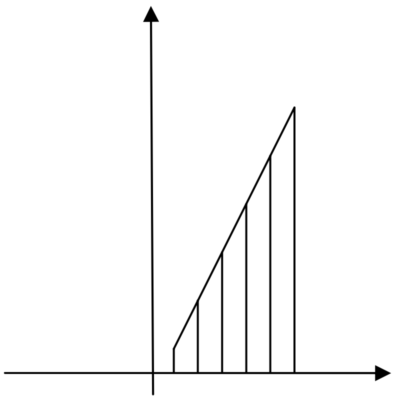
Skup vrijednosti slučajne varijable može biti konačan ili prebrojivo beskonačan skup. Skup svih uređenih parova $(x_i, p(x_i))$, i = 1,2,… zove se **distribucija** ili **razdioba slučajne varijable** x, a funkcija $p(x_i)$ **zakon** ili **funkcija vjerojatnosti slučajne varijable** x.

Funkcija $F(x_k) = \sum_{i=1}^{k} p(x_i)$ zove se **funkcija distribucije slučajne varijable** x. Ona nam, dakle, kaže kolika je vjerojatnost da varijabla x poprimi bilo koju vrijednost manju od x_k ili jednaku od x_k, tj. $F(x_k) = p(x \le x_k)$.

<u>Primjer 1</u>. Neka se bacaju dvije kocke i neka je x veći od brojeva koji pokazuju te dvije kocke. Tada je x zaista slučajna varijabla: $x \in \{1,2,3,4,5,6\}$ i svakoj toj vrijednosti pripada određena vjerojatnost: $p(1) = \frac{kard\{(1,1)\}}{36} = \frac{1}{36}$, $p(2) = \frac{kard\{(1,2),(2,2),(2,1)\}}{36} = \frac{3}{36}$, $p(3) = \frac{5}{36}$, $p(4) = \frac{7}{36}$, $p(5) = \frac{9}{36}$, $p(6) = \frac{11}{36}$. Zaista vrijedi $\sum_{i=1}^{6} p(i) = \frac{36}{36} = 1$. Distribuciju ove slučajne varijable možemo prikazati tablicom:

x_i	1	2	3	4	5	6
$p(x_i)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

ili grafički:



Primjer 2. Neka se opet bacaju dvije kocke i neka je y suma brojeva koje pokazuju kocke. Analognim zaključivanjem pokazuje se da je y slučajna varijabla i da je njena distribucija:

x_i	2	3	4	5	6	7	8	9	10	11	12
$p(x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Funkcija vjerojatnosti slučajne varijable y je:

$$p(y) = \begin{cases} \frac{1}{36}y - \frac{1}{36}, & y = 2,3,4,5,6,7 \\ \frac{-1}{36}y + \frac{1}{36}, & y = 7,8,9,10,11,12 \end{cases} = \left| \frac{1}{36}y - \frac{1}{36} \right|$$

Za slučajnu varijablu definiraju se karakteristične veličine analogno kao i karakteristične veličine distribucije frekvencija empiričkih podataka, samo što se umjesto relativne frekvencije f_{ri} uzimaju vjerojatnosti $p(x_i)$, a umjesto aritmetičke sredine $\overline{x}$ uzima očekivanje μ. Ipak, navedimo ih posebno.

2. OČEKIVANJE, MOMENTI I VARIJANCE

Ako je $(x_i, p(x_i))$ distribucija slučajne varijable *x*, onda se broj $E(x) = \mu = \sum_i x_i \cdot p(x_i)$ zove **očekivanje slučajne varijable** *x*.

Budući da vjerojatnostima $p(x_i)$ kod slučajnih varijabli odgovaraju relativne frekvencije f_{ri} kod empiričkih distribucija, vidimo da je očekivanje $E(x) = \sum_i x_i \cdot p(x_i)$ analogan pojam pojmu **aritmetičke sredine** $\overline{x} = \sum_i x_i \cdot f_{ri}$. Naime, znamo da je relativna frekvencija f_{ri} empirička vjerojatnost da varijabla *x* poprimi vjerojatnost x_i.

Neka je zadana slučajna varijabla *x*, pa pomoću nje definirajmo novu slučajnu varijablu $ax + b$ gdje su a, b neki realni brojevi. Varijabla $ax + b$ poprima vrijednost $ax_i + b$ jedino onda kada varijabla *x* poprimi vrijednost x_i, a to znači da vrijednost $ax_i + b$ ima istu vrijednost kao i vrijednost x_i: $p(ax_i + b) = p(x_i)$.

***Teorem*.** Vrijedi $E(ax + b) = a \cdot E(x) + b$.

Dokaz:

$$E(ax+b) = \sum (ax_i + b) \cdot p(ax_i + b) = \sum (ax_i + b) \cdot p(x_i) = a \sum x_i \cdot p(x_i) + b \overbrace{\sum p(x_i)}^{=1} = a \cdot E(x) + b$$

Centralni moment *r*-tog reda diskretne slučajne varijable *x* je očekivanje varijable $(x - \mu)^r$, gdje je $\mu = E(x)$, tj. $M_r = E\left[(x - \mu)^r\right] = \sum_i (x_i - \mu)^r \cdot p(x_i)$.

Posebno, moment drugog reda: $M_2 = V(x) = \sigma^2 = E\left[(x - \mu)^2\right] = \sum_i (x_i - \mu)^2 \cdot p(x_i)$ zove se **varijanca** slučajne varijable *x*, a sam σ **standardna devijacija** slučajne varijable *x*.

Nešto kasnije pokazat ćemo da vrijedi: $V(x) = \sum_i x_i^2 \cdot p(x_i) - \mu^2$.

***Teorem*.** Varijanca slučajne varijable $ax + b$ je $V(ax + b) = a^2 V(x)$.

Dokaz:

$$V(ax+b)=E\{[(ax+b)-E(ax+b)]^2\}=E\{[(ax+b)-aE(x)-b]^2\}=E\{[ax-a\mu]^2\}=$$
$$=E\{a^2(x-\mu)^2\}=a^2E\{(x-\mu)^2\}=a^2V(x).$$

I kod slučajnih varijabli definira se:

- $\alpha_3=\dfrac{M_3}{\sigma^3}$ = koeficijent asimetrije,
- $\alpha_4=\dfrac{M_4}{\sigma^4}$ = koeficijent spljoštenosti,
- $m_r=E(x^r)=\sum_i x_i^{\ r}\cdot p(x_i)$ = pomoćni moment reda *r*.

Pomoćni momenti služe za lakše računanje centralnih momenata jer vrijedi (isto kao i kod momenata empiričke distribucije):

$$M_r=\sum_{k=0}^{r}(-1)^k\binom{r}{k}m_{r-k}m_1^{\ k}.$$

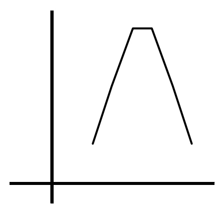
Posebno, $M_2=V(x)=m_2-m_1^{\ 2}=E(x^2)-\mu^2=\sum_i x_i^{\ 2}\cdot p(x_i)-\mu^2$, što je formula koja se češće koristi za računanje varijance.

Primjer 1. Zadana je diskontinuirana slučajna varijabla:

x	2	3	4	5	6	7
$p(x)$	$\frac{2}{30}$	$\frac{5}{30}$	$\frac{8}{30}$	$\frac{8}{30}$	$\frac{5}{30}$	$\frac{2}{30}$

a) Prikaži grafički distribuciju varijable *x*;
b) Izračunaj vjerojatnost $p(x\le 4)=F(4)$;
c) Prikaži grafički funkciju distribucije $F(x_i)$;
d) Izračunaj očekivanje i varijancu varijable *x*.

a)



b) $p(x \le 4) = p(2) + p(3) + p(4) = \frac{2}{30} + \frac{5}{30} + \frac{8}{30} = \frac{15}{30} = \frac{1}{2}$

c)

x_i	$F(x_i)$
2	$\frac{2}{30}$
3	$\frac{7}{30}$
4	$\frac{15}{30}$
5	$\frac{23}{30}$
6	$\frac{28}{30}$
7	$\frac{30}{30}$

d)

$$E(x) = \sum_i x_i \cdot p(x_i) = 2 \cdot \frac{2}{30} + 3 \cdot \frac{5}{30} + 4 \cdot \frac{8}{30} + 5 \cdot \frac{8}{30} + 6 \cdot \frac{5}{30} + 7 \cdot \frac{2}{30} = 4.5$$

$$V(x) = \sum_i x_i^{\,2} \cdot p(x_i) - \mu^2 = 4 \cdot \frac{2}{30} + 9 \cdot \frac{5}{30} + 16 \cdot \frac{8}{30} + 25 \cdot \frac{8}{30} + 36 \cdot \frac{5}{30} + 49 \cdot \frac{2}{30} - 4.5^2 = 1.717$$

Primjer 2. Događaj A pojavljuje se u jednom pokusu s vjerojatnošću $p = 0.3$. Neka je *x* broj nastupa događaja A u seriji od 5 pokusa.

a) Nađi zakon vjerojatnosti varijable *x*;
b) Izračunaj $p(x \le 3)$;
c) Izračunaj očekivanje i varijancu varijable *x*.

a) Iz teorije vjerojatnosti znamo da je vjerojatnost da taj događaj u seriji od 5 pokusa nastupi npr. 2 puta jednaka $\binom{5}{2} p^2 q^3$, odnosno da nastupi *x* puta $\binom{5}{x} p^x q^{5-x}$. Dakle,

$$p(x) = \binom{5}{x} p^x q^{5-x} = \binom{5}{x} 0.3^x \cdot 0.7^{5-x}.$$

b) $p(x \le 3) = p(0) + p(1) + p(2) + p(3) =$

$$= \binom{5}{0} 0.3^0 \cdot 0.7^5 + \binom{5}{1} 0.3^1 \cdot 0.7^4 + \binom{5}{2} 0.3^2 \cdot 0.7^3 + \binom{5}{3} 0.3^3 \cdot 0.7^2 = \ldots = 0.96922$$

c) $E(x) = \sum_i x_i \cdot p(x_i) = 0 \cdot p(0) + 1 \cdot p(1) + 2 \cdot p(2) + 3 \cdot p(3) + 4 \cdot p(4) + 5 \cdot p(5) =$

$$= 0 + 1 \cdot \binom{5}{1} 0.3^1 \cdot 0.7^4 + \cdots + \binom{5}{5} 0.3^5 \cdot 0.7^0 = \ldots = 1.5$$

$$V(x) = \sum_i x_i^{\,2} \cdot p(x_i) - \mu^2 = \cdots = 1.05$$

Primjer 3. U kutiji se nalazi 5 dobrih i 7 loših proizvoda. Iz kutije se izvlači slučajnim izborom uzorak od 4 proizvoda. Neka je x broj dobrih proizvoda u uzorku. Odredi zakon vjerojatnosti varijable x i prikaži grafički distribuciju.

$$x = 0,1,2,3,4$$

$$p(0) = \frac{\binom{5}{0}\binom{7}{4}}{\binom{12}{4}} = \frac{7}{99},$$

$$p(1) = \frac{\binom{5}{1}\binom{7}{3}}{\binom{12}{4}} = \frac{35}{99},$$

$$p(2) = \frac{42}{99}, \quad p(3) = \frac{14}{99}, \quad p(4) = \frac{1}{99}$$

$$\Rightarrow \quad p(x) = \frac{\binom{5}{x}\binom{7}{4-x}}{\binom{12}{4}}$$

Primjer 4. Zadan je niz brojeva: 1,2,3,4,5,6,7,8. Iz tog niza uzima se slučajnim izborom uzorak od 3 broja. Označimo li sa x najveći broj u uzorku tada je x slučajna varijabla.

a) Koje vrijednosti može poprimiti varijabla x;

b) Koje vjerojatnosti pripadaju pojedinim vrijednostima;

c) Kako glasi zakon vjerojatnosti varijable x;

d) Izračunaj očekivanje $E(x)$.

a) 3,4,5,6,7,8

b)

$$p(3)=\frac{1}{\binom{8}{3}}=\frac{1}{56},\ p(4)=\frac{\binom{3}{2}}{\binom{8}{3}}=\frac{3}{56},\ p(5)=\frac{\binom{4}{2}}{\binom{8}{3}}=\frac{6}{56}\ ,$$

$$p(6)=\frac{\binom{5}{2}}{\binom{8}{3}}=\frac{10}{56},\ p(7)=\frac{\binom{6}{2}}{\binom{8}{3}}=\frac{15}{56},\ p(8)=\frac{\binom{7}{2}}{\binom{8}{3}}=\frac{21}{56}$$

c)

$$p(x)=\frac{\binom{x-1}{2}}{\binom{8}{3}}$$

d)

$$E(x)=\sum_{i=3}^{8}x_i\,p(x_i)=3\cdot p(3)+4\cdot p(4)+\cdots+8\cdot p(8)=$$

$$=\frac{1}{56}(3+4\cdot 3+5\cdot 6+6\cdot 10+7\cdot 15+8\cdot 21)=\frac{378}{56}=6.75$$

Zadatak 1. Slučajna varijabla poprima vrijednosti $1,2,\ldots,n$ s vjerojatnošću koja je proporcionalna s vrijednošću varijable. Odredi funkciju vjerojatnosti.

Rješenje:

$$p(x_i)=kx_i,\quad p(1)+p(2)+\cdots+p(n)=1\ \Rightarrow\ k+2k+\cdots+nk=\frac{n(n+1)}{2}k=1\ \Rightarrow$$

$$k=\frac{2}{n(n+1)};\quad p(x_i)=\frac{2}{n(n+1)}x_i\,.$$

Zadatak 2. Jedna posuda sadrži 4 bijele i 3 crvene kuglice. Uzmemo nasumce uzorak od 3 kuglice. Neka je slučajna varijabla *x* broj bijelih kuglica u tom uzorku.

a) Kolike vjerojatnosti pripadaju brojevima $x = 0,1,2,3$

b) Kako glasi zakon vjerojatnosti varijable *x*;

c) Izračunaj očekivanje i varijancu varijable *x*.

Rješenje:

a) $x = 0,1,2,3$

x	0	1	2	3
$p(x)$	$\frac{1}{35}$	$\frac{12}{35}$	$\frac{18}{35}$	$\frac{4}{35}$

$$p(0) = \frac{\binom{3}{3}}{\binom{7}{3}} = \frac{1}{35}, \quad p(1) = \frac{\binom{4}{1}\binom{3}{2}}{\binom{7}{3}} = \frac{12}{35}, \quad p(2) = \frac{\binom{4}{2}\binom{3}{1}}{\binom{7}{3}} = \frac{18}{35}, \quad p(3) = \frac{\binom{4}{3}}{\binom{7}{3}} = \frac{4}{35}$$

b) $p(x) = \dfrac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}$

c) $E(x) = \sum_{i=3}^{8} x_i \, p(x_i) = \dfrac{12}{35} + \dfrac{36}{35} + \dfrac{12}{35} = \dfrac{60}{35} = \dfrac{12}{7}$

$$V(x) = \sum_i x_i^{\,2} \cdot p(x_i) - \mu^2 = \frac{12}{35} + \frac{4 \cdot 18}{35} + \frac{9 \cdot 4}{35} - \frac{144}{49} = \cdots = \frac{24}{49}$$

Zadatak 3. U kutiji se nalazi 5 dobrih i 7 loših žarulja. Neka je slučajna varijabla *x* broj dobrih žarulja u uzorku od 4 žarulje. Izračunaj a), b) i c) kao u zadatku 2.

$$\Rightarrow \quad p(x) = \frac{\binom{5}{4}\binom{7}{4-x}}{\binom{12}{4}}$$

3. BINOMNA RAZDIOBA

Binomna razdioba je najvažnija diskretna razdioba.

Neka se događaj A pojavljuje u nizu od *n* pokusa svaki put s konstantnom vjerojatnošću *p*. Takav događaj zove se Bernoullijev. Vjerojatnost da se događaj A neće desiti u jednom od tih pokusa je $q = 1 - p$.

Varijabla *x* neka je broj nastupa događaja A u tom nizu od *n* pokusa. To je slučajna varijabla i to diskretna jer svakoj njenoj vrijednosti $0,1,2,\ldots,n$ pripada odgovarajuća vjerojatnost. Iz teorije vjerojatnosti slijedi da je ta vjerojatnost

$$P(x) = \binom{n}{x} p^x \cdot q^{n-x}.$$

Ovom relacijom dana je vjerojatnost da u seriji od *n* pokusa događaj A nastupi *x* puta, a time je potpuno određena slučajna varijabla *x*.

Skup svih parova $(x, P(x))$; $x = 0,1,2,...,n$ čini **binomnu razdiobu** i kaže se da slučajna varijabla ima binomnu razdiobu ili da se distribuira po binomnoj razdiobi.

Zbog svojstava binomnih koeficijenata $\binom{n}{x}$, lako se pokaže da vrijedi $\sum_{x=0}^{n} P(x) = 1$, a što je nužno da *x* bude diskretna slučajna varijabla.

Vidimo da je binomna razdioba potpuno određena brojevima $n \in \mathbf{N}$ i $p \in \langle 0,1 \rangle$, pa se kraće označava sa $B\{n, p\}$.

Računanje vrijednosti $P(x)$ po gornjoj formuli nije jednostavno, posebno za veće vrijednosti od *x* i *n*, pa je $P(x)$ računamo po rekurzivnoj formuli

$$P(x) = \frac{n - x + 1}{x} \cdot \frac{p}{q} \cdot P(x-1).$$

Po toj formuli lako se izračuna $P(x)$ ako se zna $P(x-1)$, odnosno ako se zna $P(0)$ (vrijedi: $P(0) = q^n$), lako se izračuna $P(1)$, $P(2)$, itd.

Inače, postoje tablice u kojima su ispisane vjerojatnosti za najčešće binomne razdiobe (vidi str. **53**), pa te vjerojatnosti ni ne treba računati već samo iščitati iz tih tablica.

Može se pokazati da vrijede teoremi:

<u>Teorem 1</u>. Slučajna varijabla *x* binomne razdiobe ima **očekivanje** $E(x) = np$ i **varijancu** $V(x) = npq$.

Teorem 2. Vjerojatnosti $P(x)$ binomne razdiobe čine padajući niz ako je $p \le \frac{1}{n+1}$, a rastući ako je $p \ge \frac{n}{n+1}$. U ostalim slučajevima, tj. $\frac{1}{n+1} < p < \frac{n}{n+1}$, binomne vjerojatnosti $P(x)$ u početku rastu, a zatim padaju, a $P(x)$ ima maksimum za onaj $x \in \mathbf{N}$ koji zadovoljava $np - q \le x \le np + p$.

Dakle, pri binomnoj razdiobi $B\{n, p\}$ najveća vjerojatnost pripada onoj vrijednosti slučajne varijable x koja zadovoljava nejednakost $np - q \le x \le np + p$.

Primjer 1. Stroj koji radi ujednačeno tokom vremena 5% defektnih proizvoda. Kolika je vjerojatnost da u slučajnom uzorku od 5 proizvoda budu 2 defektna?

Vjerojatnost izrade defektnog proizvoda je $p = 0.05$, a dobrog proizvoda je $q = 0.95$.Broj loših proizvoda u uzorku je slučajna binomna razdioba $B\{5, 0.05\}$, pa je

$$P(2) = \binom{5}{2} 0.05^2 \cdot 0.95^3 = 0.021 \text{ (ovaj broj možemo iščitati i iz tablice).}$$

Što znači ovaj rezultat? On znači da ako bismo proizvode pakirali u kutije po 5 proizvoda (bez prethodnog odvajanja dobrih od loših proizvoda), onda bi 2.1% kutija sadržavalo 2 defektna proizvoda.

Analogno bi izračunali $P(0) = 0.774$, $P(1) = 0.203$, itd.

Primjer 2. Stroj daje 4% loših proizvoda. Proizvodi se stavljaju u kutije od po 50 komada. U koliko posto kutija će biti najviše jedan loš proizvod?

Ovdje je $p = 0.04$, $q = 0.96$, $n = 50$, pa je vjerojatnost da bude 0 loših proizvoda:

$$P(0) = \binom{50}{0} 0.04^0 \cdot 0.96^{50} = 0.96^{50},$$

a da bude jedan loš proizvod:

$$P(1) = \binom{50}{1} 0.04^1 \cdot 0.96^{49}.$$

Dakle,

$$P(x \le 1) = P(0) + P(1) = 0.96^{50} + 50 \cdot 0.04 \cdot 0.96^{49} = \ldots = 0.40 \;\Rightarrow\; 40\%.$$

Primjer 3. Od svakih 10 proizvoda njih 9 zadovoljava standard. Odredi vjerojatnost da od 50 uzetih proizvoda broj onih koji zadovoljavaju standard bude veći od 42, a manji od 48.

$$p = \frac{9}{10} = 0.9\,,\ q = 0.1\,,\ n = 50\ \Rightarrow\ P(x) = \binom{n}{x} p^x \cdot q^{n-x} = \binom{50}{x} 0.9^x \cdot 0.1^{n-x}$$

$$P(42 < x < 48) = P(43) + P(44) + P(45) + P(46) + P(47) =$$

$$= \binom{50}{43} 0.9^{43} \cdot 0.1^7 + \binom{50}{44} 0.9^{44} \cdot 0.1^6 + \binom{50}{45} 0.9^{45} \cdot 0.1^5 + \binom{50}{46} 0.9^{46} \cdot 0.1^4 + \binom{50}{47} 0.9^{47} \cdot 0.1^3 =$$

$$= \cdots = 0.7652$$

Primjer 4. U kutiji ima 200 plavih i 300 crvenih kuglica. Izvlači se nasumce 5 puta po jedna kuglica (s vraćanjem u kutiju) i neka je x broj plavih kuglica među njima. Ako se taj pokus ponovi 3125 puta, izračunaj pripadne frekvencije.

$$p = \frac{200}{500} = 0.4\,,\ q = 0.6\,,\ n = 50\,,\ \ x = 0,1,2,3,4,5$$

Ovo je zaista binomna razdioba $B\{n, p\} = B\{5, 0.4\}$, pa je zakon vjerojatnosti $P(x) = \binom{5}{x} 0.4^x \cdot 0.6^{5-x}$. Vjerojatnosti $P(x)$ možemo iščitati iz tablice, a pripadne frekvencije f_x (koliko smo puta u tih 3125 pokusa izvukli x plavih kuglica) bit će f_x=$3125 \cdot P(x)$:

x	$P(x)$	f_x
0	0.0778	243
1	0.2592	810
2	0.3456	1080
3	0.2314	720
4	0.0768	240
5	0.0102	32
	1.00	3125

Primjer 5. Vjerojatnost pojavljivanja događaja A u svakom ispitivanju iznosi 0.75. Kolika je vjerojatnost da se kod 8 ispitivanja događaj A pojavi više od 6 puta?

$$P(x > 6) = P(7) + P(8)\,;\ B\{8, 0.75\};\ \ P(x) = \binom{8}{x} 0.75^x \cdot 0.25^{8-x}$$

$$P(x > 6) = 8 \cdot 0.75^7 \cdot 0.25 + 0.75^8 = 0.75^7 (2 + 0.75) = 0.3671$$

4. PRILAGOĐAVANJE BINOMNE RAZDIOBE EMPIRIČKIM PODACIMA

Pretpostavimo da cjelobrojnim empiričkim podacima x pripadaju empiričke frekvencije f_x i neka je $\sum f_x = N$. Moguće je nekad naslutiti da su vrijednosti x distribuirane po binomnoj razdiobi. Tada će svakoj vrijednosti x pripadati vjerojatnost $P(x)$ po binomnoj razdiobi, a prema tome i neka teorijska frekvencija f_{tx}. Riješimo problem nalaženja f_{tx}.

Budući da je binomna razdioba određena sa n i p, odredimo te parametre.

Broj $\boldsymbol{n}$ je najveća moguća vrijednost od x i to teorijski najveća, pa se ne može podudarati s najvećom vrijednošću $x - a$ iz empiričkih podataka. Broj n se dakle mora odrediti posebnim rasuđivanjem.

Parametar $\boldsymbol{p}$ računa se iz formule za očekivanje $E(x) = np$ tako da se $E(x)$ zamijeni aritmetičkom sredinom $\bar{x}$ empiričkih podataka: $\bar{x} = np \Rightarrow p = \dfrac{\bar{x}}{n}$.

Kad znamo n i p, vjerojatnost $P(x)$ se izračuna po formuli za binomnu razdiobu (ili iz tablice) i teorijske frekvencije $f_{tx} = N \cdot P(x)$.

Ukoliko razlike između empiričkih i teorijskih frekvencija nisu velike, može se zaključiti da varijabla x zaista slijedi zakon binomne razdiobe.

<u>Zadatak 1.</u> Rad jednog stroja kontrolira se u uzorcima od 15 proizvoda tako da se u svakom uzorku broji broj loših proizvoda. Pošto je uzeto 200 uzoraka dobiveni su podaci:

x	0	1	2	3	4	5	6
f_x	77	81	31	7	2	1	1

Ovoj empiričkoj distribuciji prilagodi binomnu razdiobu, tj. pomoću binomne razdiobe $B\{n, p\}$ izračunaj teorijske frekvencije za pojedine vrijednosti varijable x.

Rješenje:

Budući da svih 15 proizvoda mogu biti defektni, slijedi da je $n = 15$. Drugim riječima, varijabla x može poprimiti vrijednosti: 0,1,2,…,15.

$$\Rightarrow \quad \bar{x} = \frac{1}{N}\sum_i f_i x_i = \text{vidi 3. stupac donje tablice} = \frac{183}{200} = 0.915$$

$$\Rightarrow \quad p = \frac{\bar{x}}{n} = \frac{0.915}{15} = 0.061 \approx 0.06$$

Iz tablice vjerojatnosti za binomnu razdiobu $B\{n=15, p=0.06\}$ čitamo vrijednosti $P(x)$ i ispisujemo ih u 4. stupac donje tablice, a teorijske frekvencije računamo po formuli

$$f_{tx} = N \cdot P(x) = 200 \cdot P(x)$$

i ispisujemo ih u 5. stupac:

x	f_x	$x \cdot f_x$	$P(x)$	f_{tx}
0	77	0	0.39529	79.058
1	81	81	0.37847	75.694
2	31	62	0.16911	33.822
3	7	21	0.04677	8.354
4	2	8	0.00896	1.792
5	1	5	0.00125	0.250
6	1	6	0.00014	0.028
	200	183		198.998

Iz 2. i 5. stupca vidimo da se empiričke i teorijske frekvencije prilično slažu, pa možemo reći da se ovi empirički podaci ravnaju po binomnoj razdiobi. One bi se još bolje slagale da smo promatrali još veći broj uzoraka.

Zadatak 2. Promatran je skup od 64 peteročlane obitelji. Ako je $x = 0,1,2,3,4,5$ broj zaposlenih u pojedinoj obitelji, dobivene su slijedeće frekvencije: f_x= 3,9,19,21,19,2. Ako ovaj statistički skup pripada binomnoj razdiobi, izračunaj teorijske frekvencije.

Zadatak 3. Baca se istovremeno 5 novčića 1000 puta. Ako je $x = 0,1,2,3,4,5$ broj novčića koji su pali na "grb", dobivene su slijedeće frekvencije: f_x= 39,144,342,287,164,25. Ako ovaj statistički skup pripada binomnoj razdiobi izračunaj teorijske frekvencije.

Vjerojatnosti pri binomnoj razdiobi $P(x) = \binom{n}{x} p^x \cdot q^{n-x}$

Svakoj vrijednosti u ovoj tablici prethodi decimalni zarez, dakle, 0,… .

5. POISSONOVA RAZDIOBA

Poissonova razdioba je granični slučaj binomne razdiobe kada n (broj pokusa) raste u beskonačnost ali tako da je umnožak $n \cdot p$ konstantan. Pri tome može p težiti k nuli: $n \to \infty$, $p \to 0$.

Izvršimo li taj granični prijelaz u relaciji $P(x) = \binom{n}{x} p^x \cdot q^{n-x}$, dobit ćemo $P(x) = \frac{m^x}{x!} e^{-m}$, gdje je $m = n \cdot p$, a $x = 0,1,2,3,\ldots$.

Stoga kažemo da je slučajna varijabla x distribuirana po zakonu Poissona, ako su njene vrijednosti $x = 0,1,2,3,\ldots$, a pripadne vjerojatnosti određene sa $P(x) = \frac{m^x}{x!} e^{-m}$. Skup svih parova $\{x, P(x)\}$ čini Poissonovu razdiobu. Ona se susreće kada se radi o događajima koji imaju malu vjerojatnost.

Vidimo da je Poissonova razdioba potpuno određena samo s parametrom *m*.

Teorem. Očekivanje i varijanca varijable x pri Poissonovoj razdiobi su: $E(x) = m$, $V(x) = m$.

Dokaz:

$$E(x) = \sum_{x=0}^{\infty} x \cdot P(x) = \sum_{x=0}^{\infty} x \cdot \frac{m^x}{x!} e^{-m} = \ldots = m$$

Ponekad je jednostavnije vjerojatnosti $P(x)$ računati po rekurzivnoj formuli:

$$P(x) = \frac{m}{x} P(x-1), \quad x = 1,2,3,\ldots,$$

dok za $x = 0$ vrijedi $P(0) = e^{-m}$.

Za neke vrijednosti parametra *m*, vjerojatnosti $P(x)$ mogu se iščitati iz tablice (vidi str.**57.**)

Primjer 1. Zadana je Poissonova razdioba s parametrom $m = 2$. Kolika je vjerojatnost da u toj razdiobi varijabla x primi vrijednost 3?

Iz formule za vjerojatnost Poissonove razdiobe za $m = 2$, $x = 3$ $\Rightarrow$ $P(3) = \frac{2^3}{3!} e^{-2} = \ldots = 0.1804$ (ovaj rezultat može se naći i u tablici).

Da je u ovom primjeru trebalo izračunati $P(0), P(1), P(2),...$, i da pri ruci nemamo tablicu vjerojatnosti Poissonove razdiobe, te vjerojatnosti računali bi po rekurzivnoj formuli $P(x) = \frac{m}{x} P(x-1)$:

$$P(0) = e^{-2} = 0.1353, P(1) = \frac{2}{1} \cdot P(0) = 2 \cdot 0.1353 = 0.2706, \; P(2) = \frac{2}{2} \cdot P(1) = 0.2706, \dots .$$

I Poissonovu razdiobu moguće je **prilagoditi empiričkim podacima** ako se procijeni da neki cjelobrojni skup empiričkih podataka slijedi Poissonov zakon (tj. ako je p malen broj, a n velik broj). Potrebno je samo izračunati aritmetičku sredinu $\bar{x}$ tog skupa podataka i staviti da je $\bar{x} = E(x) = m$ (jer je očekivanje $E(x) = m$, a očekivanje $E(x)$ odgovara $\bar{x}$). Nakon toga teorijske frekvencije f_{tx} pojedinih vrijednosti x računaju se po formuli $f_{tx} = N \cdot P(x)$, gdje je N ukupni broj empiričkih podataka, tj. $N = \sum_x f_x$, a $P(x)$ računa se po formuli

$$P(x) = \frac{m^x}{x!} e^{-m} \quad \text{ili} \quad P(x) = \frac{m}{x} P(x-1).$$

<u>Zadatak 1</u>. u Njemačkoj je tijekom 69 godina promatrano rođenje četvorki. Dobiveni su ovi rezultati (prva dva stupca tablice):

x	f_x	f_{tx}
0	14	14.2
1	24	22.5
2	17	17.7
3	9	9.3
4	2	3.7
5	2	1.2
6	1	0.3
	69	

Iz ovih empiričkih podataka slijedi da je $m = \bar{x} = 1.58$, pa su teorijske frekvencije $f_{tx} = 69 \cdot P(x) = 69 \cdot \frac{1.58^x}{x!} e^{-1.58}$. Pomoću te formule izračunali smo treći stupac tablice, pa vidimo dobro slaganje 2. i 3. stupca.

Zadatak 2. U 200 vremenskih razdoblja vođena je statistika o broju vojnika poginulih od pogrešnog rukovanja oružjem. Dobiveni su rezultati:

u 109 razdoblja	0 poginulih
65	1
22	2
3	3
1	4

Ovo je tipičan primjer rijetkog događaja. Ako se izračuna $m = \overline{x} = \dfrac{65+44+9+4}{200} = 0.61$,

$P(x) = \dfrac{0.61^x}{x!} e^{-0.61}$, $f_{tx} = 200 \cdot P(x)$, dobit ćemo 3. i 4. stupac slijedeće tablice.

x	f_x	$P(x)$	f_{tx}
0	109	0.5435	108.7
1	65	0.3311	66.2
2	22	0.1011	20.2
3	3	0.0207	4.1
4	1	0.0032	0.6

Opet se vidi dobro slaganje empiričkih i teorijskih rezultata.

Zadatak 3. Za vrijeme 2. svjetskog rata Nijemci su bombardirali London letećim bombama V1. Na područje Južnog Londona koje je podijeljeno na 576 sektora palo je 537 bombi i to po slijedećem rasporedu (prva 2 stupca):

x = broj bombi	f_x = broj sektora	f_{tx}
0	229	226.7
1	211	211.4
2	93	98.5
3	35	30.6
4	7	7.1
5 i više	1	1.6
	576	

Treći stupac slijedi iz: $m=\bar{x}=\dfrac{0\cdot 229+1\cdot 211+\cdots+5\cdot 1}{576}=0.9323$, $P(x)=\dfrac{0.9323^{x}}{x!}e^{-0.9323}$,

$f_{tx}=576\cdot P(x)$.

Opet se vidi dobro slaganje, pa se ovdje zaista radi o Poissonovoj razdiobi.

Vjerojatnosti pri Poissonovoj razdiobi $P(x)=\dfrac{m^{x}}{x!}e^{-m}$

Svakoj vrijednosti u ovoj tablici prethodi decimalni zarez!

6. KONTINUIRANE SLUČAJNE VARIJABLE

Kontinuirane slučajne varijable služe za izučavanje neprekinutih ili kontinuiranih obilježja nekog skupa. Njihovo područje vrijednosti je neki interval $\langle a, b \rangle$ realnih brojeva ili čitav skup **R** odnosno brojevni pravac $\langle -\infty, +\infty \rangle$. Kod kontinuiranih slučajnih varijabli vjerojatnosti ćemo pridruživati intervalima, a pojedinim vrijednostima varijable *x* pripada vjerojatnost 0.

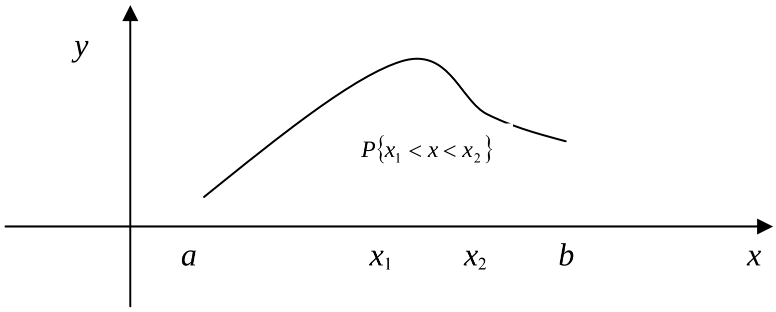
Sjetimo se, diskontinuirana ili diskretna slučajna varijabla poprimala je diskretne vrijednosti $x_1, x_2, x_3, \ldots$ s vjerojatnostima $p(x_1), p(x_2), p(x_3), \ldots$. Ulogu zakona vjerojatnosti ima $p(x_i)$ kod diskretne, a kod kontinuirane slučajne varijable ima funkcija $f(x)$, odnosno $P\{x_1 < x < x_2\}$ definirana na slijedeći način:

Funkcija vjerojatnosti ili **funkcija gustoće vjerojatnosti** kontinuirane slučajne varijable *x* je funkcija $f(x)$ koja ima svojstva:

1. $f(x) \geq 0$, $\forall x$;
2. $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$;
3. $\int\limits_{x_1}^{x_2} f(x)dx = P\{x_1 < x < x_2\}$ (tu je $x_1 < x_2$).

Zorna interpretacija ovih svojstava slijedi iz grafa funkcije vjerojatnosti $f(x)$. Relacija 1. izriče da je čitava krivulja vjerojatnosti iznad osi *x*, relacija 2. kaže da površina između osi *x* i krivulje $f(x)$ mora biti jednaka 1, a relacija 3. kaže da svakom intervalu $\langle x_1, x_2 \rangle$ pridružujemo vjerojatnost $P\{x_1 < x < x_2\}$ koja je jednaka površini ispod krivulje vjerojatnosti nad intervalom $\langle x_1, x_2 \rangle$.

Ako područje vrijednosti varijable *x* nije čitav skup **R** već interval $\langle a, b \rangle$, tada se uzima da je $f(x) = 0$ za sve vrijednosti *x* izvan tog intervala. Tada relacija 2. ima oblik $\int\limits_{a}^{b} f(x)dx = 1$.



Funkcija distribucije kontinuirane slučajne varijable *x* je funkcija $F(x) = \int_{-\infty}^{x} f(x)dx$.

Broj $F(x)$ je dakle vjerojatnost da varijabla *x* primi bilo koju vrijednost manju od *x*.

Vjerojatnost $P\{x_1 < x < x_2\}$ moguće je izraziti pomoću funkcije distribucije $F(x)$:

$$P\{x_1 < x < x_2\} = F(x_2) - F(x_1).$$

Za kontinuirane slučajne varijable *x* definira se također očekivanje, varijanca i momenti. Te su definicije analogne onima za diskretnu slučajnu varijablu:

$$E(x) = \mu = \int_{-\infty}^{+\infty} x \cdot f(x)dx,$$

$$V(x) = \sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot f(x)dx,$$

$$m_r = \int_{-\infty}^{+\infty} x^r \cdot f(x)dx,$$

$$M_r = \int_{-\infty}^{+\infty} (x-\mu)^r \cdot f(x)dx.$$

Vrijedi također: $E(ax+b) = aE(x) + b$, $V(ax+b) = a^2 V(x)$.

Primjer 1.

a) Odredi konstantu *c* tako da funkcija $f(x) = cx$ bude funkcija vjerojatnosti varijable *x* koja prima vrijednosti iz intervala $\langle 0,4 \rangle$.

Zbog uvjeta 1. mora biti *c*>0, a zbog uvjeta 2. mora biti

$$\int_0^4 cxdx = c\int_0^4 xdx = c \cdot \frac{x^2}{2}\Big|_0^4 = 8c = 1, \text{ tj. } c = \frac{1}{8}.$$

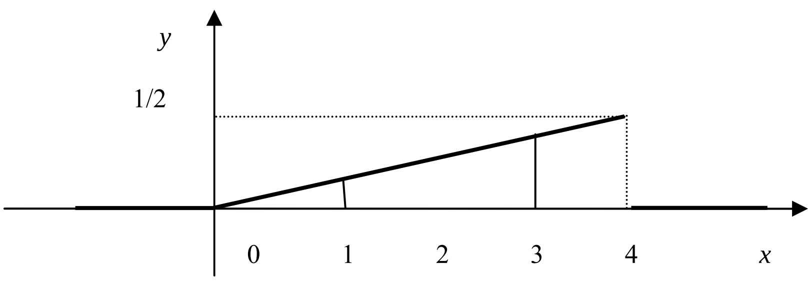
Prema tome

$$f(x)=\begin{cases}\frac{1}{8}x, x\in[0,4]\\ 0, x<0 \ \& \ x>4\end{cases}$$

b) Izračunaj vjerojatnost da slučajna varijabla *x* poprimi vrijednost iz intervala $\langle 1,3\rangle$.

$$P\{1<x<3\}=\int_1^3 f(x)dx=\frac{1}{8}\int_1^3 xdx=\ldots=\frac{1}{2}$$

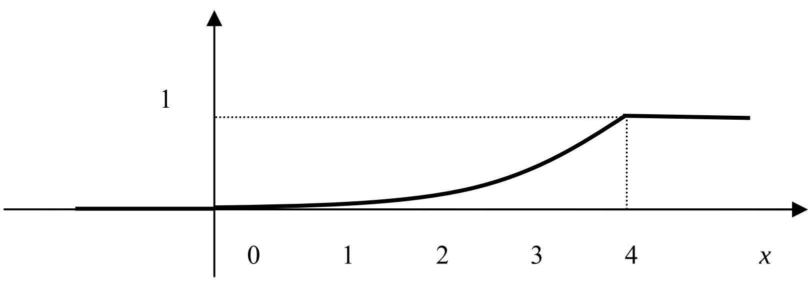
Površina koja odgovara toj vjerojatnosti je:



c) Funkcija distribucije je:

za $0<x<4$: $F(x)=\int_0^x \frac{1}{8}xdx=\frac{1}{16}x^2$.

Naravno vrijedi: $F(x)=0$ za $x\le 0$, $F(x)=1$ za $x\ge 0$



$$\Rightarrow \ F(x)=\begin{cases}0, x<0\\ \frac{1}{16}x^2, x\in[0,4]\\ 1, x>1\end{cases}$$

d) Izračunajmo očekivanje i varijancu:

$$E(x)=\mu=\int_{-\infty}^{+\infty} x\cdot f(x)dx=\int_0^4 x\cdot\frac{1}{8}xdx=\frac{1}{8}\cdot\frac{1}{3}x^3\Big|_0^4=\frac{8}{3}$$

$$V(x)=\sigma^2=\int_{-\infty}^{+\infty}(x-\mu)^2\cdot f(x)dx=\int_0^4(x-\frac{8}{3})^2\cdot\frac{1}{8}xdx=\ldots=\frac{8}{9}$$

Primjer 2. Slučajna varijabla može poprimiti vrijednost x s vjerojatnošću koja je proporcionalna s $\frac{1}{(1+x^2)^2}$.

a) Kako glasi funkcija vjerojatnosti ako x može poprimiti sve vrijednosti iz $\langle-\infty,+\infty\rangle$?

b) Kako glasi funkcija distribucije?

c) Odredi očekivanje i varijancu.

d) Odredi vjerojatnost da x poprimi vrijednost iz intervala $\langle 0,\sqrt{3}\rangle$.

a) $f(x)=c\cdot\frac{1}{(1+x^2)^2}$; $c>0$ zbog uvjeta 1.

$$\int_{-\infty}^{+\infty} c\cdot\frac{1}{(1+x^2)^2}dx=1$$

$$c\int_{-\infty}^{+\infty}\frac{1}{(1+x^2)^2}dx=\begin{vmatrix} x=tgt \\ dx=\frac{dt}{\cos^2 t}\end{vmatrix}=c\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\frac{\frac{1}{\cos^2 t}}{\frac{1}{\cos^4 t}}dt=c\cdot\frac{1}{2}(t+\sin t\cos t)\Big|_{-\frac{\rho}{2}}^{\frac{\pi}{2}}=c\cdot\frac{1}{2}(\frac{\pi}{2}+\frac{\pi}{2})=\frac{c\pi}{2}=1$$

$$\Rightarrow\ c=\frac{2}{\pi}\ \Rightarrow\ f(x)=\frac{2}{\pi}\cdot\frac{1}{(1+x^2)^2}.$$

b) $$F(x)=\int_{-\infty}^{x} f(x)dx=\frac{2}{\pi}\int_{-\infty}^{x}\frac{1}{(1+x^2)^2}dx=\frac{2}{\pi}\cdot\frac{1}{2}(t+\sin t\cos t)\Big|_{-\frac{\pi}{2}}^{arctgx}=\begin{vmatrix}\sin t=\frac{tgt}{\sqrt{1+tg^2 t}} \\ \cos t=\frac{1}{\sqrt{1+tg^2 t}}\end{vmatrix}=$$

$$=\frac{1}{\pi}(arctgx+\frac{x}{\sqrt{1+x^2}}\cdot\frac{1}{\sqrt{1+x^2}}+\frac{\pi}{2})=\frac{1}{\pi}(arctgx+\frac{x}{1+x^2}+\frac{\pi}{2}).$$

c)

$$E(x)=\int_{-\infty}^{+\infty} x\cdot f(x)dx=\int_{0}^{4} x\cdot\frac{2}{\pi}\cdot\frac{1}{(1+x^2)^2}dx=\begin{vmatrix}1+x^2=t\\2xdx=dt\end{vmatrix}=\frac{1}{\pi}\int\frac{dt}{t^2}=\frac{1}{\pi}\frac{t^{-1}}{-1}=-\frac{1}{\pi}\frac{1}{1+x^2}\Bigg|_{-\infty}^{+\infty}=0$$

$$V(x)=\sigma^2=\int_{-\infty}^{+\infty}(x-\mu)^2\cdot f(x)dx=\int_{0}^{4}(x-0)^2\cdot\frac{2}{\pi(1+x^2)^2}dx=\frac{2}{\pi}\int_{-\infty}^{+\infty}\frac{x^2}{(1+x^2)^2}=\begin{vmatrix}x=tgt\\dx=\dfrac{dt}{\cos^2 t}\end{vmatrix}=$$

$$=\frac{2}{\pi}\int_{0}^{\pi/3}\cos^2 t dt=\frac{2}{\pi}\frac{1}{2}(t+\sin t\cos t)\Bigg|_{0}^{\pi/3}=\frac{1}{\pi}\left(\frac{\pi}{3}+\frac{1}{2}\frac{\sqrt{3}}{2}\right)=\frac{1}{3}+\frac{\sqrt{3}}{4\pi}$$

ILI:

$$P\{x_1<x<x_2\}=\int_{x_1}^{x_2} f(x)dx$$

$$P\{0<x<\sqrt{3}\}=F(\sqrt{3})-F(0)=\frac{1}{\pi}\left(arctg\sqrt{3}+\frac{\sqrt{3}}{1+3}+\frac{\pi}{2}\right)-\frac{1}{\pi}\left(arctg0+0+\frac{\pi}{2}\right)=$$

$$=\frac{1}{\pi}\frac{\pi}{3}+\frac{1}{\pi}\frac{\sqrt{3}}{4}+\frac{1}{2}-\frac{1}{2}=\frac{1}{3}+\frac{\sqrt{3}}{4\pi}$$

Zadatak. Kontinuirana slučajna varijabla x ima funkciju vjerojatnosti:

$$f(x)=\begin{cases}a\cos x, -\dfrac{\pi}{2}<x<\dfrac{\pi}{2}\\ 0, x\le-\dfrac{\pi}{2}\ \&\ x\ge\dfrac{\pi}{2}\end{cases}$$

a) Odredi konstantu a.

b) Odredi funkciju distribucije.

c) Izračunaj vjerojatnost da slučajna varijabla x bude u $\left\langle 0,\frac{\pi}{4}\right\rangle$.

Rješenje:

a) $a>0$

$$\int_{-\pi/2}^{\pi/2} a\cos xdx=+a\sin x\Bigg|_{-\pi/2}^{\pi/2}=+a\left(\sin\frac{\pi}{2}+\sin\frac{\pi}{2}\right)=2a=1 \quad\Rightarrow\quad a=\frac{1}{2}$$

b)

$$-\infty<x\le-\frac{\pi}{2}\Rightarrow\ F(x)=\int_{-\infty}^{x} f(x)dx=0$$

$$-\frac{\pi}{2} < x < \frac{\pi}{2} \quad \Rightarrow \quad F(x) = \int_{-\infty}^{-\pi/2} f(x)dx + \int_{-\pi/2}^{x} f(x)dx = 0 + \int_{-\pi/2}^{x} \frac{1}{2}\cos x dx = \frac{1}{2}\sin x\Big|_{-\pi/2}^{x} = \frac{1}{2}(\sin x + 1)$$

$$\frac{\pi}{2} \le x < \infty \quad \Rightarrow \quad F(x) = \int_{-\infty}^{-\pi/2} f(x)dx + \int_{-\pi/2}^{\pi/2} f(x)dx + \int_{\pi/2}^{\infty} f(x)dx = 0 + \frac{1}{2}(\sin\frac{\pi}{2} + 1) + 0 = 1$$

Dakle,

$$F(x) = \begin{cases} 0, x \le -\pi/2 \\ \frac{1}{2}(\sin x + 1), -\pi/2 < x < \pi/2 \\ 1, x \ge \pi/2 \end{cases}$$

c)

$$P\{0 < x < \pi/4\} = F\left(\frac{\pi}{4}\right) - F(0) = \frac{1}{2}\left(\sin\frac{\pi}{4} + 1\right) - \frac{1}{2}(\sin 0 + 1) = \frac{1}{2}\frac{\sqrt{2}}{2} = \frac{\sqrt{2}}{4}$$

ILI

$$P\{0 < x < \pi/4\} = \int_{0}^{\pi/4} \frac{1}{2}\cos x dx = \frac{1}{2}\sin x\Big|_{0}^{\pi/4} = \frac{\sqrt{2}}{4}$$

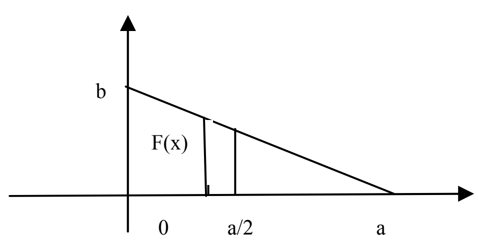
Zadatak. Slučajna varijabla (kontinuirana) x raspoređena je po "zakonu pravokutnog trokuta" u intervalu $\langle 0, a \rangle$ (vidi sliku). Izračunajte:

a) funkciju gustoće $f(x)$;

b) funkciju distribucije $F(x)$;

c) vjerojatnost da slučajna varijabla primi vrijednosti iz intervala $\left\langle \frac{a}{2}, a \right\rangle$;

d) standardnu devijaciju.



a) $\frac{x}{m}+\frac{y}{n}=1 \Rightarrow \frac{x}{a}+\frac{f(x)}{b}=1 \Rightarrow f(x)=b-\frac{b}{a}x$

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \Rightarrow \int_0^a \left(b-\frac{b}{a}x\right)dx = bx-\frac{b}{2a}x^2\Big|_0^a = ba-\frac{1}{2}ba=\frac{ba}{2}=1 \Rightarrow b=\frac{2}{a}$$

$$\Rightarrow \quad f(x)=\begin{cases}\frac{2}{a}\left(1-\frac{1}{a}x\right), x\in\langle 0,a\rangle \\ 0, x\notin\langle 0,a\rangle\end{cases}$$

Ovo se može riješiti i **elementarnim putem**. Kako uvjet $\int_{-\infty}^{+\infty} f(x)dx = 1$ geometrijski znači površinu ispod luka krivulje gustoće, a mi ovdje imamo pravokutni trokut, to je $\frac{ab}{2}=1 \Rightarrow b=\frac{2}{a}$

b) $F(x)=\int_{-\infty}^{x} f(x)dx = \int_0^x \frac{2}{a}\left(1-\frac{1}{a}x\right)dx = \frac{2}{a}\left(x-\frac{1}{2a}x^2\right)\Big|_0^x = \frac{x}{a}\left(2-\frac{x}{a}\right)$

$$\Rightarrow \quad F(x)=\begin{cases}0, x<0 \\ \frac{x}{a}\left(2-\frac{x}{a}\right), x\in(0,a) \\ 1, x>a\end{cases}$$

Ili: $F(x)=\int_{-\infty}^{x} f(x)dx$ geometrijski znači površinu ispod luka krivulje gustoće nad intervalom $(-\infty, x)$. U našem slučaju je to površina trapeza, pa imamo:

$$F(x)=\frac{f(0)+f(x)}{2}x = \frac{x}{2}\left[\frac{2}{a}+\frac{2}{a}\left(1-\frac{x}{a}\right)\right] = \frac{x}{a}\left(2-\frac{x}{a}\right)$$

c) $\int_{x_1}^{x_2} f(x)dx = P\{x_1<x<x_2\} \quad \Rightarrow \quad P\left\{\frac{a}{2}<x<a\right\} = \int_{a/2}^{a}\frac{2}{a}\left(1-\frac{1}{a}x\right) = \frac{2}{a}\left(x-\frac{1}{2a}x^2\right)\Big|_{a/2}^{a} = \ldots = \frac{1}{4}$

Ili: $P\{x_1<x<x_2\}=F(x_2)-F(x_1) \quad \Rightarrow$

$$P\left\{\frac{a}{2}<x<a\right\}=F(a)-F\left(\frac{a}{2}\right)=\frac{a}{a}\left(2-\frac{a}{a}\right)-\frac{a}{2a}\left(2-\frac{a}{2a}\right)=\frac{1}{4}$$

Ili: $\int_{x_1}^{x_2} f(x)dx = P\{x_1 < x < x_2\}$ geometrijski znači površinu ispod luka krivulje gustoće nad intervalom (x_1, x_2). U našem slučaju je to površina trokuta, pa imamo:

$$P\left\{\frac{a}{2} < x < a\right\} = \frac{\frac{a}{2}\cdot\frac{1}{a}}{2} = \frac{1}{4}$$

d) $V(x) = \sigma^2 = m_2 - m_1^{\,2}$

$$m_1 = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^a x\cdot\frac{2}{a}\left(1-\frac{1}{a}x\right)dx = \frac{a}{3}$$

$$m_2 = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_0^a x^2\cdot\frac{2}{a}\left(1-\frac{1}{a}x\right)dx = \frac{a^2}{6}$$

$$\Rightarrow\ V(x) = \sigma^2 = \frac{a^2}{6} - \frac{a^2}{9} = \frac{a^2}{18} \quad \Rightarrow\ \sigma = \frac{a}{3\sqrt{2}}$$

Ili: $V(x) = \sigma^2 = \int_{-\infty}^{+\infty}(x-\mu)^2 f(x)dx$, $\mu = E(x) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^a x\cdot\frac{2}{a}\left(1-\frac{1}{a}x\right)dx = \frac{a}{3}$

$$\Rightarrow\ V(x) = \sigma^2 = \int_0^a\left(x-\frac{a}{3}\right)^2\frac{2}{a}\left(1-\frac{1}{a}x\right)dx = \frac{a^2}{18} \quad \Rightarrow\ \sigma = \frac{a}{3\sqrt{2}}$$

7. NORMALNA RAZDIOBA

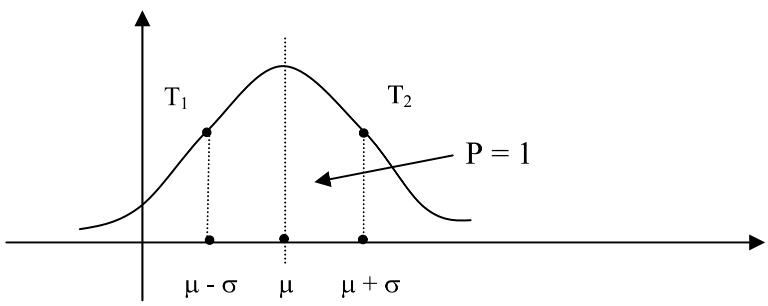
Normalna razdioba je najvažnija kontinuirana razdioba jer se najčešće susreće pri obradi nekog statističkog skupa, npr. pri promatranju visine ili debljine ili opsega lubanja nekog skupa ljudi, ili pri promatranju težine ili duljine nekog skupa istovrsnih poljoprivrednih proizvoda. Promatrajući na taj način svojstva nekog skupa, često se nailazi na veliku pravilnost i simetričnost.

Za kontinuiranu slučajnu varijablu x kaže se da ima **normalnu ili Gaussovu razdiobu** ako je njezina funkcija vjerojatnosti zadana formulom

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

Vidimo da je ta razdioba sasvim određena s očekivanjem μ i varijancom σ^2, pa se za normalnu razdiobu uvodi kraća oznaka $N\{\mu, \sigma^2\}$.

Iz gornje formule slijedi da je krivulja vjerojatnosti simetrična obzirom na pravac $x = \mu$, da je zvonolikog oblika, da ima maksimum $\frac{1}{\sigma\sqrt{2\pi}}$ u točki $x = \mu$ i da ima točke infleksije u $x = \mu - \sigma$ i $x = \mu + \sigma$:



Uvede li se supstitucija $u = \frac{x-\mu}{\sigma}$, dobiva se $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$. Ako se uvede nova funkcija $\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$, onda vrijedi $f(x) = \frac{1}{\sigma}\varphi(u)$. Vrijednosti funkcije $\varphi(u)$ date su u tablicama (vidi str. 69), pa se pomoću njih lako izračunaju vrijednosti funkcije $f(x) = \frac{1}{\sigma}\varphi(u)$.

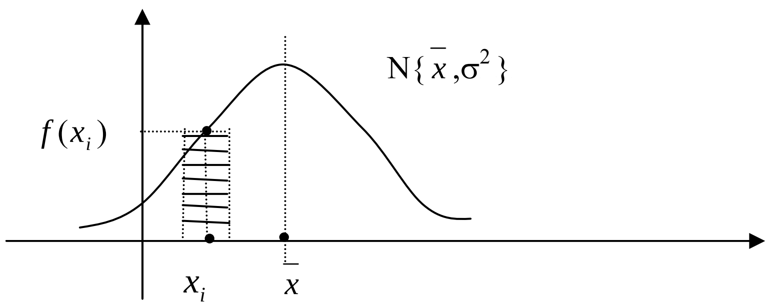
Lako se vidi da je $\varphi(u)$ funkcija vjerojatnosti jedinične normalne razdiobe N{μ=0,σ^2=1} i da vrijedi $\varphi(-u) = \varphi(u)$.

- **Prilagođavanje empiričkim podacima**

Pretpostavimo da je N empiričkih podataka grupirano u razrede širine *i* te da je $\overline{x}$ njihova aritmetička sredina, a σ^2 varijanca. Da bismo toj empiričkoj distribuciji prilagodili normalnu razdiobu, aproksimirat ćemo očekivanje μ aritmetičkom sredinom $\overline{x}$, a varijancu s varijancom empiričkih podataka.

Teorijske frekvencije za svaki razred računat ćemo formulom funkcije vjerojatnosti $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\overline{x}}{\sigma}\right)^2}$ ili jednostavnije (pomoću tablica) na sljedeći način.

Neka je x_i sredina i-tog razreda koji ima empiričku frekvenciju f_i te neka tim empiričkim podacima pripada normalna razdioba N{$\overline{x}$,σ^2}. Vjerojatnost P_i da varijabla x primi vrijednost iz i-tog razreda jednaka je površini ispod krivulje vjerojatnosti nad tim razredom:



Zato vrijedi: $P_i = i \cdot f(x_i)$, gdje je $f(x_i) = \frac{1}{\sigma}\varphi(u_i)$, a $u_i = \frac{x_i - \overline{x}}{\sigma}$. Prema tome, na bazi normalne razdiobe očekivana teorijska frekvencija i-tog razreda je:

$$f_{ti} = N \cdot P_i = N \cdot i \cdot f(x_i) = N \cdot i \cdot \frac{1}{\sigma}\varphi(u_i) = \frac{N \cdot i}{\sigma}\varphi(u_i).$$

Ukratko, postupak računanja teorijske frekvencije je:

$$x_i \to u_i = \frac{x_i - \overline{x}}{\sigma} \xrightarrow{tablica} \varphi(u_i) \to f_{ti} = \frac{N \cdot i}{\sigma}\varphi(u_i).$$

Primjer. Izmjeren je kapacitet od 485 istovrsnih kondenzatora. Dobiveni podaci dani su u prva 3 stupca donje tablice:

Razredi	Sredina x_i	Frekvencija f_i	$x_i - \bar{x}$	$\lvert u_i \rvert$	$\varphi(u_i)$	f_{ti}
19,58-19,62	19,60	3	-0,333	3,05	0,00381	0,8
19,63-19,67	19,65	5	-0,283	2,60	0,01358	3,0
19,68-19,72	19,70	5	-0,233	2,14	0,04041	9,0
19,73-19,77	19,75	20	-0,183	1,68	0,09728	21,6
19,78-19,82	19,80	35	-0,133	1,22	0,18954	42,2
19,83-19,87	19,85	74	-0,083	0,76	0,29887	66,5
19,88-19,92	19,90	92	-0,033	0,30	0,38139	84,7
19,93-19,97	19,95	83	0,017	0,16	0,39387	87,6
19,98-20,02	20,00	70	0,067	0,61	0,33121	73,6
20,03-20,07	20,05	54	0,117	1,07	0,22506	50,0
20,08-20,12	20,10	27	0,167	1,53	0,12376	27,5
20,13-20,17	20,15	12	0,217	1,99	0,05508	12,2
20,18-20,22	20,20	2	0,267	2,45	0,01984	4,4
20,23-20,27	20,25	3	0,317	2,91	0,00578	1,3
		485				484,4

Prilagodimo normalnu razdiobu ovim empiričkim podacima. Ako se izračuna aritmetička sredina i standardna devijacija ovih empiričkih podataka, dobit će se: $\bar{x} = 19{,}933$; σ = 0,109.

4. stupac tablice sadrži razlike $x_i - \bar{x}$ koje su potrebne za računanje vrijednosti u_i po formuli $u_i = \dfrac{x_i - \bar{x}}{\sigma}$ (**5. stupac**). **6. stupac** $\varphi(u_i)$ očitan je iz tablice – pri tom vrijedi $\varphi(-u) = \varphi(u)$,

npr. $\varphi(3{,}05) = 0{,}00381$ nalazi se u retku u_i= 3,0 i stupcu 5. Posljednji stupac sadrži teorijske frekvencije f_{ti} izračunate po formuli $f_{ti} = \dfrac{N \cdot i}{\sigma}\varphi(u_i) = \dfrac{485 \cdot 0{,}05}{0{,}109}\varphi(u_i) = 222{,}5 \cdot \varphi(u_i)$.

Vidimo da se kapaciteti kondenzatora zaista pokoravaju zakonu normalne razdiobe.

Ordinate jedinične normalne razdiobe

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$$

u	0	1	2	3	4	5	6	7	8	9
0,0	39894	39892	39886	39876	39862	39844	39822	39797	39767	39733
0,1	39695	39654	39608	39559	39505	39448	39387	39322	39253	39181
0,2	39104	39024	38940	38853	38762	38667	38568	38466	38361	38251
0,3	38139	38023	37903	37780	37654	37524	37391	37255	37115	36973
0,4	36827	36678	36526	36371	36213	36053	35889	35723	35553	35381
0,5	35207	35029	34849	34667	34482	34294	34105	33912	33718	33521
0,6	33322	33121	32918	32713	32506	32297	32086	31874	31659	31443
0,7	31225	31006	30785	30563	30339	30114	29887	29659	29431	29200
0,8	28969	28737	28504	28269	28034	27798	27562	27324	27086	26848
0,9	26609	26369	26129	25888	25647	25406	25164	24923	24681	24439
1,0	24197	23955	23713	23471	23230	22988	22747	22506	22265	22025
1,1	21785	21546	21307	21069	20831	20594	20357	20121	19886	19652
1,2	19419	19186	18954	18724	18494	18265	18037	17810	17585	17360
1,3	17137	16915	16694	16474	16256	16038	15822	15608	15395	15183
1,4	14973	14764	14556	14350	14146	13943	13742	13542	13344	13147
1,5	12952	12758	12566	12376	12188	12001	11816	11632	11450	11270
1,6	11092	10915	10741	10567	10396	10226	10059	09893	09728	09566
1,7	09405	09246	09089	08933	08780	08628	08478	08329	08183	08038
1,8	07895	07754	07614	07477	07341	07206	07074	06943	06814	06687
1,9	065620	06439	06316	06195	06077	05959	05844	05730	05618	05508
2,0	05399	05292	05186	05082	04980	04879	04780	04682	04586	04491
2,1	04398	04307	04217	04128	04041	03955	03871	03788	03706	03626
2,2	03547	03470	03394	03319	03246	03174	03103	03034	02965	02898
2,3	02833	02768	02705	02643	02582	02522	02463	02406	02349	02294
2,4	02239	02186	02134	02083	02033	01984	01936	01889	01842	01797
2,5	01753	01709	01667	01625	01585	01545	01506	01468	01431	01394
2,6	01358	01323	01289	01256	01223	01191	01160	01130	01100	01071
2,7	01042	01014	00987	00961	00935	00909	00885	00861	00837	00814
2,8	00792	00770	00748	00727	00707	00687	00668	00649	00631	00613
2,9	00595	00578	00562	00545	00530	00514	00499	00485	00471	00457
3,0	00443	00430	00417	00405	00393	00381	00370	00358	00348	00337
3,1	00327	00317	00307	00298	00288	00279	00271	00262	00254	00246
3,2	00238	00231	00224	00216	00210	00203	00196	00190	00184	00178
3,3	00172	00167	00161	00156	00151	00146	00141	00136	00132	00127
3,4	00123	00119	00115	00111	00107	00104	00100	00097	00094	00090
3,5	00087	00084	00081	00079	00076	00073	00071	00068	00066	00063
3,6	00061	00059	00057	00055	00053	00051	00049	00047	00046	00044
3,7	00042	00041	00039	00038	00037	00035	00034	00033	00031	00030
3,8	00029	00028	00027	00026	00025	00024	00023	00022	00021	00021
3,9	00020	00019	00018	00018	00017	00016	00016	00015	00014	00014

Svakoj vrijednosti u ovoj tablici prethodi decimalni zarez; tako je npr. φ (1,33) = 0,16474

- **Vjerojatnosti normalne razdiobe**

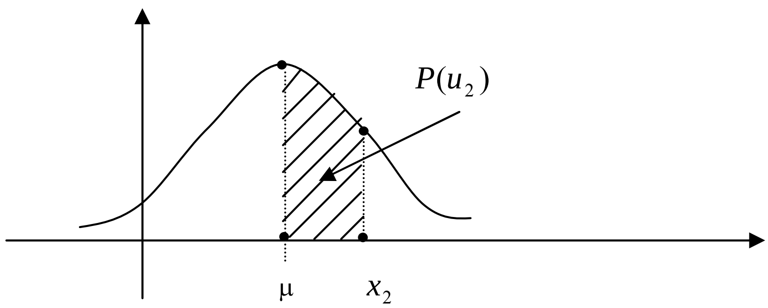
Vjerojatnost da varijabla x normalne razdiobe N{$\overline{x}$,σ²} poprimi vrijednost iz intervala $\langle x_1, x_2 \rangle$ je:

$$P\{x_1 < x < x_2\} = \int_{x_1}^{x_2} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Izvedimo jednostavniju formulu za računanje te vjerojatnosti. Provedemo li u integralu na desnoj strani supstituciju $u = \dfrac{x-\mu}{\sigma}$:

$$P\{x_1 < x < x_2\} = \frac{1}{\sqrt{2\pi}} \int_{u_1}^{u_2} e^{-\frac{1}{2}u^2} du$$

gdje je $u_1 = \dfrac{x_1 - \mu}{\sigma}$, $u_2 = \dfrac{x_2 - \mu}{\sigma}$. Ovaj integral predstavlja površinu ispod krivulje vjerojatnosti jedinične normalne razdiobe N{0,1} nad intervalom $\langle u_1, u_2 \rangle$. Ako uzmemo $x_1 = \mu$ slijedi da je $u_1 = 0$ i $P\{\mu < x < x_2\} = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^{u_2} e^{-\frac{1}{2}u^2} du = P(u_2)$.



Za različite vrijednosti u_2 vjerojatnosti $P(u_2)$ dane su u tablici na str. 72. Pomoću njih moguće je odgovarajućim slaganjem izračunati bilo koju vjerojatnost za normalnu razdiobu. U tablicama imamo samo pozitivne vrijednosti od u jer vrijedi: $\varphi(-u) = \varphi(u)$, $P(-u) = P(u)$.

Zato ćemo uzimati da je $u_i = \dfrac{|x_i - \mu|}{\sigma}$. Može se izračunati: $P\{\mu - 2\sigma < x < \mu + 2\sigma\} = 0{,}9545$, tj. pri normalnoj razdiobi 95,45% svih vrijednosti varijable x zadovoljava nejednakost: $\mu - 2\sigma < x < \mu + 2\sigma$ (odnosno pada u interval širine 4σ simetričan s obzirom na x os).

Primjer. Zadana je normalna razdioba N{ $\mu = 20$,$\sigma^2 = 4$}.

a) Izračunajmo $P\{17 < x < 21{,}5\}$

$$P\{17 < x < 21{,}5\} = P(u_1) + P(u_2) = 0{,}43319 + 0{,}27337 = 0{,}70656\,,$$

(vidi tablicu)

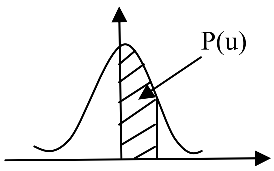
$$u_1 = \frac{|x_1 - \mu|}{\sigma} = \frac{|17 - 20|}{2} = |-1{,}5| = 1{,}5\ ;\quad P(u_1) = P(1{,}5) = 0{,}43319$$

$$u_2 = \frac{|x_2 - \mu|}{\sigma} = \frac{|21{,}5 - 20|}{2} = 0{,}75 \qquad ;\quad P(u_2) = P(0{,}75) = 0{,}27337$$

b) $P(x < 21{,}5) = \frac{1}{2} + P(u_2) = 0{,}5 + 0{,}27337 = 0{,}77337$

c) $P(x > 21{,}5) = \frac{1}{2} - P(u_2) = 0{,}5 - 0{,}27337 = 0{,}22663$

Površina ispod normalne krivulje



$$P(u)=\frac{1}{\sqrt{2\pi}}\int_0^u e^{-\frac{1}{2}x^2}\,dx$$

u	0	1	2	3	4	5	6	7	8	9
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	09383	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	39795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976
3,5	4997674									
4,0	4999683									
4,5	4999966									
5,0	499999713									

Svakoj vrijednosti u ovoj tablici prethodi decimalni zarez; tako je npr. P(1,71) = 0,45637

Zadatak. Mjerne vrijednosti karakteristične dimenzije proizvoda izrađenih na jednom stroju pokoravaju se zakonu normalne razdiobe $N\{\mu = 20mm; \sigma^2 = 0{,}04mm^2\}$. Za tu dimenziju propisano je polje tolerancije (19,45;20,40). Koliko će posto proizvoda imati dimenziju izvan tog polja?

Rješenje:

$$P = 1 - [P(u_1) - P(u_2)]$$

$$u_1 = \frac{19{,}45 - 20}{0{,}2} = -2{,}75 \;\; ; \;\; P(u_1) = 0{,}49702$$

$$u_2 = \frac{20{,}40 - 20}{0{,}2} = 2 \;\; ; \;\; P(u_2) = 0{,}47725$$

$P = 1 - [P(u_1) - P(u_2)] = 1 - 0{,}97427 = 0{,}02573$, ili u postocima 2,57%.

Zadatak. x je varijabla normalne razdiobe $N\{\mu = 30; \sigma^2 = 4\}$. Izračunaj:

a) $P(x \le 27)$

b) $P(x > 27)$

c) $P(26 < x < 29)$

d) $P(25 < x < 31)$

e) Odredi x_0 tako da bude $P(x < x_0) = 0{,}35$

Rješenje:

1. $P(x \le 27) = 0{,}5 - P(1{,}5) = 0{,}5 - 0{,}43319 = 0{,}06681$

$$u_1 = \frac{|27 - 30|}{2} = -1{,}5$$

2. $P(x > 27) = 0{,}5 + P(1{,}5) = 0{,}5 + 0{,}43319 = 0{,}93319$

3. $P(26 < x < 29) = P(u_1) - P(u_2) = P(2) - P(0{,}5) = ... = 0{,}287579$

$$u_1 = \frac{|26 - 30|}{2} = |-2| \;\; ; \;\; u_2 = \frac{|29 - 30|}{2} = |-0{,}5|$$

4. $P(25 < x < 31) = P(u_1) + P(u_2) = P(2{,}5) + P(0{,}5) = ... = 0{,}68525$

$$u_1 = \frac{|25 - 30|}{2} = 2{,}5 \;\; ; \;\; u_2 = \frac{|31 - 30|}{2} = 0{,}5$$

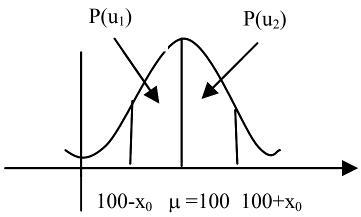
5. $P(x < x_0) + P(u_0) = 0{,}5$; $u_0 = 0{,}385$ (iz tablica)

$$u_0 = \frac{|x_0 - \mu|}{\sigma} \Rightarrow 0{,}385 = \frac{|x_0 - 30|}{2} / \cdot 2 \Rightarrow |x_0 - 30| = 0{,}77$$

$\Rightarrow 30 - x_0 = 0{,}77$ (jer je $30 > x_0$) $\Rightarrow x_0 = 29{,}23$.

Zadatak. x je varijabla normalne razdiobe $N\{\mu = 100; \sigma^2 = 9\}$. Odredi interval simetričan s obzirom na μ koji sadrži 95% svih vrijednosti varijable x.

Rješenje:



$$P(\mu - x_0 < x < \mu + x_0) = 0{,}95$$

$$P(u_1) + P(u_1) = 0{,}95 \Rightarrow P(u_1) = 0{,}475 \Rightarrow u_1 = 1{,}96$$

$$u_i = \frac{x_i - \bar{x}}{\sigma} \Rightarrow 1{,}96 = \frac{100 + x_0 - 100}{3} \Rightarrow x_0 = 3 \cdot 1{,}96 = 5{,}88$$

Rj. (94,22;105,88)

Zadatak. Izračunaj standardnu devijaciju normalne razdiobe ako je $E(x) = 2$ i $P\{x < 3\} = 0{,}9$

Rješenje:

$$P\{x < 3\} = 0{,}9$$

$$0{,}5 + P(u_0) = 0{,}9 \Rightarrow P(u_0) = 0{,}4 \Rightarrow u_0 = 1{,}28$$

$$u_i = \frac{|x_i - \bar{x}|}{\sigma}, \quad u_0 = \frac{x_0 - \bar{x}}{\sigma} \Rightarrow 1{,}28 = \frac{3-2}{\sigma} \Rightarrow \sigma = \frac{1}{1{,}28} = 0{,}781$$

Zadatak. Izračunaj onu vrijednost x_0 normalne razdiobe ako je $E(x)=10$, $V(x)=1$ i $P(x<x_0)=0{,}30$.

Rješenje:

$\mu = 10, \sigma = 1$

$$P(x<x_0)+P(u_0)=0{,}5 \Rightarrow 0{,}3+P(u_0)=0{,}5 \Rightarrow P(u_0)=0{,}2 \Rightarrow u_0=0{,}525$$

$$u_0=\frac{x_0-\bar{x}}{\sigma} \Rightarrow \frac{|x_0-10|}{1}=0{,}525 \Rightarrow 10-x_0=0{,}525 \Rightarrow x_0=9{,}475$$

- **Aproksimacija binomne razdiobe normalnom**

Može se pokazati da binomna razdioba teži k normalnoj razdiobi kad $n\to\infty$.

Vjerojatnosti pri binomnoj razdiobi $B\{n,p\}$ moguće je aproksimirati odgovarajućim vjerojatnostima pri normalnoj razdiobi $N\{\mu=np;\sigma^2=npq\}$ ako su zadovoljeni uvjeti:

1. $\frac{1}{n+1}<p<\frac{n}{n+1}$ i
2. $npq>9$.

Ako su ovi uvjeti zadovoljeni približno vrijedi:

$$P(r)=\binom{n}{r}p^r q^{n-r}=P_N\{r-0{,}5<x<r+0{,}5\}$$

Pri tom je $P(r)$ vjerojatnost pri binomnoj razdiobi $B\{n,p\}$ da varijabla x primi vrijednost r, a $P_N\{r-0{,}5<x<r+0{,}5\}$ vjerojatnost po normalnoj razdiobi $N\{\mu=np;\sigma^2=npq\}$ da varijabla x primi vrijednost iz intervala $(r-0{,}5;r+0{,}5)$.

Ako su r_1 i r_2 prirodni brojevi manji od n i $r_1<r_2$ tada je:

$$P_B\{r_1\le x\le r_2\}=\sum_{x=r_1}^{r_2}P(x)\approx P_N\{r-0{,}5<x<r+0{,}5\}.$$

Ako su uvjeti 1. i 2. ispunjeni onda i sva ostala svojstva normalne razdiobe aproksimativno vrijede za binomnu razdiobu, npr. $P\{\mu-2\sigma<x<\mu+2\sigma\}=0{,}9545$.

Primjer. U jednom eksperimentu događaj A nastupa uz vjerojatnost $p=0{,}5$. U kojim granicama će oscilirati broj nastupa događaja A u serijama od 40 eksperimenata?

Broj x nastupa događaja A u seriji od 40 eksperimenata je slučajna varijabla od $B\{n = 40, p = 0{,}5\}$.

1. $\dfrac{1}{n+1} < p < \dfrac{n}{n+1} \;\Rightarrow\; \dfrac{1}{41} < 0{,}5 < \dfrac{40}{41}$

2. $npq > 9 \;\Rightarrow\; 40 \cdot \dfrac{1}{2} \cdot \dfrac{1}{2} = 10 > 9$

Vidimo da su zadovoljena oba uvjeta za aproksimaciju binomne razdiobe sa normalnom razdiobom $N\{\mu = np = 20; \sigma^2 = npq = 10\}$. Zbog toga uz vjerojatnost od 0,9545 smijemo tvrditi da će broj nastupa x događaja A zadovoljavati nejednakost:

$$\mu - 2\sigma < x < \mu + 2\sigma$$

$$20 - 2\sqrt{10} < x < 20 + 2\sqrt{10}$$

$$13{,}68 < x < 26{,}32$$

$$14 \le x \le 26$$

Izračunamo li zbroj vjerojatnosti po binomnoj razdiobi $\sum_{r=14}^{26} P(r)$ dobit ćemo vrijednost: 0,96152.

Primjer. Automat daje u prosjeku 12% defektnih proizvoda. U kojim će se granicama kretati broj defektnih proizvoda u uzorcima od 200 proizvoda?

Broj defektnih proizvoda u uzorku je slučajna varijabla binomne razdiobe $B\{n = 200, p = 0{,}12\}$. Može se i ovdje vidjeti da su zadovoljeni uvjeti za aproksimaciju sa normalnom razdiobom $N\{\mu = np = 24; \sigma^2 = npq = 21{,}12\}$. Zbog toga će u oko 95% slučajeva broj x defektnih proizvoda u uzorku zadovoljavati nejednakost:

$$\mu - 2\sigma < x < \mu + 2\sigma$$

$$24 - 2\sqrt{21{,}12} < x < 24 + 2\sqrt{21{,}12}$$

$$14{,}8 < x < 33{,}2$$

$$15 \le x \le 33$$

Zadatak. Uz vjerojatnost 0,95 odredi granice broja x, ako je x broj pojavljivanja grba u 500 bacanja novčića.

Rješenje:

$$B\left\{n = 500,\ p = \frac{1}{2}\right\}$$

Provjerimo li uvjete 1. i 2. vidimo da se može izvršiti aproksimacija sa normalnom razdiobom $N\{\mu = np = 250; \sigma^2 = npq = 125\}$, pa će se uz vjerojatnost 0,95 x nalaziti u granicama:

$\mu - 2\sigma < x < \mu + 2\sigma \;\Rightarrow\; 250 - 2\cdot 11{,}18 < x < 250 + 2\cdot 11{,}18 \;\Rightarrow\; 228 \le x \le 272$

8. GAMA-RAZDIOBA

- **Gama funkcija**

Za pozitivan cijeli broj *n* definira se relacijom: $\Gamma(n) = \int_0^\infty x^{n-1} e^{-x}\,dx$.

Može se pokazati da taj nepravi integral konvergira za svaki pozitivni *n*. Parcijalnom integracijom pokazuje se da vrijedi rekurzivna formula:

$$\Gamma(n) = (n-1)\cdot\Gamma(n-1),\quad n > 1.$$

Vrijedi:

$\Gamma(1) = 1$, odnosno $\Gamma(n) = (n-1)!$

$\Gamma(n) = \frac{1}{n}\Gamma(n+1)$ - pomoću ove relacije može se područje definicije Γ-funkcije proširiti i na negativne *n*.

$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ (izračunava se pomoću beta funkcije (vidi niže)) $\Rightarrow$ $\Gamma\left(\frac{1}{2} - 1\right) = \dfrac{\Gamma(1)}{\frac{1}{2} - 1}$;

$\Gamma\left(\frac{1}{2} - 2\right) = \dfrac{\Gamma\left(\frac{1}{2} - 1\right)}{\frac{1}{2} - 2}$; …; $\Gamma\left(\frac{9}{2}\right) = \frac{7}{2}\cdot\Gamma\left(\frac{7}{2}\right) = \ldots = \frac{105}{16}\sqrt{\pi}$

- **Beta funkcija:** $\beta(m,n)=\int_0^1 x^{m-1}(1-x)^{n-1}dx$; $m>0, n>0$

Supstitucijom $x=\cos^2\varphi$ gornji izraz se prevodi u trigonometrijski oblik:

$$\beta(m,n)=2\int_0^{\pi/2}\cos^{2m-1}\varphi\cdot\sin^{2n-1}\varphi d\varphi$$

$$\Rightarrow \beta\left(\frac{1}{2},\frac{1}{2}\right)=2\int_0^{\pi/2}d\varphi=\pi$$

$$\Rightarrow \beta(m,n)=\frac{\Gamma(m)\cdot\Gamma(n)}{\Gamma(m+n)}$$

- **Gama razdioba**

Za varijablu x kažemo da je distribuirana po zakonu gama razdiobe ako je područje njenih vrijednosti interval $\langle 0,+\infty\rangle$, a funkcija vjerojatnosti:

$$f(x)=C_k x^{\frac{k-2}{2}}e^{-\frac{x}{2}},$$

gdje je $C_k=\dfrac{1}{2^{k/2}\cdot\Gamma\left(\frac{k}{2}\right)}$ i pri tom je k prirodni broj.

Parametar k kojim je gama funkcija jednoznačno određena zovemo **stupnjem slobode** razdiobe.

Očekivanje i varijanca varijable x koja je distribuirana po gama razdiobi su:

$$E(x)=k,\quad V(x)=2k.$$

Osim toga, α_3 je uvijek veći od nule i teži k nuli kada $k\to\infty$.

9. χ^2 – TEST

Kad smo prilagođavali pojedine razdiobe empiričkim podacima vidjeli smo da se empiričke i teorijske frekvencije u potpunosti ne podudaraju. Razlike $|f_i - f_{ti}|$ bile su u nekim razredima veće, a u nekima manje.

Postavlja se pitanje da li su te razlike prevelike da bi se moglo smatrati da se varijabla x pokorava dotičnoj prilagođenoj razdiobi. Ili, da li je razumno prihvatiti hipotezu H da se varijabla x na koju se odnose podaci pokorava prilagođenom teorijskom zakonu.

Važno je napomenuti da se hipoteza H odnosi samo na oblik razdiobe varijable x,a ne i na njene parametre. Odgovor daje **teorem**:

Veličina:

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_i - f_{ti})^2}{f_{ti}}$$

distribuirana je približno po zakonu gama razdiobe sa stupnjem slobode *k* koji ovisi o broju razreda *n* i o razdiobi koju smo prilagodili. Vrijedi za prilagođavanje:

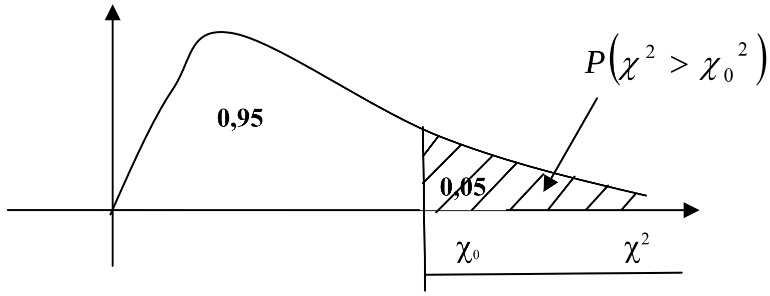
1. normalne razdiobe: $k = n - 3$;
2. binomne razdiobe: $k = n - 2$;
3. Poissonove razdiobe: $k = n - 2$.

Broj stupnjeva slobode gama razdiobe općenito je jednak $n - r - 1$, gdje je *r* broj parametara koji određuju prilagođenu razdiobu, a procjenjuju se iz empiričkih podataka. Za normalnu razdiobu je $r = 2$ ($\mu = \bar{x}$ i σ^2; tj. razredi nisu nezavisni, imaju određen $\bar{x}$ i σ^2).

Ako su razlike $|f_i - f_{ti}|$ velike, bit će velik i χ^2. Ako je, dakle, χ^2 velik razumno je odbaciti hipotezu H kao neispravnu. Po dogovoru uzeti ćemo da je χ^2 prevelik (to se kaže signifikantan) ako padne izvan područja kojem pripada vjerojatnost 0,95.

Postupak testiranja hipoteze:

Izračunamo ${\chi_0}^2$ po formuli za χ^2, a nakon toga za odgovarajući stupanj slobode *k* očitamo iz tablice VII vjerojatnost: $P(\chi^2 > {\chi_0}^2)$. Ako je ta vjerojatnost veća od 0,05 prihvaćamo hipotezu H, a ako je manja od 0,05 hipotezu ćemo odbaciti. U tom se slučaju ${\chi_0}^2$ nalazi u području signifikantnosti. (To odbacivanje hipoteze u 5% slučajeva bit će pogrešno)



Područje signifikantnosti

χ^2 - test se primjenjuje jedino kad su teorijske frekvencije f_{ti} veće od 5. Ako je $f_{ti} < 5$ za početne ili posljednje razrede izvrši se pregrupacija podataka tako da se spoji nekoliko prvih odnosno posljednjih razreda u jedan (frekvencije tih razreda jednostavno zbrojimo). U tom je slučaju *n* broj pregrupiranih razreda.

Zadatak. Mjerenjem kapaciteta 485 kondenzatora dobiveni su podaci kao u prva dva stupca tablice. Tim empiričkim podacima prilagodili smo normalnu razdiobu i dobili teorijske frekvencije f_{ti} (3. stupac):

x_i	f_i	f_{ti}	$\lvert f_i - f_{ti} \rvert$	$(f_i - f_{ti})^2 / f_{ti}$
19,60 19,65 19,70	3 5 } 13 5	0,8 3,0 } 12,8 9,0	0,2	0,003
19,75	20	21,6	1,6	0,118
19,80	35	42,2	7,2	1,227
19,85	74	66,5	7,5	0,846
19,90	92	84,7	7,3	0,629
19,95	83	87,6	4,6	0,241
20,00	70	73,6	3,6	0,176
20,05	54	50,0	4,0	0,320
20,10	27	27,5	0,5	0,009
20,15	12	12,2	0,2	0,003
20,20 20,25	2 3 } 5	4,4 1,3 } 5,7	0,7	0,086
				3,658

Postavljamo hipotezu H: kapaciteti serije kondenzatora iz koje je uzet naš uzorak pokoravaju se zakonu normalne razdiobe.

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_i - f_{ti})^2}{f_{ti}} = 3{,}658$$

Ovoj vrijednosti za χ^2 odgovara stupanj slobode $k = n - 3$ jer je riječ o prilagođavanju normalne razdiobe: $n = 11$ - broj pregrupiranih razreda, $k = 11 - 3 = 8$.

Za 8 stupnjeva slobode nalazimo u tablici VII: $P(\chi^2 > 3{,}490) = 0{,}90$ i $P(\chi^2 > 4{,}594) = 0{,}80$.

Budući da je naš $\chi^2 =$ 3,658 između 3,490 i 4,594 to je: $0{,}80 < P(\chi^2 > 3{,}658) < 0{,}90$.

Vjerojatnost je dakle veća od 0,05 pa prihvaćamo hipotezu H. Štoviše, jer je vjerojatnost dosta velika, mogli bismo očekivati da uz ispravnost naše hipoteze dobijemo još i veće razlike među empiričkim i teorijskim frekvencijama. Drugim riječima, naši empirički podaci nisu u neskladu sa hipotezom da se kapaciteti kondenzatora pokoravaju zakonu normalne razdiobe.

<u>Zadatak.</u> Kontrolor ispituje 100 uzoraka od po 20 proizvoda da bi ustanovio broj x loših proizvoda. Provjerimo hipotezu H da za empirički osnovni skup vrijedi zakon binomne razdiobe. Na osnovu podataka dobivamo tablicu:

x_i	f_i	$f_i\ x_i$	$p(x_i)$	f_{ti}	$\lvert f_i - f_{ti} \rvert$	$(f_i - f_{ti})^2 / f_{ti}$
0	14	0	0,1216	12,16	1,84	0,279
1	25	25	0,2701	27,01	2,01	0,150
2	27	54	0,2852	28,52	1,52	0,082
3	23	69	0,1901	19,01	3,99	0,837
4	7	28	0,0898	8,98		
5	3 } 11	15	0,0319	3,19 } 13,06	2,06	0,325
6	1	6	0,0089	0,89		
	100	197		99,76		1,673

$$\bar{x} = \frac{1}{N} \sum f_i x_i = \frac{197}{100} = 1{,}97 \ ; \quad p = \frac{\bar{x}}{n} = \frac{1{,}97}{20} = 0{,}1098 = 0{,}1$$

$p(x_i)$ iz tablice za $p = 0{,}1; n = 20$

$f_{ti} = N \cdot p(x_i) = 100 \cdot p(x_i)$

$\chi^2 =$ 1,673 ; $k = n - 2 = 5 - 2 = 3$ (broj stupnjeva slobode)

$0{,}50 < P\left(\chi^2 > 1{,}673\right) < 0{,}70$

Pošto je $P\left(\chi^2 > 1{,}673\right) > 0{,}05$ hipoteza H je točna, tj. empirički osnovni skup pripada binomnoj razdiobi. Razlike frekvencija nisu prevelike i one su slučajne.

<u>Zadatak</u>. Predan je isti oglas u različitim novinama A, B, C. Nakon izvjesnog vremena stiglo je 13, 23, 18 odgovora. Može li se na temelju ovih podataka tvrditi da su novine B efikasnije u oglašavanju od ostalih?

Postavimo hipotezu H: novine A, B i C su jednako efikasne.

Ako je ova hipoteza ispravna onda svim novinama pripadaju iste teorijske frekvencije:

$$f_{ti} = \frac{54}{3} = 18,\ i = 1,2,3.$$

Napravimo tablicu i izračunajmo χ^2:

Novine	f_i	f_{ti}	$\lvert f_i - f_{ti}\rvert$	$(f_i - f_{ti})^2 / f_{ti}$
A	13	18	5	1,389
B	23	18	5	1,389
C	18	18	0	0
				2,778

U izračunavanju teorijskih frekvencija nismo koristili niti jedan od parametara koje bismo trebali procijeniti iz empiričkih podataka, pa je $r = 0$ i $k = n - r - 1 = 3 - 0 - 1 = 2$.

Iz tablice VII: $0{,}20 < P\left(\chi^2 > 2{,}778\right) < 0{,}30$. Budući da je ova vjerojatnost veća od 0,05 zaključujemo da razlike u prispjelim odgovorima nisu signifikantne, tj. empirički podaci nisu kontradiktorni s postavljenom hipotezom. Drugim riječima, ne bi bilo razumno tvrditi da su novine B efikasnije od ostalih.

χ^2 - test moguće je primijeniti kad god možemo izračunati teorijske frekvencije na bazi postavljene hipoteze. Često se javlja slučaj da su empiričke frekvencije svrstane u tablicu sa n redaka i m stupaca. Pokazat ćemo na primjeru kako se u tom slučaju određuju teorijske frekvencije. U tom slučaju stupanj slobode pripadne gama razdiobe treba računati po formuli:

$$k = (n-1)(m-1),$$

uz uvjet $m, n \geq 2$.

Zadatak. Iste finalne proizvode kontrolirala su tri kontrolora A, B i C. Svaki od njih pregledao je različitu količinu proizvoda i našao različit broj loših.

	Kontrolori			
Proizvodi	A	B	C	Ukupno
Dobri	549	368	186	1102
Loši	373	330	129	832
Ukupno	922	698	315	1935

Hipoteza H: sva tri kontrolora jednako dobro odvajaju dobre i loše proizvode.

Pretpostavimo da je kvaliteta proizvoda bio isti u sva tri vremenska intervala u kojima je provedena kontrola. Uz tu pretpostavku testirajmo hipotezu H.

Ako je hipoteza istinita broj loših proizvoda kod svakog kontrolora mora biti proporcionalan broju pregledanih proizvoda. Konstanta proporcionalnosti je $\frac{832}{1935}$. Slično vrijedi i za dobre proizvode gdje je konstanta proporcionalnosti $\frac{1103}{1935}$. Na osnovu toga su izračunate teorijske frekvencije u donjoj tablici:

	Kontrolori			
Proizvodi	A	B	C	Ukupno
Dobri	$\frac{1103}{1935}\cdot 922=525$	$\frac{1103}{1935}\cdot 698=398$	$\frac{1103}{1935}\cdot 315=180$	1102
Loši	$\frac{832}{1935}\cdot 922=397$	$\frac{832}{1935}\cdot 698=300$	$\frac{832}{1935}\cdot 315=135$	832
Ukupno	922	698	315	1935

Tablica pokazuje: teorijski ukupni broj dobrih, odnosno loših proizvoda jednak je empiričkom. Isto tako se slažu i ukupno pregledani proizvodi za svakog kontrolora. To služi kao kontrola računa.

Pomoću jedne i druge tablice izračunat ćemo χ^2.

	Kontrolori			
Proizvodi	A	B	C	Ukupno
Dobri	$\frac{(549-525)^2}{525}=1{,}10$	2,26	0,20	3,56
Loši	1,46	3,00	0,27	4,73
Ukupno	2,56	5,26	0,47	8,29

Dakle, χ^2 = 8,29. Stupanj slobode pripadne gama razdiobe je:

$k=(n-1)(m-1)=(2-1)(3-1)=2$. Iz tablice VII: $0{,}01 < P\left(\chi^2 > 8{,}29\right) < 0{,}02$.

Vjerojatnost je manja od 0,05, pa zaključujemo da se kontrolori signifikantno razlikuju u odvajanju dobrih i loših proizvoda. To znači da hipotezu o njihovoj jednakosti valja odbaciti.

Zadatak. Kod Poissonove razdiobe promatrali smo primjer "bombardiranja Londona" i dobili ove rezultate (prva 3 stupca):

x = broj bombi	f_x=broj sektora	f_{tx}	f_{tx} - f_x	(f_{tx} - f_x)^2	$(f_{tx}-f_x)^2 / f_{tx}$
0	229	226,74	2,26	5,1076	0,0225
1	211	211,39	-0,39	0,1521	0,007
2	93	98,54	-5,54	30,6916	0,3115
3	35	30,62	4,38	19,1844	0,6265
4	$\left.\begin{matrix}7\\1\end{matrix}\right\}8$	$\left.\begin{matrix}7{,}14\\1{,}57\end{matrix}\right\}8{,}71$			
5 i više			0,71	0,5041	0,0579
	576				χ^2 = 1,0191

U ovom smo primjeru spojili 4. i 5. razred zato jer u njima ima malo članova, pa je broj pregrupiranih razreda $n=5$. $k=n-2=5-2=3$.

Iz tablice za vrijednost χ^2 čitamo da je za χ^2 = 1,005 $P=0{,}80$.

Budući da je taj χ^2 blizu vrijednosti 1,0191, možemo uzeti da je vjerojatnost da se premaši χ^2 = 1,0191 jednaka 80%. Smatra se, prema tome, da je hipoteza potvrđena i da se stvarno radi o Poissonovoj razdiobi.

Zadatak. Empirički podaci o broju šibica (x_i) u kutijama (f_i) dani su u slijedećoj tablici:

x_i	43-44	45-46	47-48	49-50	51-52	53-54	55-56	57-58
f_i	4	6	16	22	24	18	6	4

Hipoteza: empirički podaci slijede zakon normalne razdiobe.

Provjerite hipotezu χ^2 testom.

Razr.	x_i	f_i	d_i	$f_i\ d_i$	$f_i d_i^2$	$\lvert x_i - \overline{x}\rvert$	u_i	$\varphi(u_i)$	f_{ti}	$\lvert f_i - f_{ti}\rvert$	$(f_i - f_{ti})^2 / f_{ti}$
43-44	43,5	4 } 10	-4	-16	64	7,08	2,19	0,09626	2,25 } 9,45	0,55	0,03201
45-46	45,5	6	-3	-18	54	5,08	1,57	0,11632	7,20		
47-48	47,5	16	-2	-32	64	3,08	0,95	0,25406	15,73	0,27	0,00463
49-50	49,5	22	-1	-22	22	1,08	0,33	0,37780	23,40	1,40	0,08376
51-52	51,5	24	0	0	0	0,92	0,28	0,38361	23,75	0,25	0,00263
53-54	53,5	18	1	18	18	2,92	0,90	0,26609	16,48	1,52	0,14019
55-56	55,5	6 } 10	2	12	24	4,92	1,52	0,12566	7,78 } 10,28		
57-58	57,5	4	3	12	36	6,92	2,14	0,04041	2,50	0,28	0,00763
		N=100		-46	282						$\chi_0^2 = 0{,}27085$

$$x_i = c \cdot d_i + x_0,$$

$$c = 2,\ d_i = \frac{x_i - x_0}{c},\ x_0 = 51{,}5$$

$$\overline{x} = x_0 + \frac{c}{N}\sum f_i d_i = 51{,}5 + \frac{2}{100}\cdot(-46) = 51{,}5 - 0{,}92 = \mathbf{50{,}58}$$

$$\sigma^2 = c^2\left[\frac{1}{N}\sum f_i d_i^2 - \left(\frac{1}{N}\sum f_i d_i\right)^2\right] = 4\left[\frac{282}{100} - \left(\frac{46}{100}\right)^2\right] = 4(2{,}82 - 0{,}2116) = 10{,}4336$$

σ = **3,23**

$$u_i = \frac{\lvert x_i - \overline{x}\rvert}{\sigma}$$

$$f_{ti} = \frac{N \cdot c}{\sigma}\cdot\varphi(u_i) = 61{,}92\cdot\varphi(u_i)$$

$$k = n - 3 = 6 - 3 = 3$$

$$0{,}95 < P\left(\chi^2 > 0{,}27085\right) < 0{,}98$$

Zadatak. Podaci o težini promatranog broja osoba dana je u slijedećoj tablici (prva dva stupca donje tablice).

Hipoteza: empirički podaci pokoravaju se normalnoj razdiobi.

Provjeriti hipotezu χ^2 testom.

Razr.	x_i	f_i	$f_i x_i$	$f_i x_i^2$	$\lvert x_i - \overline{x} \rvert$	u_i	$\varphi(u_i)$	f_{ti}	$\lvert f_i - f_{ti} \rvert$	$(f_i - f_{ti})^2 / f_{ti}$
45-49	47	1 } 11	47	2209	22,11	2,54	0,01885	0,69 } 9,86		
50-54	52	3 } 11	156	8112	17,11	1,96	0,05844	2,55 } 9,86	1,14	0,1318
55-59	57	7 } 11	399	22743	12,11	1,39	0,15183	6,62 } 9,86		
60-64	62	10	620	38440	7,11	0,82	0,28504	12,44	2,44	0,4785
65-69	67	19	1273	85291	2,11	0,24	0,38762	16,91	2,09	0,2586
70-74	72	15	1080	77760	2,89	0,33	0,37780	16,48	1,48	0,1329
75-79	77	12	924	71148	7,89	0,91	0,26369	11,50	0,50	0,0021
80-84	82	6 } 9	492	40344	12,89	1,48	0,13334	5,82 } 7,95		
85-89	87	3 } 9	261	22707	17,89	2,05	0,04879	2,13 } 7,95	1,05	0,1386
		N=76	5252	368754						$\chi_0^2 = 0{,}1425$

$$c = 5$$

$$\overline{x} = \frac{1}{N}\sum f_i x_i = \frac{5252}{76} = 69{,}11$$

$$\sigma^2 = \frac{1}{N}\sum_i f_i x_i^2 - \overline{x}^2 = \frac{368754}{76} - 69{,}11^2 = 75{,}8342$$

σ = 8,71

$$u_i = \frac{\lvert x_i - \overline{x} \rvert}{\sigma}$$

$$f_{ti} = \frac{N \cdot c}{\sigma} \cdot \varphi(u_i) = 43{,}63 \cdot \varphi(u_i)$$

$$k = n - 3 = 6 - 3 = 3$$

$$0{,}70 < P\left(\chi^2 > 1{,}142\right) < 0{,}80$$

10. FUNKCIJA IZVODNICA

Neka je kontinuirana slučajna varijabla x zadana funkcijom vjerojatnosti $f(x)$. Radi izračunavanja momenata varijable x uvodi se tzv. **funkcija izvodnica**:

$$g_x(t) = \int_{-\infty}^{+\infty} e^{tx} f(x)dx .$$

Ako je x diskontinuirana slučajna varijabla tada treba integral zamijeniti sumom preko svih vrijednosti x_i, a $f(x)dx$ vjerojatnošću $p(x_i)$.

Napomenimo odmah da za neke slučajne varijable integral na desnoj strani **divergira**, pa te varijable nemaju funkciju izvodnicu. Ako funkcija $g_x(t)$ postoji ona je jednoznačno određena funkcijom vjerojatnosti $f(x)$.

Funkciju izvodnicu moguće je pisati i u drugom obliku. Naime, iz desne strane izraza za $g_x(t)$ vidimo da je to očekivanje varijable e^{tx}, tj. $g_x(t) = \mathrm{E}(e^{tx})$.

Budući da je $e^{tx} = 1 + \frac{tx}{1!} + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \ldots$, vrijedi:

$$g_x(t) = \int_{-\infty}^{+\infty} f(x)dx + \frac{t}{1!}\int_{-\infty}^{+\infty} xf(x)dx + \frac{t^2}{2!}\int_{-\infty}^{+\infty} x^2 f(x)dx + \ldots = 1 + \frac{m_1}{1!}t + \frac{m_2}{2!}t^2 + \ldots$$

Pri tom su m_i ($i = 1,2,\ldots$) pomoćni momenti i-tog reda varijable x.

$\Rightarrow$ $[g_x(t)]_{t=0} = m_i$.

Time je ujedno opravdan naziv funkcije $g_x(t)$ jer se iz te funkcije **izvode pomoćni momenti** varijable x.

$$\Rightarrow \qquad M_1 = m_1 = \mu = E(x) = g_x'(t)\big|_{t=0}$$

$$M_2 = \sigma^2 = V(x) = m_2 - m_1^2 = g_x''(0) - g_x'(0)^2$$

$$\left.\begin{aligned} M_3 &= m_3 - 3m_2m_1 + 2m_1^3 \\ M_4 &= m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4 \end{aligned}\right\} \qquad \begin{aligned} m_3 &= g_x'''(0) \\ m_4 &= g_x''''(0) \end{aligned}$$

Teorem: $g_{ax+b}(t) = e^{bt} \cdot g_x(at)$.

Zadatak . Diskontinuirana slučajna varijabla prima vrijednosti $x = 1,2,3,...$ s vjerojatnošću $f(x) = A \cdot 3^{-x}$. Odredite A, $g_x(t)$ i $V(x)$.

Rješenje:

a) $\sum_i f(x_i) = 1$

$$A\left(3^{-1} + 3^{-2} + 3^{-3} + ...\right) = 1$$

$$A\left(\frac{1}{3} + \frac{1}{9} + \frac{1}{27} + ...\right) = 1$$

radi se o geometrijskom redu: $q = \frac{1}{3}, a_1 = \frac{1}{3} \Rightarrow S = \frac{\frac{1}{3}}{1 - \frac{1}{3}} = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$

$$\Rightarrow A \cdot \frac{1}{2} = 1 \Rightarrow A = 2 \Rightarrow f(x) = 2 \cdot 3^{-x}$$

b) $g_x(t) = \sum_i e^{tx_i} f(x_i) = 2\left(e^t 3^{-1} + e^{2t} 3^{-2} + e^{3t} 3^{-3} + ...\right) = 2\left(\frac{e^t}{3} + \frac{e^{2t}}{9} + \frac{e^{3t}}{27} + ...\right) =$

$$= 2 \cdot \frac{\frac{e^t}{3}}{1 - \frac{e^t}{3}} = 2 \cdot \frac{e^t}{3 - e^t}$$

c) $m_1 = \left[g_x{}'(t)\right]_{t=0}$

$$g_x'(t) = 2 \cdot \frac{e^t(3 - e^t) - e^t(-e^t)}{(3 - e^t)^2} = 2 \cdot \frac{3e^t - e^{2t} + e^{2t}}{(3 - e^t)^2} = \frac{6e^t}{(3 - e^t)^2}$$

$$m_1 = \left[\frac{6e^t}{(3 - e^t)^2}\right]_{t=0} = \frac{6}{4} = \frac{3}{2}$$

$$g_x''(t) = \frac{6e^t(3 - e^t)^2 - 6e^t \cdot 2(3 - e^t)(-e^t)}{(3 - e^t)^4} = \frac{18e^t - 6e^{2t} + 12e^{2t}}{(3 - e^t)^3} = \frac{18e^t + 6e^{2t}}{(3 - e^t)^3}$$

$$m_2 = \left[g_x{}''(t)\right]_{t=0} = \frac{18 + 6}{8} = \frac{24}{8} = 3$$

$$V(x) = m_2 - m_1^2 = 3 - \frac{9}{4} = \frac{3}{4}$$

Zadatak. $f(x)=\frac{Aa^x}{(1+a)^{x+1}}$; $x=0,1,2,3,...$; A, $g_x(t), V(x)=?$

Rješenje:

$$\sum_{x=0}^{\infty} f(x)=\sum_{x=0}^{\infty}\frac{Aa^x}{(1+a)^{x+1}}=\frac{A}{1+a}\cdot\sum_{x=0}^{\infty}\left(\frac{a}{1+a}\right)^x=\frac{A}{1+a}\cdot\frac{1}{1-\frac{a}{1+a}}=\frac{A}{1+a\cdot\frac{1+a-a}{1+a}}=A=1$$

$$g_x(t)=\sum e^{tx}f(x)=\sum_{x=0}^{\infty}e^{tx}\frac{a^x}{(1+a)^{x+1}}=\frac{1}{1+a}\cdot\frac{1}{1-\frac{e^t a}{1+a}}=\frac{1}{1+a-e^t a}$$

$$g_x'(t)=\frac{ae^t}{(1+a-e^t a)^2}$$

$$g_x''(t)=\frac{ae^t(1+a-e^t a)-2ae^t(1+a-e^t a)(-e^t a)}{(1+a-e^t a)^4}$$

$$\Rightarrow \qquad m_1=\left[g_x{}'(t)\right]_{t=0}=\frac{a}{1^2}=a$$

$$m_2=\left[g_x{}''(t)\right]_{t=0}=\frac{a\cdot 1-2a(-a)}{1^4}=a+2a^2$$

$$V(x)=m_2-m_1^2=a+2a^2-a^2=a+a^2$$

Zadatak. Funkcija distribucije slučajne varijable je: $F(x)=\begin{cases}0, & x<0\\ \frac{x(2a-x)}{a^2}, & 0\le x\le a\\ 1, & x>a\end{cases}$.

Izračunaj funkciju gustoće $f(x), m_1, m_2$ i $V(x)$.

Rješenje:

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{0}^{x} f(x)dx \quad\Rightarrow\quad \frac{dF(x)}{dx}=f(x)$$

$$\left.\begin{array}{l}\lim\limits_{x\to 0}F(x)=0\\ \lim\limits_{x\to a}F(x)=\frac{a\cdot a}{a^2}=1\end{array}\right\}$$ dakle, konstanta *a* može biti bilo koja pozitivna vrijednost

$$f(x)=\frac{dF(x)}{dx}=\frac{1}{a^2}\left[2a-x+x(-1)\right]=\frac{1}{a^2}(2a-2x)=\frac{2}{a^2}(a-x)$$

$$f(x)=\begin{cases}0, & x<0\\ \frac{2}{a^2}(a-x), & 0\le x\le a\\ 0, & x>a\end{cases}$$

Ispunjen je osnovni uvjet $\int\limits_{-\infty}^{+\infty} f(x)dx=1$:

$$\int\limits_0^a \frac{2}{a^2}(a-x)dx=\frac{2}{a^2}\left(ax-\frac{1}{2}x^2\right)\Bigg|_0^a=\frac{2}{a^2}\left(a^2-\frac{a^2}{2}\right)=\frac{2}{a^2}\cdot\frac{a^2}{2}=1$$

$$g_x(t)=\int\limits_{-\infty}^{+\infty} e^{tx} f(x)dx=\int\limits_0^a e^{tx}\cdot\frac{2}{a^2}(a-x)dx=\frac{2}{a}\int\limits_0^a e^{tx}dx-\frac{2}{a^2}\int\limits_0^a xe^{tx}dx=...=\frac{2}{a^2t^2}\left(e^{at}-1-at\right)$$

$$m_1=\left[g_x'(t)\right]_{t=0}=...=\frac{a}{3}$$

$$m_2=\left[g_x''(t)\right]_{t=0}=...=\frac{a^2}{6}$$

$$V(x)=m_2-m_1^2=\frac{a^2}{6}-\frac{a^2}{9}=\frac{a^2}{18}$$