Verovatnoća i statistika 2 + 3, 7 ESPB

FIMEK - Saobraćajno inženjerstvo

školska 2023/24

Literatura

[1] Kovačević I, Verovatnoća i statistika sa zbirkom zadataka, Univerzitet Singidunum, 2011

[2] Crvenković Z. L., Carić M., Subotić B., Poslovna statistika, FIMEK Novi Sad 2009.

[3] Ghilezan et. al., Zbirka rešenih zadataka iz Verovatnoće i statistike, CMS, NS, 2009.

Bodovi i datumi

	Kol. 1	Kol. 2	Predavanja	Aktivnost	Ispit	Σ
MAX	25	25	10	10	30	100
MIN	10	10	0	0	0	51
Datumi	28. XI	16. I	1+2+3+4	5. XII	2024.	2024.

Prostor verovatnoće

Posmatramo neprazan **skup ishoda** Ω i skup njegovih podskupova, **prostor događaja** $\mathcal{A}$, za koji važi: $\Omega \in \mathcal{A}$, $A \in \mathcal{A} \Rightarrow \overline{A} := \Omega \backslash A \in \mathcal{A}$, $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$.

Elemente $A, B, \ldots \in \mathcal{A}$ nazivamo **događaji**. Ω je **siguran događaj**. Za događaj A, njegov komplement $\Omega \backslash A =: \overline{A}$ je **suprotan događaj**. $\emptyset$ je **nemoguć događaj**.

PRIMER 1 *Za skup* $\Omega = \{1,2,3,4,5,6\}$, $\mathcal{A} = \{\emptyset, \Omega, \{1,2\}, \{3,4,5,6\}\}$ *je* σ*-polje događaja.*

Ako je $A\,B := A \cap B = \emptyset$, kažemo da su događaji A i B **disjunktni**.

Verovatnoća je funkcija koja prostoru događaja $\mathcal{A}$ pridružuje nenegativne brojeve i zadovoljava ($A_1, A_2 \in \mathcal{A} \ \wedge \ A_1 A_2 = \emptyset) \ \Rightarrow \ P(A_1 \cup A_2) = P(A_1) + P(A_2)$ i $P(\Omega) = 1$.

Uređenu trojku $(\Omega, \mathcal{A}, P)$ nazivamo **prostor verovatnoće**.

PRIMER 2 *Za skup* $\Omega = \{1,2,3,4,5,6\}$ *i* $\mathcal{A} = \{\emptyset, \Omega, \{1,2\}, \{3,4,5,6\}\}$, *funkcija*

$$P = \begin{pmatrix} \emptyset & \Omega & \{1,2\} & \{3,4,5,6\} \\ 0 & 1 & 1/3 & 2/3 \end{pmatrix}$$

je verovatnoća, odnosno, $(\Omega, \mathcal{A}, P)$ *je prostor verovatnoće.*

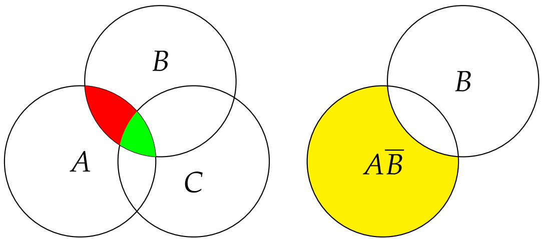
Osobine verovatnoće

Za proizvoljne događaje A i B:

- $P(A) \leq 1$
- $P(\emptyset) = 0$
- $P(\overline{A}) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(AB)$
- $A \subseteq B \Rightarrow P(A) \leq P(B)$

Primer 3 *Neka je:*
$P(A) = 0{,}6$; $P(B) = 0{,}6$;
$P(AB) = 0{,}4$; $P(BC) = 0{,}25$;
$P(ABC) = 0{,}15$;

Izračunati $P(A\overline{B})$, $P(AB\overline{C})$ *i* $P(\overline{A}BC)$.



PRIMER 4 *Neka je skup ishoda konačan: $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, i neka je σ-polje skup svih podskupova $\mathcal{A} = \mathcal{P}(\Omega)$.*

Neka je $p_k = P(\{\omega_k\}) \geq 0$, $k = 1, 2, \dots, n$ i $\sum_{k=1}^{n} p_k = 1$.

Verovatnoća je definisana

$$P(A) = \sum_{k:\omega_k \in A} p_k.$$

Ovaj prostor zovemo ***diskretni prostor verovatnoće****.*

PRIMER 5 *Ako u primeru 4 važi i $p_1 = p_2 = \cdots = p_n = 1/n$, dobijamo $P(A) = \#A/\#\Omega$, to je* ***klasična definicija*** *verovatnoće. U klasičnoj definiciji verovatnoće se kaže da je verovatnoća* broj povoljnih podeljen sa brojem mogućih ishoda. *Klasična definicija verovatnoće se koristi kada imamo konačno mnogo jednako verovatnih ishoda, odnosno, kada se vrši slučajan izbor.*

PRIMER 6 *U kutiji se nalazi 10 kuglica sa brojem 0, 11 kuglica sa brojem 1 i 12 kuglica sa brojem 2. Na slučajan način se izvlači kuglica i posmatra izvučeni broj. Opisati skup ishoda i prostor verovatnoće.*

PRIMER 7 *Ako se u prethodnom primeru na slučajan način izvlače 3 kuglice bez vraćanja, kolika je verovatnoća da su izvučena tri različita broja?*

Primer 8 *Baca se kockica za igru. Događaji su $A =$ pao je paran broj, $B =$ pao je broj manji od 4. Rečima opisati događaje $A\overline{B}$ i $A \cup B$ i izračunati njihove verovatnoće.*

Neka je $\Omega \subseteq \mathbb{R}^n$ Euklidski prostor.

Neka je $\mathcal{A}$ skup podskupova od Ω koji su merljivi merom m i neka je $m(\Omega) > 0$.

Geometrijsku verovatnoću za $A \subseteq \Omega$ definišemo: $P(A) = m(A)/m(\Omega)$.

Onda je $(\Omega, \mathcal{A}, P(\cdot))$ prostor verovatnoće.

Primer 9 *Oko kocke je opisana lopta. Na slučajan način se bira tačka u lopti. Kolika je verovatnoća da je izabrana tačka u kocki?*

Skup čija verovatnoća je 0 zovemo **nemoguć događaj**. Na primer, $\emptyset$ je nemoguć događaj.

Ako je $P(A) = 1$, kažemo da je A **skoro siguran skup (događaj)**. Ako nešto važi na skupu verovatnoće 1, kažemo da **skoro sigurno važi**.

Primer 10 *Tri dečaka i tri devojčice sedaju na slučajan način u red sa 6 mesta. Kolika je verovatnoća da nema dve osobe istog pola koje sede jedna do druge?*

Primer 11 *Iz špila od 52 karte na slučajan način se izvlači jedna karta. Kolika je verovatnoća da je izvučena karta dama ili herc?*

Ako u prostoru verovatnoće $(\Omega, \mathcal{A}, P(\cdot))$ za događaje $A, B \in \mathcal{A}$ važi $P(AB) = P(A)P(B)$, kažemo da su A i B **nezavisni**.

PRIMER 12 *Da li su u prethodnom primeru događaji $A =$ izvučena karta je dama i $B =$ izvučena karta je herc nezavisni?*

PRIMER 13 *Novčić se baca tri puta. Bacanja su nazavisna. Izračunati verovatnoće p_k da će pasti k grbova za $k = 0, 1, 2, 3$.*

PRIMER 14 *Novčić se baca dok se ne dobije grb. Izračunati ver. da bude paran broj bacanja.*

PRIMER 15 *(Bernulijeva shema) Pozitivna realizacija eksperimenta u svim pokušajima ima istu verovatnoću $p \in (0, 1)$. Eksperiment se vrši n puta. Kolika je verovatnoća da će biti k, za $0 \leq k \leq n$ pozitivnih realizacija?*

PRIMER 16 *U odeljenju od 30 đaka ima 18 dečaka. Na slučajan način se bira petočlana komisija. Kolika je verovatnoća da u komisiji ima (barem) 2 dečaka?*

Uslovna verovatnoća

Neka je $(\Omega, \mathcal{A}, P(\cdot))$ prostor verovatnoće i neka je $A \in \mathcal{A}$ i $P(A) > 0$.
Definišemo za $B \in \mathcal{A}$ **verovatnoću pod uslovom da se desio događaj** A:

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

Onda je $(\Omega, \mathcal{A}, P(\cdot|A))$ prostor verovatnoće.
Prebrojiva familija događaja $H_1, H_2, \ldots$ čini **potpun sistem događaja** ako su događaji disjunktni po parovima i ako važi $\bigcup_{j=1}^{\infty} H_j = \Omega$.

Formula totalne verovatnoće za potpuni sistem događaja $H_1, H_2, \ldots$ i događaj A:

$$P(A) = \sum_{j=1}^{\infty} P(H_j)\,P(A|H_j).$$

Bejzova (Bayes) formula: $\quad P(H_j|A) = \dfrac{P(H_j)\,P(A|H_j)}{\sum_{j=1}^{\infty} P(H_j)\,P(A|H_j)}.$

Primer 17 *Simptom X se pojavljuje usled bolesti A, B i C. Poznato je da se bolest A, B i C pojavljuju kod redom 10%, 5%, 20% populacije. Bolesti A, B i C isključuju jedna drugu. Simptom X se u slučaju bolesti A razvija u 90% slučajeva, u slučaju bolesti B razvija se u 95% slučajeva, i u slučaju bolesti C razvija u 75% slučajeva.*

Kolika je verovatnoća da će se kod slučajno odabranog čoveka pojaviti simptom X?

Ako se pojavio simptom X, kolika je verovatnoća da ima bolest A, B, odnosno C?

Primer 18 *Od n novčića jedan je neispravan: ima grb sa obe strane. Na slučajan način se bira novčić i baca k puta. Kolika je verovatnoća da svih k puta padne grb?*

Ako je svih k puta pao grb, kolika je verovatnoća da je u pitanju neispravan novčić?

Uopštena formula preseka:

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

Primer 19 *Ako se u primeru 6 izvlače 3 kuglice jedna za drugom, kolika je verovatnoća da izvučeni brojevi čine rastući niz?*

Primer 20 *Iz špila od 52 karte se na slučajan način, jedna po jedna, izvlače dve karte. Kolika je verovatnoća da je druga izvučena karta herc?*

PRIMER 21 *Na deonici pravog puta su postavljena dva semafora. Vozila stižu u slučajnim momentima. Prvi semafor propušta 80% vozila. Drugi semafor propušta 75% vozila koja nisu stala na prvom semaforu i 60% vozila koja su stala na prvom semaforu.*

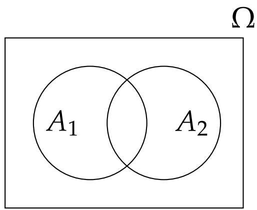
Kolika je verovatnoća da će vozilo proći deonicu sa tačno jednim zaustavljanjem?

Označimo događaje:
$A_1 =$ vozilo je zaustavljeno na 1. semaforu, $A_2 =$ vozilo je zaustavljeno na 2. semaforu.
$\overline{A_1} =$ vozilo je propušteno na 1. semaforu, $\overline{A_2} =$ vozilo je propušteno na 2. semaforu.

Događaj da će vozilo proći deonicu sa jednim zaustavljanjem je $A = \overline{A_1}\,A_2 \cup A_1\,\overline{A_2}$.

U tekstu je dato: $P(\overline{A_1}) = 0{,}80$; $P(\overline{A_2}|\overline{A_1}) = 0{,}75$; $P(\overline{A_2}|A_1) = 0{,}60$.

$$\begin{aligned}
P(A) &= P(\overline{A_1}\,A_2) + P(A_1\,\overline{A_2}) = \\
&= P(\overline{A_1}) \cdot P(A_2|\overline{A_1}) + P(A_1) \cdot P(\overline{A_2}|A_1) = \\
&= 0{,}80 \cdot 0{,}25 + 0{,}20 \cdot 0{,}60 = 0{,}32.
\end{aligned}$$



Za vežbu popuniti verovatnoćama polja na Venovom dijagramu.

Slučajne promenljive

Ako $X : \Omega \to \mathbb{R}$, za $S \subseteq \mathbb{R}$, **inverzna sliku** od S je

$$X^{-1}(S) = \{\omega \in \Omega : X(\omega) \in S\} .$$

Definicija

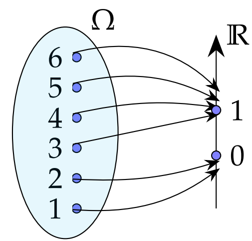
Ako je $(\Omega, \mathcal{A}, P)$ prostor verovatnoće i $X : \Omega \to \mathbb{R}$ zadovoljava:

$$\forall x \in \mathbb{R},\ X^{-1}((-\infty, x]) \in \mathcal{A} ,$$

kažemo da je X **slučajna promenljiva**.

Primer 22 *Za prostor verovatnoće iz primera 2, možemo definisati slučajnu promenljivu koja registruje sa 1 da li je broj veći od 2 (inače je nula):*

$$X = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} .$$



Funkcija raspodele

Za slučajnu promenljivu X nad $(\Omega, \mathcal{A}, P)$ definišemo **funkciju raspodele**:

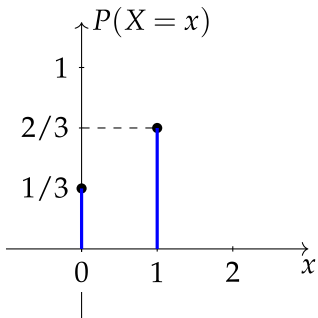
$$F(x) = P\left(X^{-1}((-\infty, x])\right) = P(\{\omega : X(\omega) \leq x\}) = P(X \leq x)\,.$$

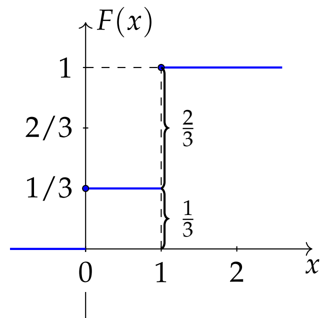
Primer 23 *Za slučajnu promenljivu X iz primera 22 funkcija raspodele je*

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{3}, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

Zakon raspodele

$$X : \begin{pmatrix} 0 & 1 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$





Osobine funkcije raspodele

1. $\lim\limits_{x \to -\infty} F(x) = 0$
2. $\lim\limits_{x \to \infty} F(x) = 1$
3. $x_1 < x_2 \;\Rightarrow\; F(x_1) \leq F(x_2)$
4. $\forall a \in \mathbb{R},\ \lim\limits_{x \to a^+} F(x) = F(a)$
5. $P(a < X \leq b) = F(b) - F(a)$

Diskretne slučajne promenljive

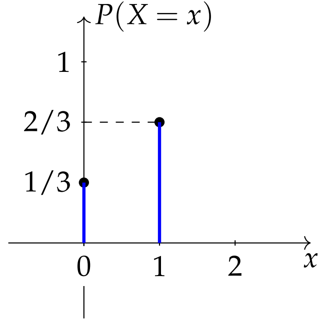
Neka je X slučajna promenljiva nad $(\Omega, \mathcal{A}, P)$. Sliku skupa ishoda Ω označavamo sa $\mathcal{R}_X$. Kažemo da je X **diskretna slučajna promenljiva** ako je slika $\mathcal{R}_X$ konačan ili prebrojiv skup. Ako je $\mathcal{R}_X = \{x_1, x_2, \ldots\}$ onda $\Omega = \bigcup\limits_{n=1}^{\infty} (X = x_n)$ sledi da je $1 = \sum\limits_{n=1}^{\infty} P(X = x_n)$. Možemo uvesti oznake $p_n = P(X = x_n)$. Funkciju $x_n \mapsto p_n$ zovemo **zakon raspodele** slučajne promenljive X i zapisujemo $X : \begin{pmatrix} x_1 & x_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix}$. Važi $F(x) = \sum\limits_{n : x_n \leq x} p_n$.

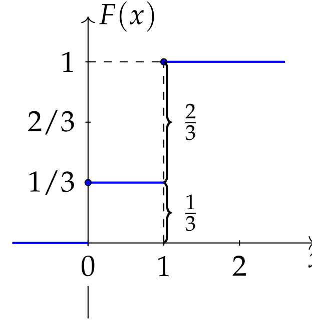
Bernulijeva raspodela sa parametrom $p \in (0,1)$ je $X : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$.

PRIMER 24 *U primeru 22 posmatra se slučajna promenljiva X koja uzima vrednost* 1 *ako u bacanju kocke za igru padne broj veći u 2, inače uzima vrednost 0. Ovde je skup* $\Omega = \{1,2,3,4,5,6\}$, $A = X^{-1}(\{1\}) = \{3,4,5,6\}$. *Po klasičnoj definiciji verovatnoće:* $p = P(A) = \frac{\#A}{\#\Omega} = \frac{4}{6} = \frac{2}{3}$. *Zakon raspodele je* $X : \begin{pmatrix} 0 & 1 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$.

Funkcija raspodele

$$F(x) = \begin{cases} 0 & x < 0, \\ 1/3 & 0 \le x < 1, \\ 1 & x \ge 1. \end{cases}$$





Binomna raspodela sa parametrima $p \in (0,1)$ i $n \in \mathbb{N}$ u oznaci $\mathcal{B}(n,p)$:

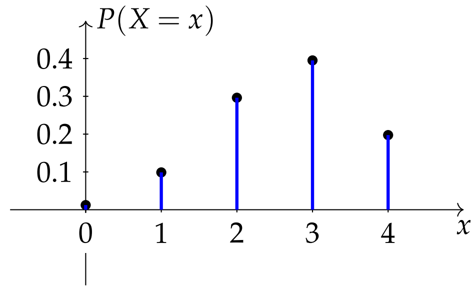
$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \; k \in \{0,1,\ldots,n\}.$$

PRIMER 25 *U primeru 15 (Bernulijeva shema) je opisana ova raspodela. Recimo, ako posmatramo koliko puta je pao broj veći od dva u četiri nezavisna bacanja kockice za igru, dobija se Binomna raspodela sa* $p = \frac{2}{3}, n = 4$. $X : \mathcal{B}\left(4, \frac{2}{3}\right)$. *Zakon raspodele je:*

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \binom{4}{0}\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^4 & \binom{4}{1}\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^3 & \binom{4}{2}\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^2 & \binom{4}{3}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^1 & \binom{4}{4}\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^0 \end{pmatrix}$$

Kad se izračuna i nacrta:

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 0.0123 & 0.0988 & 0.2963 & 0.3951 & 0.1975 \end{pmatrix}$$



$$E(X) = 0 \cdot 0.0123 + 1 \cdot 0.0988 + 2 \cdot 0.2963 + 3 \cdot 0.3951 + 4 \cdot 0.1975 = 2.6667 = 4 \cdot \frac{2}{3}$$

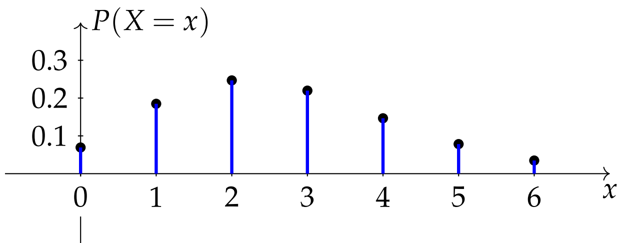
Poasonova raspodela sa parametrom $\lambda \in (0,\infty)$ u oznaci $\mathcal{P}(\lambda)$:

$$P(X=k) = \frac{\lambda^k}{k!}e^{-\lambda}, k = 0,1,\ldots$$

Kada se posmatra učestalo pojavljivanje događaja kroz vreme i kad pojava jednog događaja ne utiče na vreme pojave sledećeg, onda broj događaja koji su se pojavili do nekog fiksnog momenta ima Poasonovu raspodelu. Parametar λ predstavlja učestalost tih događaja.

Primer 26 *Broj klijenata koji se pojave na šalterima za kupovinu karata ima Poasonovu raspodelu sa prosekom* 160 *klijenata u na sat. Zakon raspodele slučajne promenljive* $X =$ *broj klijenata u minutu ima Poasonovu raspodelu sa* $\lambda = 160/60 = 2.6667$.

$$X : \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & \ldots \\ 0.0693 & 0.1849 & 0.2468 & 0.2197 & 0.1466 & 0.0783 & 0.0348 & \ldots \end{pmatrix}$$

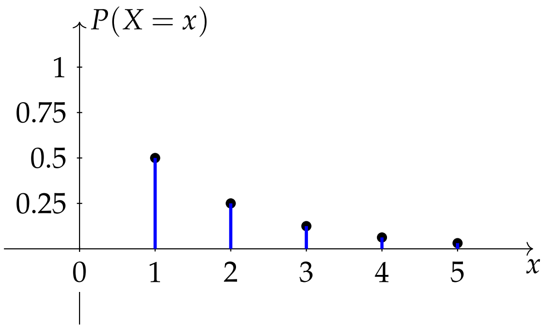


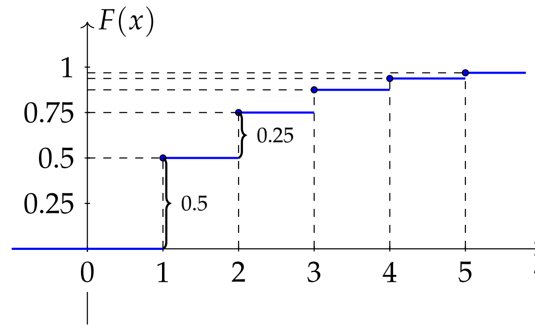
Geometrijska raspodela sa parametrom $p \in (0,1)$ u oznaci $\mathcal{G}(p)$:

$$P(X = k) = p\,(1-p)^{k-1},\ k = 1,2,\ldots$$

Primer 27 *Novčić se baca dok se ne dobije grb. Slučajna promenljiva X predstavlja broj bacanja. Naći zakon raspodele X. Izračunati verovatnoću da je bio paran broj bacanja.*

$$X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & \ldots \\ 0.5 & 0.25 & 0.125 & 0.0625 & 0.03125 & \ldots \end{pmatrix}$$





Paran broj bacanja $=: A = \{2,4,6,\ldots\}$. $P(A) = \sum_{k=1}^{\infty} P(X = 2k) = \sum_{k=1}^{\infty} \left(\frac{1}{4}\right)^k = \frac{\frac{1}{4}}{1-\frac{1}{4}} = \frac{1}{3}$.

Primer 28 *U kutiji se nalazi 10 kuglica sa brojem 0, 11 kuglica sa brojem 1 i 12 kuglica sa brojem 2. Na slučajan način se izvlači dve kuglice. Slučajna promenljiva X predstavlja zbir izvučenih brojeva.*

Naći zakon raspodele i očekivanje slučajne promenljive X.

$$P(X=0)=P(\{0+0\})=\tfrac{10}{33}\cdot\tfrac{9}{32},$$
$$P(X=1)=P(\{0+1,1+0\})=2\cdot\tfrac{10}{33}\cdot\tfrac{11}{32},$$
$$P(X=2)=P(\{0+2,2+0,1+1\})=2\cdot\tfrac{10}{33}\cdot\tfrac{12}{32}+\tfrac{11}{33}\cdot\tfrac{10}{32},$$
$$P(X=3)=P(\{1+2,2+1\})=2\cdot\tfrac{11}{33}\cdot\tfrac{12}{32},$$
$$P(X=4)=P(\{2+2\})=\tfrac{12}{33}\cdot\tfrac{11}{32}.$$

$$X:\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{10}{33}\cdot\frac{9}{32} & 2\cdot\frac{10}{33}\cdot\frac{11}{32} & 2\cdot\frac{10}{33}\cdot\frac{12}{32}+\frac{11}{33}\cdot\frac{10}{32} & 2\cdot\frac{11}{33}\cdot\frac{12}{32} & \frac{12}{33}\cdot\frac{11}{32}\end{pmatrix}$$

$$X:\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{90}{33\cdot 32} & \frac{220}{33\cdot 32} & \frac{350}{33\cdot 32} & \frac{264}{33\cdot 32} & \frac{132}{33\cdot 32}\end{pmatrix}$$

$$E(X)=0\cdot\tfrac{10}{33}\cdot\tfrac{9}{32}+1\cdot 2\cdot\tfrac{10}{33}\cdot\tfrac{11}{32}+2\cdot\left(2\cdot\tfrac{10}{33}\cdot\tfrac{12}{32}+\tfrac{11}{33}\cdot\tfrac{10}{32}\right)+$$
$$+3\cdot 2\cdot\tfrac{11}{33}\cdot\tfrac{12}{32}+4\cdot\tfrac{12}{33}\cdot\tfrac{11}{32}=\tfrac{2240}{33\cdot 32}=2.1212$$

Neprekidne slučajne promenljive

Kažemo da je slučajna promenljiva **apsolutno neprekidnog tipa** ako postoji funkcija $\varphi : \mathbb{R} \to [0,\infty)$ takva da je

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

Funkciju φ nazivamo **funkcija gustine raspodele** slučajne promenljive X.

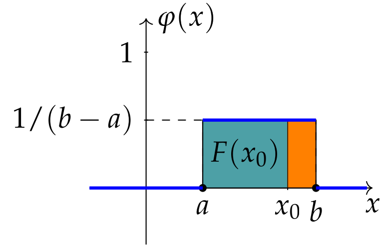
Osobine gustine raspodele

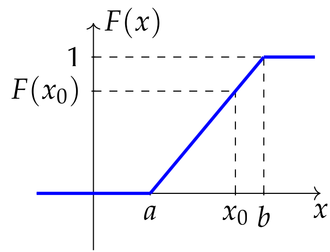
1. $\varphi(x) = F'(x)$
2. $\int_{-\infty}^{\infty} \varphi(t)dt = 1$
3. $\int_a^b \varphi(x)dx = F(b) - F(a) = P(a < X < b) = P(a \leq X < b) = P(a \leq X \leq b) = P(a < X \leq b)$

Uniformna raspodela $\mathcal{U}(a,b)$, $a < b \in \mathbb{R}$

$$\varphi(x) = \begin{cases} \frac{1}{b-a}, & x \in (a,b), \\ 0, & x \notin (a,b); \end{cases}$$

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



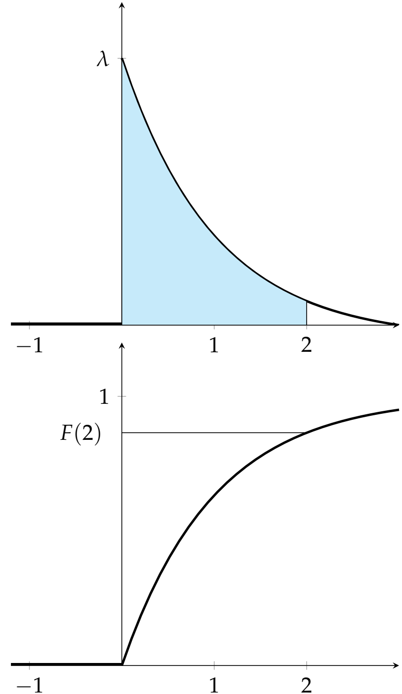


Eksponencijalna raspodela

$$\mathcal{E}(\lambda),\ \lambda \in (0,\infty)$$

$$\varphi(x) = \begin{cases} 0, & x < 0, \\ \lambda e^{-\lambda x}, & x \geq 0 \end{cases}$$

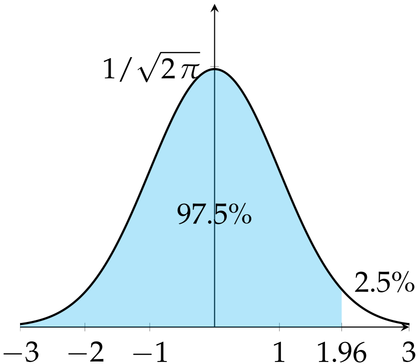
$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

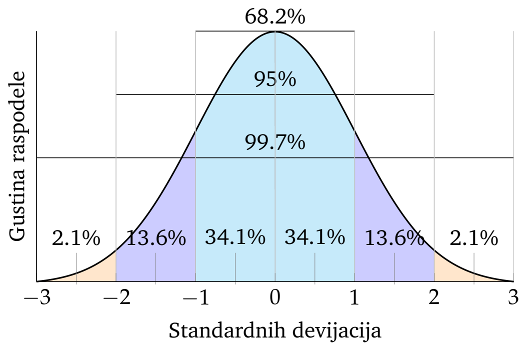


Normalna (Gausova) raspodela $\mathcal{N}(\mu,\sigma)$, $m,\sigma \in \mathbb{R}$, $\sigma > 0$

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x \in \mathbb{R}; \qquad F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt.$$

Za $\mu = 0$ i $\sigma = 1$, $\mathcal{N}(0,1)$, standardna normalna raspodela $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$.





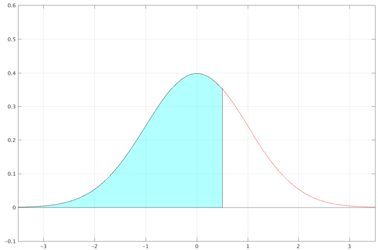
Tablice Na primer $P(X \leq 1.960) = \Phi(1.960) = 0.975$, $\Phi^{-1}(0.975) = 1.960$.

$p = P(X \leq z)$

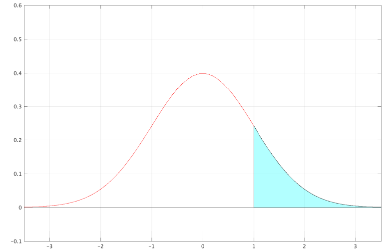
p	.75	.90	.95	.975	.990	.995	.9995
z	.674	1.282	1.645	1.960	2.326	2.576	3.291

Primer 29 *Ako $X : \mathcal{N}(0,1)$, iz tablica funkcije Φ očitati vrednosti $P(X \leq 0.55), P(X > 1), P(|X| < 2)$ i naći vrednost x za koju je $P(X \leq x) = F(x) = 0.975$.*

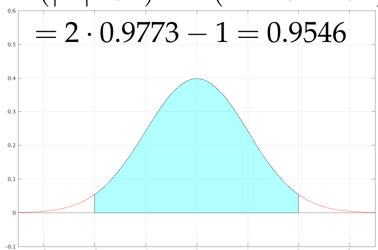
$$P(X \leq 0.55) = 0.7088$$



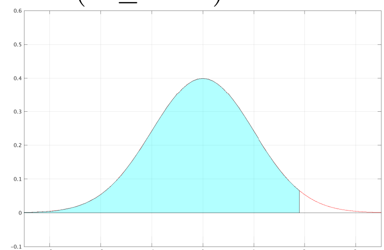
$$P(X > 1) = 1 - P(X \leq 1) = 1 - 0.8413 = 0.1587$$



$$P(|X| < 2) = P(-2 < X < 2) =$$
$$= 2 \cdot 0.9773 - 1 = 0.9546$$



$$P(X \leq 1.960) = 0.975$$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981

Standardizacija Normalne raspodele

Neka $X:\mathcal{N}(0,1)$ i neka je $Y=aX+b$, $a\neq 0$. Onda slučajna promenljiva Y ima Normalnu (Gausovu) raspodelu $Y:\mathcal{N}(b,|a|)$.

Ako $X:\mathcal{N}(\mu,\sigma)$, onda $X^{*}=\dfrac{X-\mu}{\sigma}:\mathcal{N}(0,1)$.

Primer 30 *Neka* $X:\mathcal{N}(0.5,2)$. *Izračunati verovatnoće* $P(X\leq 0.55), P(X>1), P(|X|<2)$.

$$P(X\leq 0.55)=P((X-0.5)/2\leq(0.55-0.5)/2)=P(X^{*}\leq 0.025)=0.510,$$

$$\begin{aligned}P(X>1)&=1-P(X\leq 1)=1-P((X-0.5)/2\leq(1-0.5)/2)=1-P(X^{*}\leq 0.25)=\\&=1-\Phi(0.25)=1-0.5987=0.4013,\end{aligned}$$

$$\begin{aligned}P(|X|<2)&=P(-2<X<2)=P((-2-0.5)/2<(X-0.5)/2<(2-0.5)/2)=\\&=P(-1.25<X^{*}<0.75)=\Phi(0.75)-\Phi(-1.5)=0.7734-(1-0.8944)=0.6678.\end{aligned}$$

Numeričke karakteristike slučajnih promenljivih

Matematičko očekivanje

Za diskretnu slučajnu promenljivu $X : \begin{pmatrix} x_1 & x_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix}$, definišemo $E(X) = \sum\limits_{i=1}^{\infty} x_i \, p_i$.

Za neprekidnu slučajnu promenljivu $X : \varphi(x)$, definišemo $E(X) = \int\limits_{-\infty}^{\infty} x \, \varphi(x) \, dx$.

Osobine matematičkog očekivanja

1. $X = c, \, c = const \;\Rightarrow\; E(X) = c$

2. $Y = f(X), \, E(Y) = E(f(X)) = \begin{cases} \sum\limits_{i=1}^{\infty} f(x_i) \, p_i, & X \text{ diskretna} \\ \int\limits_{-\infty}^{\infty} f(x) \, \varphi(x) \, dx, & X \text{ neprekidna} \end{cases}$

3. $E(X + Y) = E(X) + E(Y)$

4. $E(c \, X) = c \, E(X)$

Disperzija (Varijansa)

$$D(X) = E\left((X - E(X))^2\right) = E\left(X^2\right) - (E(X))^2 = \begin{cases} \sum\limits_{i=1}^{\infty} (x_i - E(X))^2 p_i \\ \int\limits_{-\infty}^{\infty} (x - E(X))^2 \varphi(x) dx \end{cases}$$

Osobine disperzije

1. $X = c, c = const \Leftrightarrow D(X) = 0$
2. $D(X) \geq 0$
3. $c = const \Rightarrow D(cX) = c^2 D(X), D(X + c) = D(X)$
4. X i Y nezavisne $\Rightarrow D(X + Y) = D(X) + D(Y)$

Standardna devijacija

$\sigma(X) = \sqrt{D(X)}$

Standardizacija slučajne promenljive

$X^* = \frac{X - E(X)}{\sigma(X)} \Rightarrow E(X^*) = 0, D(X^*) = 1$

PRIMER 31 *Posmatra se X, broj koji padne u slučajnom bacanju kocke za igru. Naći zakon raspodele, očekivanje i disperziju slučajne promenljive X.*

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

$$E(X) = 1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{6} + 3 \cdot \tfrac{1}{6} + 4 \cdot \tfrac{1}{6} + 5 \cdot \tfrac{1}{6} + 6 \cdot \tfrac{1}{6} = 3.5$$
$$D(X) = (1 - 3.5)^2 \cdot \tfrac{1}{6} + \cdots + (6 - 3.5)^2 \cdot \tfrac{1}{6} = 2.916667$$

$X : Ber(p)$	$E(X) = p$	$D(X) = p(1-p)$	$p \in (0,1)$
$X : \mathcal{B}(n,p)$	$E(X) = np$	$D(X) = np(1-p)$	$n \in \mathbb{N},\ p \in (0,1)$
$X : \mathcal{P}(\lambda)$	$E(X) = \lambda$	$D(X) = \lambda$	$\lambda > 0$
$X : \mathcal{G}(p)$	$E(X) = \frac{1}{p}$	$D(X) = \frac{1}{p^2}$	$p \in (0,1)$
$X : \mathcal{U}(a,b)$	$E(X) = \frac{a+b}{2}$	$D(X) = \frac{(b-a)^2}{12}$	$a < b$
$X : \mathcal{E}(\lambda)$	$E(X) = \frac{1}{\lambda}$	$D(X) = \frac{1}{\lambda^2}$	$\lambda > 0$
$X : \mathcal{N}(\mu,\sigma)$	$E(X) = m$	$D(X) = \sigma^2$	$\sigma > 0$

Normalna raspodela $\mathcal{N}(\mu,\sigma)$ (Gausova)

Slučajna promenljiva ima **Normalnu raspodelu**, $\mathcal{N}(\mu,\sigma)$, $\mu,\sigma \in \mathbb{R}$, $\sigma > 0$ ako

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \qquad E(X) = \mu, \qquad D(X) = \sigma^2.$$

Centriranje normalne raspodele: $X : \mathcal{N}(\mu,\sigma) \;\Leftrightarrow\; X^* = \dfrac{X-\mu}{\sigma} : \mathcal{N}(0,1)$.

Teorema: Ako su X i Y nezavisne slučajne promenljive sa normalnim raspodelama $X : \mathcal{N}(\mu_1,\sigma_1)$, $Y : \mathcal{N}(\mu_2,\sigma_2)$, onda $X \pm Y : \mathcal{N}\left(\mu_1 \pm \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$

Posledica 1: Ako nezavisne slučajne promenljive imaju raspodelu $X_k : \mathcal{N}(\mu,\sigma)$, $k = 1,\ldots,n$, onda slučajna promenljiva $X_1 + X_2 + \cdots + X_n$ ima normalnu raspodelu $\mathcal{N}(n\mu, \sqrt{n}\sigma)$.

Posledica 2: Ako nezavisne slučajne promenljive imaju raspodelu $X_k : \mathcal{N}(\mu,\sigma)$, $k = 1,\ldots,n$, onda slučajna promenljiva $Z = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ ima normalnu raspodelu $Z : \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$, a slučajna promenljiva $Z^* = \frac{Z-\mu}{\sigma}\sqrt{n} = \frac{\frac{1}{n}\sum_{k=1}^{n} X_k - \mu}{\sigma}\sqrt{n}$ ima normalnu raspodelu $Z^* : \mathcal{N}(0,1)$.

Centralna granična teorema

Ako je $X_1, X_2, \ldots$ niz nezavisnih slučajnih promenljivih sa istom raspodelom čije su očekivanje i disprezija redom $E(X_k) = a$ i $D(X_k) = s^2$, $0 < s < \infty$, onda za svako x

$$\lim_{n\to\infty} P\left(\frac{\sum\limits_{k=1}^{n} X_k - E\left(\sum\limits_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum\limits_{k=1}^{n} X_k\right)}} \le x\right) = \lim_{n\to\infty} P\left(\frac{\sum\limits_{k=1}^{n} X_k - n\,a}{s\sqrt{n}} \le x\right) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x).$$

Moavr-Laplasova teorema

$$X : \mathcal{B}(n,p) \;\Rightarrow\; \lim_{n\to\infty} P\left(\frac{X - n\,p}{\sqrt{n\,p\,(1-p)}} \le x\right) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x).$$

Poasonova aproksimacija

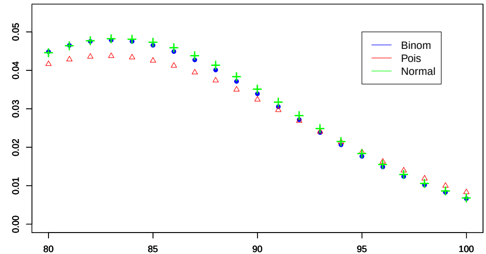
Za konačno k, ako je $\lim\limits_{n\to\infty} n\,p = \lambda = const$, važi $\lim\limits_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} = \dfrac{\lambda^k}{k!} e^{-\lambda}$.

Primer 32 *Kolika je verovatnoća da je broj šestica u 500 bacanja kocke između 80 i 100?*

Označimo X broj šestica u 500 bacanja. Onda $X : \mathcal{B}(500, \frac{1}{6})$.

$$P = \sum_{k=80}^{100} \binom{500}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{500-k} = \underline{0.6518}.$$

(R: sum(dbinom((80:100),500,1/6)))



Moavr-Laplasova aproksimacija

$$P(80 < X \leq 100) = P\left(\frac{80-500\frac{1}{6}}{\sqrt{500\frac{1}{6}\frac{5}{6}}} < \frac{X-500\frac{1}{6}}{\sqrt{500\frac{1}{6}\frac{5}{6}}} \leq \frac{100-500\frac{1}{6}}{\sqrt{500\frac{1}{6}\frac{5}{6}}}\right) \approx \Phi\left(\frac{100-500\frac{1}{6}}{\sqrt{500\frac{1}{6}\frac{5}{6}}}\right) - \Phi\left(\frac{80-500\frac{1}{6}}{\sqrt{500\frac{1}{6}\frac{5}{6}}}\right) =$$
$$= \Phi(2) - \Phi(-0.4) = \Phi(2) - 1 + \Phi(0.4) =$$
$$= 0.9772 - 1 + 0.6554 = 0.6326$$

Poasonova aproksimacija, $\lambda = n \cdot p = 500 \cdot \frac{1}{6} = 83.33333$

$$P = \sum_{k=80}^{100} \binom{500}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{500-k} \approx \sum_{k=80}^{100} \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=80}^{100} \frac{83.33333^k}{k!} \exp(-83.33333) = 0.6242$$

Statistika

Normalna raspodela $\mathcal{N}(\mu,\sigma)$ (Gausova)

Slučajna promenljiva ima **Normalnu raspodelu**, $\mathcal{N}(\mu,\sigma)$, $\mu,\sigma \in \mathbb{R}$, $\sigma > 0$ ako

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \qquad E(X) = \mu, \qquad D(X) = \sigma^2.$$

Centriranje normalne raspodele: $X : \mathcal{N}(\mu,\sigma) \;\Leftrightarrow\; X^* = \dfrac{X-\mu}{\sigma} : \mathcal{N}(0,1).$

Linearna kombinacija normalnih raspodela

Teorema: X i Y nezavisne i $X : \mathcal{N}(\mu_1,\sigma_1)$, $Y : \mathcal{N}(\mu_2,\sigma_2) \;\Rightarrow\; X \pm Y : \mathcal{N}\left(\mu_1 \pm \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$

Posledica: Ako nezavisne slučajne promenljive imaju raspodelu $X_k : \mathcal{N}(\mu_k,\sigma_k)$, $k = 1,2,\ldots,n$, onda slučajna promenljiva $Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$ ima normalnu raspodelu $\mathcal{N}(\mu,\sigma)$, gde je $\mu = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$ i $\sigma^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2$.

χ^2 raspodela (Pirsonova)

Ako X ima Normalnu raspodelu, $X : \mathcal{N}(0,1)$ i $Y = X^2$, onda je gustina raspodele Y:

$$\varphi_Y(y) = \begin{cases} 0, & y \leq 0, \\ \frac{1}{\sqrt{2\pi y}} e^{-y/2}, & y > 0, \end{cases}$$

kažemo da Y ima **hi-kvadrat raspodelu** sa jednim stepenom slobode, $Y : \chi^2_1$.

Slučajna promenljiva sa **hi-kvadrat raspodelom** sa n stepeni slobode, $X : \chi^2_n$, ima gustinu:

$$\varphi(x) = \frac{x^{n/2-1} e^{-x/2}}{2^{n/2}\Gamma(n/2)}, \, x > 0, \quad E(X) = n, \quad D(X) = 2n.$$

Teorema: Ako su $Y_1 : \chi^2_n$ i $Y_2 : \chi^2_m$ nezavisne slučajne promenljive, onda $Y = Y_1 + Y_2 : \chi^2_{n+m}$.

Posledica 1: Ako su $X_1, X_2, \ldots, X_n$ nezavisne slučajne promenljive sa normalnom $\mathcal{N}(0,1)$ raspodelom, onda $Y = X_1^2 + X_2^2 + \cdots + X_n^2$ ima χ^2_n raspodelu.

Posledica 2: Ako su $X_1, X_2, \ldots, X_n$ nezavisne slučajne promenljive sa normalnom $\mathcal{N}(\mu, \sigma)$ raspodelom, onda $Y = \frac{1}{\sigma^2}\left((X_1 - \mu)^2 + (X_2 - \mu)^2 + \cdots + (X_n - \mu)^2\right)$ ima χ^2_n raspodelu.

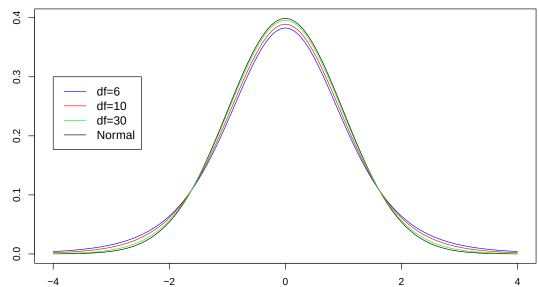
t raspodela (Studentova)

Slučajna promenljiva $T : t_n$ ima **Studentovu raspodelu** sa n stepeni slobode ako ima gustinu

$$\varphi(t) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\ \Gamma(n/2)\ (1+t^2/n)^{(n+1)/2}}.$$

Drugi zapis: $\varphi(t) = \left(\sqrt{n}\, B\left(\frac{1}{2}, \frac{n}{2}\right) \left(1 + \frac{t^2}{n}\right)^{(n+1)/2}\right)^{-1}$. $E(T) = 0$, $D(T) = \frac{n}{n-2}$, $n > 2$.

Teorema:
Ako su $X : \mathcal{N}(0,1)$ i $Y : \chi^2_n$ nezavisne sl. prom, onda $T = \frac{X}{\sqrt{\frac{Y}{n}}}$ ima t_n raspodelu.



Statistika, osnovni pojmovi

Populacija je skup svih elemenata koje ispitujemo.

Obeležje je numerička karakteristika elementa. Modeliramo ga slučajnom promenljivom.

Uzorak je odabrani deo populacije na kojem ispitujemo realizovanu vrednost obeležja X.

Prost slučajni uzorak je n-dimenzionalna slučajna promenljiva čije komponente su nezavisne i imaju raspodelu posmatranog obeležja $(X_1, X_2, \ldots, X_n)$.

Uzoračka funkcija raspodele $F(x_1, x_2, \ldots, x_n) = F(x_1)\,F(x_2)\cdots F(x_n)$

Realizovane vrednosti slučajnih promenljivih obeležavamo malim slovima $X_i \to x_i$.

Realizovana vrednost prostog slučajnog uzorka $(X_1, X_2, \ldots, X_n) \to (x_1, x_2, \ldots, x_n)$

Statistika je funkcija uzorka, $Y = h(X_1, X_2, \ldots, X_n)$.

Realizovana vrednost statistike je $y = h(x_1, x_2, \ldots, x_n)$.

Kao transformacija slučajnih promenljivih, **statistika je slučajna promenljiva**.

Raspodela statistike obeležja koje modeliramo se koristi za statističko zaključivanje.

Važne statistike uzorka $(X_1, X_2, \dots, X_n)$ obeležja X

Aritmetička sredina uzorka

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \quad \Rightarrow \quad \begin{cases} E(\bar{X}_n) = E(X), \quad D(\bar{X}_n) = \frac{1}{n} D(X) \\ X : \mathcal{N}(\mu, \sigma) \Rightarrow \bar{X}_n : \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow Z = \frac{\bar{X}_n - \mu}{\sigma}\sqrt{n} : \mathcal{N}(0,1) \end{cases}$$

Uzoračka disperzija (varijansa)

$$\bar{S}_n^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X}_n)^2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \bar{X}_n^2,\ E(\bar{S}_n^2) = \frac{n-1}{n} D(X),\ \bar{S}_n = \sqrt{\bar{S}_n^2}.$$

$$X : \mathcal{N}(\mu, \sigma) \quad \Rightarrow \quad \begin{cases} Y = \frac{n\bar{S}_n^2}{\sigma^2} : \chi^2_{n-1} = \frac{(n-1)\bar{S}_n^{2\prime}}{\sigma^2} : \chi^2_{n-1}, \\ T = \frac{\bar{X}_n - \mu}{\bar{S}_n}\sqrt{n-1} = \frac{\bar{X}_n - \mu}{\bar{S}_n{}'}\sqrt{n} : t_{n-1} \end{cases}$$

Korigovana varijansa $\bar{S}_n^{2\prime} = \dfrac{1}{n-1}\displaystyle\sum_{k=1}^{n}(X_k - \bar{X}_n)^2,\ E(\bar{S}_n^{2\prime}) = D(X),\ \bar{S}_n{}' = \sqrt{\bar{S}_n^{2\prime}}$

Intervalni uzorak

Intervalni uzorak nastaje grupisanjem elemenata početnog uzorka u intervale I_i.

Ako imamo granice intervala I_i, odnosno deobene tačke m_i, $i = 0, 1, \ldots k$ i broj elemenata uzorka u intervalu i: **frekvencije** f_i, $i = 1, 2, \ldots k$, kažemo da je to **intervalni** uzorak.

Delimična rekonstrukcija početnog uzorka sredinama intervala: smatramo da imamo f_i komada elemenata jednakih $x_i = (m_i + m_{i-1})/2$, sredini i-tog intervala.

Ponekad se anketiranjem podaci prikupljaju u intervalni uzorak.

Formule za računanje aritmetičke sredine i varijanse intervalnog uzorka sa sredinama x_i, $i = 1, 2, \ldots, k$ i frekvencijama f_i, $i = 1, 2, \ldots, k$ su:

$$n = \sum_{i=1}^{k} f_i\,, \quad \bar{x}_n = \frac{1}{n}\sum_{i=1}^{k} x_i f_i\,, \quad \bar{s}_n^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x}_n)^2 f_i\,, \quad \bar{s}_n = \sqrt{\bar{s}_n^2}\,, \quad \bar{s}_n^{2\prime} = \frac{n}{n-1}\bar{s}_n^2\,.$$

Primer: Anketirani su kupci o vremenu u godinama do prvog kvara na bojleru

I_i	$[0,1]$	$(1,2]$	$(2,3]$	$(3,5]$	$(5,10]$	$(10,20]$
f_i	15	11	7	7	6	4

$n = 15 + 11 + 7 + 7 + 6 + 4 = 50$, $\quad \bar{x}_n = (0.5 \cdot 15 + 1.5 \cdot 11 + \cdots + 15 \cdot 4)/50 = 3.49$,

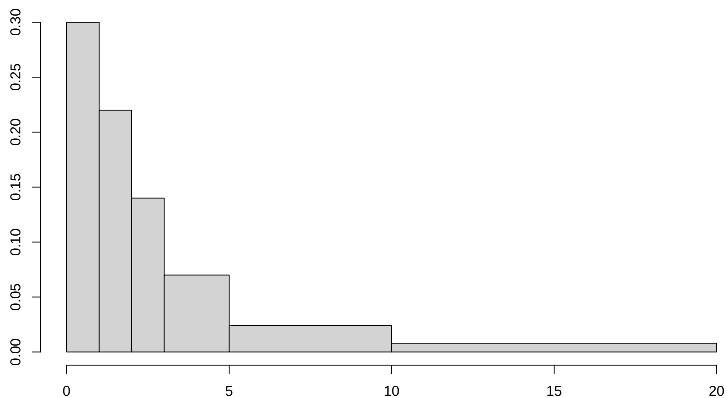
$\bar{s}_n^2 = ((0.5 - 3.49)^2 \cdot 15 + \cdots + (15 - 3.49)^2 \cdot 4)/50 = 16.2549$, $\quad \bar{s}_n = \sqrt{16.2549} = 4.0317$.

Primer: Anketirani su kupci o vremenu u godinama do prvog kvara na bojleru

Dopunili smo tabelu sa četiri nove vrste ($\bar{f}_i = f_i/h_i$, $p_i = \bar{f}_i/n$):

I_i	$[0,1]$	$(1,2]$	$(2,3]$	$(3,5]$	$(5,10]$	$(10,20]$
f_i	15	11	7	7	6	4
x_i	0.5	1.5	2.5	4	7.5	15
h_i	1	1	1	2	5	10
$\bar{f}_i$	15	11	7	3.5	1.2	0.4
p_i	0.30	0.22	0.14	0.07	0.024	0.008

Histogram



Tabelarni uzorak

Za diskretno obeležje uzorak se može zadati tabelom u kojoj se navode vrednosti obležja i **frekvencije** pojavljivanja u uzorku.

Formule za računanje aritmetičke sredine i standardne devijacije tabelarnog uzorka sa vrednostima x_i, $i = 1,2,\ldots,k$ i frekvencijama f_i, $i = 1,2,\ldots,k$ su:

$$n = \sum_{i=1}^{k} f_i \, , \quad \bar{x}_n = \frac{1}{n}\sum_{i=1}^{k} x_i f_i \, , \quad \bar{s}_n^2 = \frac{1}{n}\sum_{i=1}^{k} (x_i - \bar{x}_n)^2 f_i \, , \quad \bar{s}_n^{2\,\prime} = \frac{n}{n-1}\bar{s}_n^2 \, .$$

Primer: Kockica je bačena 100 puta. Rezultati bacanja:

x_i	1	2	3	4	5	6
f_i	15	17	18	21	15	14

$n = 15 + 17 + 18 + 21 + 15 + 14 = 100,$

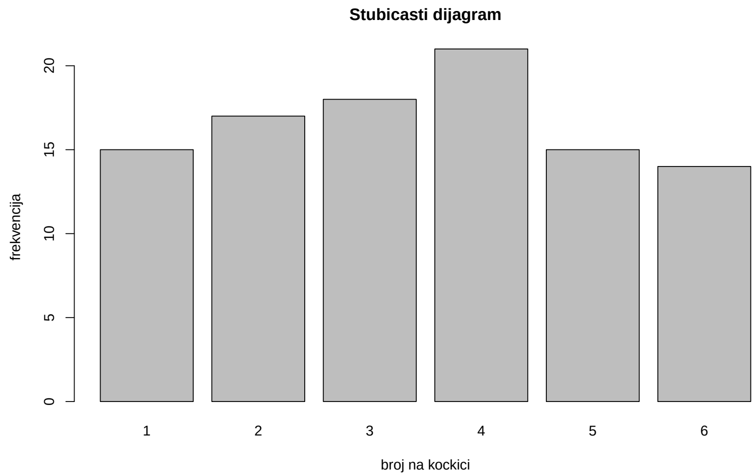
$\bar{x}_n = (1 \cdot 15 + 2 \cdot 17 + \cdots + 6 \cdot 14) \, /100 = 3.46,$

$\bar{s}_n^2 = \left((1 - 3.46)^2 \cdot 15 + (2 - 3.46)^2 \cdot 17 + \cdots + (6 - 3.46)^2 \cdot 14\right) /100 = 2.6284$

Stubičasti dijagram i pita

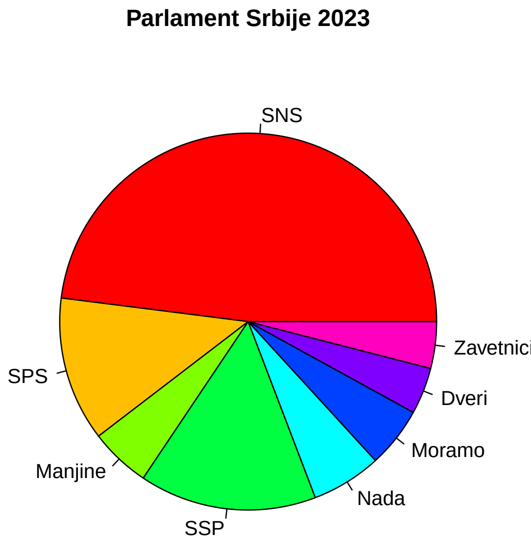
Primer: Kockica je bačena 100 puta. Rezultati bacanja:

x_i	1	2	3	4	5	6
f_i	15	17	18	21	15	14



Primer: Sadržaj Parlamenta Srbije pre raspuštanja 2023.

SNS	SPS	Manjine	SSP	Nada	Moramo	Dveri	Zavetnici
120	31	13	38	15	13	10	10



Intervali poverenja

Za obeležje X raspodele $F(x,\theta)$, sa uzorkom $(X_1, X_2, \ldots, X_n)$, ako su $U_1 = u_1(X_1, X_2, \ldots, X_n)$ i $U_2 = u_2(X_1, X_2, \ldots, X_n)$ statistike za koje važi $P(U_1 < \theta < U_2) = 1 - \alpha$, gde je $1 - \alpha$ unapred zadat **nivo poverenja**, onda je (U_1, U_2) **interval poverenja** širine $1 - \alpha$.

Za očekivanje μ obeležja $X : \mathcal{N}(\mu, \sigma)$, σ poznato

Ako $X : \mathcal{N}(\mu, \sigma)$ onda $\bar{X}_n : \mathcal{N}(\mu, \sigma/\sqrt{n})$, odnosno, onda $Z = \dfrac{\bar{X}_n - \mu}{\sigma}\sqrt{n} : \mathcal{N}(0,1)$.

Označimo sa $z_{1-\alpha}$ vrednost za koju $P\left(|Z| < z_{1-\alpha}\right) = 1 - \alpha$. $z_{1-\alpha} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ je $1 - \frac{\alpha}{2}$ kvantil.

Onda je $U_1 = \bar{X}_n - z_{1-\alpha}\dfrac{\sigma}{\sqrt{n}}$, $U_2 = \bar{X}_n + z_{1-\alpha}\dfrac{\sigma}{\sqrt{n}}$. Izraz $\dfrac{\sigma}{\sqrt{n}}$ nazivamo **standardna greška**.

Za očekivanje μ obeležja $X : \mathcal{N}(\mu, \sigma)$, σ nepoznato

Ako $X : \mathcal{N}(\mu, \sigma)$ onda $T = \dfrac{\bar{X}_n - \mu}{\bar{S}_n}\sqrt{n-1} = \dfrac{\bar{X}_n - \mu}{\bar{S}'_n}\sqrt{n} : t_{n-1}$ (ima Studentovu raspodelu).

Označimo sa t_β vrednost za koju je $P\left(|T| < t_{1-\beta}\right)$. ($t_{1-\alpha}$ je $1 - \frac{\alpha}{2}$ kvantil raspodele t_{n-1}.)

$U_1 = \bar{X}_n - t_{1-\alpha}\dfrac{\bar{S}_n}{\sqrt{n-1}}$, $U_2 = \bar{X}_n + t_{1-\alpha}\dfrac{\bar{S}_n}{\sqrt{n-1}}$. **Standardna greška** je $\dfrac{\bar{S}_n}{\sqrt{n-1}} = \dfrac{\bar{S}'_n}{\sqrt{n}}$.

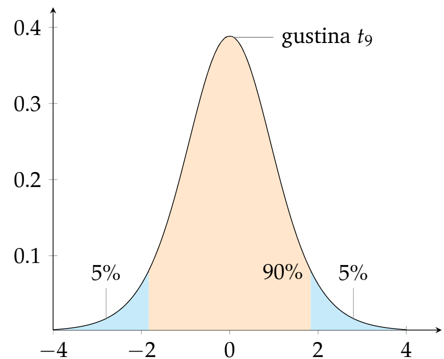
PRIMER 33 *Naći 90% interval poverenja za srednju vrednost m obeležja sa normalnom $\mathcal{N}(m,\sigma)$ raspodelom*

(a) Ako je poznato $\sigma = 3$, *(b) ako je σ nepoznato,*

za uzorak $(17.3, 12.9, 10.4, 11.9, 9.9, 8.9, 9.9, 6.3, 12.9, 9.4)$.

($n = 10$, $\bar{x}_n = 10.98$, $\bar{s}'_n = 2.973139$, $z_{0.9} = 1.645$, $t_{0.9} = 1.833$)

```
> x<-c(17.3, 12.9, 10.4, 11.9, 9.9,
        8.9, 9.9, 6.3, 12.9, 9.4)
> n<-10; xn<-mean(x); sn<-sd(x)
> z<-qnorm(.95); t<-qt(.95,9)
> xn-z*3/sqrt(10)
[1] 9.419555
> xn+z*3/sqrt(10)
[1] 12.54045
> xn-t*sn/sqrt(10)
[1] 9.256527
> xn+t*sn/sqrt(10)
[1] 12.70347
```



PRIMER 34 *Merena je visina muških studenata druge godine, rezultati su dati u tabeli:*

173	174	174.5	175	176	177	177.5	178	178.5	179	179.5	180	180.5	181	181.5	182	182.5	183.5	184.5	185	186	186.5	189	189.5
1	2	1	2	1	1	1	1	5	3	1	3	2	2	3	3	2	2	2	3	3	2	2	2

Naći 95% interval visine studenata.

Rešenje:

$1-\alpha = 0.95$, $t_{1-\alpha;49} = 2.010$, $n = 50$,

$\bar{x}_n = \frac{1}{50}(173\cdot 1 + 174\cdot 2 + 174.5\cdot 1 + 175\cdot 2 + 176\cdot 1 + 177\cdot 1 + 177.5\cdot 1 + 178\cdot 1 + 178.5\cdot 5 + 179\cdot 3 + 179.5\cdot 1 + 180\cdot 3 + 180.5\cdot 2 + 181\cdot 2 + 181.5\cdot 3 + 182\cdot 3 + 182.5\cdot 2 + 183.5\cdot 2 + 184.5\cdot 2 + 185\cdot 3 + 186\cdot 3 + 186.5\cdot 2 + 189\cdot 2 + 189.5\cdot 2) = \underline{181.21}$,

$\bar{s}_n^2 = \frac{1}{50}((173-181.21)^2\cdot 1 + (174-181.21)^2\cdot 2 + (174.5-181.21)^2\cdot(1 + (175-181.21)^2\cdot 2 + (176-181.21)^2\cdot(1 + (177-181.21)^2\cdot(1 + (177.5-181.21)^2\cdot(1 + (178-181.21)^2\cdot(1 + (178.5-181.21)^2\cdot 5 + (179-181.21)^2\cdot 3 + (179.5-181.21)^2\cdot(1 + (180-181.21)^2\cdot 3 + (180.5-181.21)^2\cdot 2 + (181-181.21)^2\cdot 2 + (181.5-181.21)^2\cdot 3 + (182-181.21)^2\cdot 3 + (182.5-181.21)^2\cdot 2 + (183.5-181.21)^2\cdot 2 + (184.5-181.21)^2\cdot 2 + (185-181.21)^2\cdot 3 + (186-181.21)^2\cdot 3 + (186.5-181.21)^2\cdot 2 + (189-181.21)^2\cdot 2 + (189.5-181.21)^2\cdot 2) = \underline{17.7309}$

$$\bar{x}_n \pm t_{1-\alpha;49}\frac{\bar{s}_n}{\sqrt{n-1}} = 181.21 \pm 2.010\cdot\frac{\sqrt{17.7309}}{\sqrt{49}} = \underline{(180.0009, 182.4191)}.$$

Studentove i Gausove tablice t i z vrednosti

Za $X : t_n$ raspodelu $P = P(X \leq t)$, za $n \to \infty$, $t_n \to \mathcal{N}$, $t \to z$

$n \backslash P$	.75	.90	.95	.975	.990	.995	.9995
1	1.000	3.078	6.314	12.706	31.821	63.657	636.619
2	.816	1.886	2.920	4.303	6.965	9.925	31.599
3	.765	1.638	2.353	3.182	4.541	5.841	12.924
...							
8	.706	1.397	1.860	2.306	2.896	3.355	5.041
9	.703	1.383	**1.833**	2.262	2.821	3.250	4.781
10	.700	1.372	1.812	2.228	2.764	3.169	4.587
...							
30	.683	1.310	1.697	2.042	2.457	2.750	3.646
40	.681	1.303	1.684	2.021	2.423	2.704	3.551
49	.680	1.299	1.677	**2.010**	2.405	2.680	3.500
60	.679	1.296	1.671	2.000	2.390	2.660	3.460
120	.677	1.289	1.658	1.980	2.358	2.617	3.373
...							
z	.674	1.282	1.645	1.960	2.326	2.576	3.291

Testiranje hipoteza

Statistički testovi

Hipoteza H_0 protiv H_1

	Usvojena H_0	Usvojena H_1
Tačna H_0	OK	Greška I vrste
Tačna H_1	Greška II vrste	OK

Parametarske hipoteze

- Zadaje se prag značajnosti α (recimo $\alpha = 5\% = 0.05$)
- Bira se parametar raspodele obeležja (θ).
- Nalazi se ocena parametra $\theta = h(x_1, \ldots, x_n)$.
- Nalazi se kritična oblast C (koja daje nedozvoljene vrednosti) parametra, takva da je $P_{H_0}(\hat{\theta} = h(X_1, \ldots, X_n) \in C) = \alpha$.
- Računa se statistika uzorka $\theta = h(x_1, x_2, \ldots, x_n)$ i ako $\theta \in C$, odbacujemo H_0 (i usvajamo H_1)

Testiranje $H_0(\mu = \mu_0)$ **protiv** $H_1(\mu \neq \mu_0)$ **za** $X : \mathcal{N}(\mu, \sigma)$, σ **poznato**

Koristimo interval poverenja za nepoznato očekivanje m obeležja sa normalnom raspodelom, σ poznato, širine $1 - \alpha$.

$$\mu_0 \in \mathbb{R} \setminus \left(\bar{x}_n \mp z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right) \Leftrightarrow z := \frac{|\bar{x}_n - \mu_0|}{\sigma} \sqrt{n} > z_{1-\alpha} \Leftrightarrow \alpha^* := P_{H_0} \left(|X^*| > \frac{|\bar{x}_n - \mu_0|}{\sigma} \sqrt{n} \right) < \alpha$$

Testiranje $H_0(\mu = \mu_0)$ **protiv** $H_1(\mu \neq \mu_0)$ **za** $X : \mathcal{N}(\mu, \sigma)$, σ **nepoznato**

Koristimo interval poverenja za nepoznato očekivanje μ obeležja sa normalnom raspodelom, σ nepoznato, širine $1 - \alpha$.

$$\mu_0 \in \mathbb{R} \setminus \left(\bar{x}_n \mp t_{1-\alpha} \frac{\bar{s}'_n}{\sqrt{n}} \right) \Leftrightarrow t := \frac{|\bar{x}_n - \mu_0|}{\bar{s}'_n} \sqrt{n} > t_{1-\alpha} \Leftrightarrow \alpha^* := P_{H_0} \left(|T| > \frac{|\bar{x}_n - \mu_0|}{\bar{s}'_n} \sqrt{n} \right) < \alpha$$

$z_{1-\alpha} = \Phi^{-1} \left(1 - \frac{\alpha}{2} \right)$ je $1 - \frac{\alpha}{2}$ kvantil.

$t_{1-\alpha}$ je $1 - \frac{\alpha}{2}$ kvantil raspodele t_{n-1}.

Testiranje jednakosti srednjih vrednosti obeležja X_1 i X_2 dva uzorka

$H_0(\mu_1 = \mu_2)$ **protiv** $H_1(\mu_1 \neq \mu_2)$**,** σ_1**,** σ_2 **poznato,** $X_1 : \mathcal{N}(\mu_1, \sigma_1)$**,** $X_2 : \mathcal{N}(\mu_2, \sigma_2)$

Koristimo statistiku $Z := \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ koja ima $\mathcal{N}(0,1)$ raspodelu

$H_0(\mu_1 = \mu_2)$ **protiv** $H_1(\mu_1 \neq \mu_2)$ **(T-test),** $X_1 : \mathcal{N}(\mu_1, \sigma_1)$**,** $X_2 : \mathcal{N}(\mu_2, \sigma_2)$

Koristimo statistiku $T := \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\bar{S}_1^{2\prime}}{n_1} + \frac{\bar{S}_2^{2\prime}}{n_2}}}$, koja približno ima t_ν raspodelu,
gde se za ν uzima $\nu = n_1 + n_2 - 2$ ili procena Welcha.

PRIMER 35 *Testirati hipotezu o jednakosti srednjih vrednosti uzorka* $(9.81, 9.83, 10.43, 11.13, 9.70, 9.59, 10.88, 10.97, 9.35, 9.34, 9.41, 9.95)$ *i uzorka* $(9.85, 9.30, 9.08, 8.07, 9.22, 9.55, 7.88, 7.84, 8.50, 11.95, 10.92, 9.78)$.

Računamo (sledeća strana) $t = 1.7536$, $\nu = 12 + 12 - 2 = 22$. Kvantil $t_{1-\alpha/2} = t_{0.975} = 2.074$ čitamo iz tablica. Kako $t = 1.7536 < t_{0.975} = 2.074$, ne odbacujemo H_0.

Račun prethodnog zadatka:

	x_1	$x_1 - \bar{x}_{1,n}$	$(x_1 - \bar{x}_{1,n})^2$	x_2	$x_2 - \bar{x}_{2,n}$	$(x_2 - \bar{x}_{2,n})^2$	
1	9.81	−0.22	0.05	9.85	0.52	0.27	
2	9.83	−0.20	0.04	9.30	−0.03	0.00	
3	10.43	0.40	0.16	9.08	−0.25	0.06	
4	11.13	1.10	1.20	8.07	−1.26	1.58	
5	9.70	−0.33	0.11	9.22	−0.11	0.01	
6	9.59	−0.44	0.20	9.55	0.22	0.05	
7	10.88	0.85	0.72	7.88	−1.45	2.10	
8	10.97	0.94	0.88	7.84	−1.49	2.22	
9	9.35	−0.68	0.47	8.50	−0.83	0.69	
10	9.34	−0.69	0.48	11.95	2.62	6.87	
11	9.41	−0.62	0.39	10.92	1.59	2.53	
12	9.95	−0.08	0.01	9.78	0.45	0.20	
$\sum$	120.39		4.696225	111.94		16.58837	
$\bar{x}_{1,n} =$	10.0325	$\bar{s}^2_{1,n} =$	0.3913521	9.328333	$= \bar{x}_{2,n}$	1.382364	$= \bar{s}^2_{2,n}$

$$t := \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\bar{s}_1^{2\prime}}{n_1} + \frac{\bar{s}_2^{2\prime}}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\bar{s}_1^2}{n_1-1} + \frac{\bar{s}_2^2}{n_2-1}}} = \frac{10.0325 - 9.328333}{\sqrt{\frac{0.3913521}{11} + \frac{1.382364}{11}}} = 1.7536$$

T-test parova

Koristi se kad imamo dva obeležja sa uparenim vrednostima ("pre" i "posle"): $x_1, x_2, \ldots, x_n$ i $y_1, y_2, \ldots, y_n$.

Nalazimo $t_1 = x_1 - y_1, t_2 = x_2 - y_2, \ldots, t_n = x_n - y_n$ i testiramo $H_0(\mu = 0)$ protiv $H_1(\mu \neq 0)$ ili $H_1(\mu < 0)$ ili $H_1(\mu > 0)$, sa σ nepoznato za uzorak $t_1, t_2, \ldots, t_n$

Primer 36 *Testirati hipotezu o jednakosti srednjih vrednosti uzorka* $(9.81, 9.83, 10.43, 11.13, 9.70, 9.59, 10.88, 10.97, 9.35, 9.34, 9.41, 9.95)$ *i uzorka* $(9.85, 9.30, 9.08, 8.07, 9.22, 9.55, 7.88, 7.84, 8.50, 11.95, 10.92, 9.78)$ *ako su vrednosti redom pre i posle tretmana na istom uzorku.*

Rešenje: Koristimo T-test parova. Dobijamo redom vrednosti: $(-0.04, 0.53, 1.35, 3.06, 0.48, 0.04, 3.00, 3.13, 0.85, -2.61, -1.51, 0.17)$, $n = 12$.

Zatim, $\bar{x}_n = 0.7042$, $\bar{s}'_n = 1.7681$, $t = \frac{|\bar{x}_n - m_0|}{\bar{s}'_n}\sqrt{n} = 1.380 \leq 2.201$, ne odbacujemo H_0.

$\bar{x}_{12} = (-0.04 + 0.53 + 1.35 + 3.06 + 0.48 + 0.04 + 3.00 + 3.13 + 0.85 - 2.61 - 1.51 + 0.17)/12 = 0.7042$,

$\bar{s}^2_{12} = (0.04^2 + 0.53^2 + 1.35^2 + 3.06^2 + 0.48^2 + 0.04^2 + 3.00^2 + 3.13^2 + 0.85^2 + 2.61^2 + 1.51^2 + 0.17^2)/12 - 0.7042^2 = 2.8659$,

$t = \frac{|\bar{x}_n - m_0|}{\bar{s}_n}\sqrt{n-1} = \frac{|0.7042 - 0|}{\sqrt{2.8659}}\sqrt{11} = 1.380$

$n \backslash P$	.75	.90	.95	.975	.990	.995	.9995
...							
11	.697	1.363	1.796	2.201	2.718	3.106	4.437
...							
22	.686	1.321	1.717	2.074	2.508	2.819	3.792
...							
z	.674	1.282	1.645	1.960	2.326	2.576	3.291

PRIMER 37 *Na jednom studijskom programu upisalo se 20 kandidata muškog pola i 14 ženskog pola. Muški su na prijemnom osvojili bodove: 48; 60; 35; 37; 43; 53; 47; 49; 51; 53; 35; 44; 54; 54; 60; 53; 60; 45; 46; 48 a ženski: 56; 42; 36; 36; 60; 56; 48; 57; 60; 60; 37; 50; 60; 50.*

Testirati hipotezu da muški i ženski kandidati imaju osvojen isti broj bodova na prijemnom.

1. *Napisati naziv testa koji se koristi.*

2. *Napisati nultu i alternativnu hipotezu.*

3. *Izračunati vrednost statistike i napisati zaključak.*

Test jednakosti proporcija dva uzorka

U tabeli su dati podaci o osnovnim školama u tri opštine u Srbiji 2015/2016.

```
skola,opstina   ,broj_ucenika,broj_nast,ps,boravak,smena,broj_odeljenja,ima_f_salu,broj_knjiga
459  ,Novi Sad  ,973         ,68       ,Ne,Ne     ,2    ,39            ,Da        ,13.99
880  ,Nis       ,1046        ,77       ,Ne,Da     ,2    ,43            ,Da        ,11.2
704  ,Kragujevac,473         ,53       ,Ne,Ne     ,2    ,23            ,Da        ,21.38
466  ,Novi Sad  ,1087        ,76       ,Da,Ne     ,2    ,42            ,Da        ,15.09
...
```

Opis atributa:

Naziv kolone	sadržaj
skola	ID škole
opstina	Opština u kojoj se škola nalazi
broj_ucenika	Ukupan broj učenika u školi
broj_nast	Broj nastavnika zaposlenih u školi
ps	Uzima vrednost da ako je u školi zaposlen pedagoški savetnik, inače uzima vrednost ne
boravak	Uzima vrednost da ako je u školi organizovan produženi boravak, inače uzima vrednost ne
smena	Broj smena u kojima je organizovan rad u školi
broj_odeljenja	Broj odeljenja u školi
ima_f_salu	Uzima vrednost da ako škola poseduje fiskulturnu salu inače uzima vrednost ne
broj_knjiga_po_uc	Broj knjiga u biblioteci po učeniku

Za atribute označiti znakom ✓ da li su odgovarajućeg tipa

	Kval.	Nom.	Ord.	Kvan.	Diskr.	Nepr.	Bin.
1.	skola	✓					
2.	opstina	✓					
3.	broj_ucenika			✓	✓		
4.	broj_nast			✓	✓		
5.	ps	✓					✓
6.	boravak	✓					✓
7.	smena	✓	✓				
8.	broj_odeljenja			✓	✓		
9.	ima_f_salu	✓					✓
10.	broj_knjiga_po_uc			✓		✓	