

STRUČNI STUDIJ INFORMATIKE

VJEROJATNOST I STATISTIKA

Dr. Igor Urbija, prof. vis. šk.

Sadržaj

1 Predgovor	4
I Deskriptivna statistika	5
2 Tabela frekvencija, histogram, kumulativna funkcija	5
3 Aritmetička sredina, mod i medijan	12
4 Kvartil. Percentil. Kvantil	16
5 Varijanca. Standardna devijacija. Čebiševljev teorem	19
6 Uspoređivanje različitih mjerena. Uspoređivanje različitih rezultata	24
7 Linearna regresija	26
II Vjerojatnost	36
8 Događaj. Vjerojatnost	37
9 Diskretna slučajna varijabla. Distribucija slučajne varijable.	65
10 Funkcija vjerojatnosti. Funkcija distribucije vjerojatnosti	70
11 Očekivanje, varijanca i standardna devijacija diskretne slučajne varijable	73
12 Diskretna uniformna distribucija	76
13 Bernoullijev pokus. Bernoullijeva shema. Binomna distribucija	80
14 Poissonova distribucija	86
15 Kontinuirana slučajna varijabla	93
16 Normalna (Gaussova) distribucija. Standardna normalna distribucija	95
17 χ^2 -distribucija	100
III Statistički testovi	101
18 Testiranje hipoteze o očekivanju uz poznatu varijancu	103
19 χ^2 test	109

20 Rješenja zadataka	122
20.1 Događaj. Vjerojatnost	122
20.2 Aritmetička sredina, mod i medijan	123
20.3 Uvjetna vjerojatnost	124
20.4 Nezavisni događaji	125
20.5 Diskretna uniformna distribucija	125
20.6 Bernoullijev pokus. Bernoullijeva shema. Binomna distribucija	127

1 Predgovor

Sadržaj ovog nastavnog materijala odgovara sadržaju kolegija Vjerojatnost i statistika koji se održava na drugoj godini stručnog studija informatike i druge godine stručnog studija računarstva na Tehničkom veleučilištu u Zagrebu. Podijeljen je na tri dijela koji obuhvaćaju sljedeće cjeline: deskriptivna statistika, uvod u vjerojatnost i uvod u statističke testove.

U prvoj cjelini se obrađuje deskriptivna statistika gdje su dani osnovni postupci izračunavanja sumarnih podataka (aritmetička sredina, mod, medijan, varijanca, standardna devijacija, kvartili, percentili) za zadani skup brojčanih podataka, kao i njihov grafički prikaz (histogram, pravac regresije).

U drugoj cjelini je dan uvod u vjerojatnost, uvedene su slučajne varijable i obradene neke diskretne (uniformna, binomna i Poissonova distribucija) i kontinuirane distribucije (normalna i χ^2 distribucija).

U trećoj cjelini su obrađena dva statistička testa (testiranje hipoteze o očekivanju uz poznatu varijancu i χ^2 -test).

Namjena materijala, koji prati predavanja i vježbe, je upoznavanje studenata s osnovama vjerojatnosti i osnovama statistike uz minimum teorije koja je dovoljna za definiranje potrebnih pojmovima i za rješavanje zadataka kao i izvor zadataka te kao takav može biti koristan svim studentima Tehničkog veleučilišta koji u programu imaju isti ili sličan kolegij.

Napomenuo bih da se skoro svi statistički izračuni spomenuti u deskriptivnoj statistici mogu provesti na većini kalkulatora, dok se na boljim kalkulatorima nalaze programi za provođenje statističkih testova ili se oni mogu lako isprogramirati. Za složenije statističke postupke postoje posebni statistički programski paketi namijenjeni osobnim računalima (npr. Statistica), a mogu se koristiti i tablični kalkulatori (npr. Open Office Calc ili Microsoft Office Excel), koji imaju implementiranu većinu potrebnih funkcija.

Dr. Igor Urbija, prof. struč. stud., 2008.

Datum zadnje izmjene: 11. listopada 2024.

Dio I

Deskriptivna statistika

2 Tabela frekvencija, histogram, kumulativna funkcija

Dani su brojevi y_1, y_2, \dots, y_n . Neka se među njima pojavljuje f_1 brojeva x_1 , f_2 brojeva x_2, \dots, f_r brojeva x_r pri čemu su brojevi x_1, x_2, \dots, x_r međusobno različiti.

Brojevi x_1, x_2, \dots, x_r su razredi i obično su sortirani uzlazno (tj. vrijedi $x_1 < x_2 < \dots < x_r$), a brojevi f_1, f_2, \dots, f_r su *frekvencije*.

Prvi korak u sređivanju dаних brojeva je izrada *tabele frekvencija*:

x	x_1	x_2	\dots	x_r
f	f_1	f_2	\dots	f_r

Primjetimo da je suma svih frekvencija jednaka broju dаних podataka n :

$$f_1 + f_2 + \dots + f_r = n$$

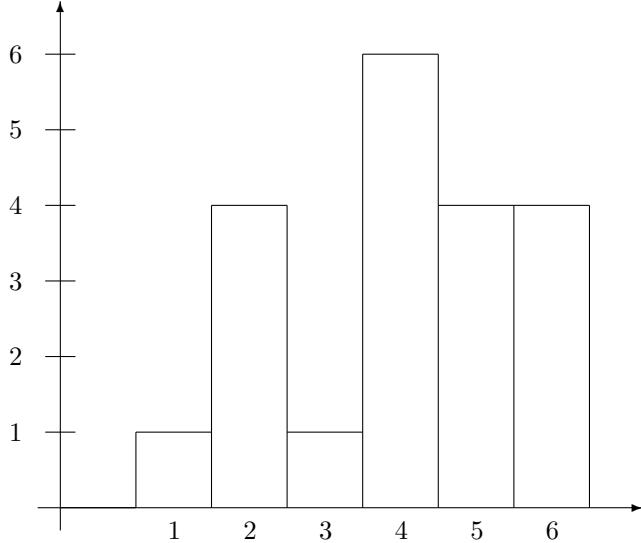
Primjer 1.

Dani su sljedeći brojevi: 6, 4, 4, 1, 6, 2, 6, 4, 5, 6, 4, 2, 3, 5, 5, 2, 2, 4, 4, 5.

Prebrojavanjem ustanovljavamo da su razredi 1, 2, 3, 4, 5, 6 pri čemu se broj 1 pojavljuje jednom,
broj 2 pojavljuje četiri puta,
broj 3 pojavljuje jednom,
broj 4 pojavljuje šest puta,
broj 5 pojavljuje četiri puta i
broj 6 pojavljuje četiri puta pa pripadna tabela frekvencija izgleda ovako:

x	1	2	3	4	5	6
f	1	4	1	6	4	4

Grafički prikaz gornje tabele (u kojoj je frekvencija svakog podatka predstavljena pravokutnikom čija je visina upravo jednaka toj frekvenciji) se zove *histogram*.



Na x -osi se navode sve različite vrijednosti danih podataka (i samo one), a na y -osi njihove frekvencije.

U tabeli frekvencija se obično nalazi još jedan redak s *relativnim frekvencijama* (u tabeli označen s rf), a to su frekvencije podijeljene brojem danih brojeva. Općenito to izgleda ovako:

x	x_1	x_2	\dots	x_r
f	f_1	f_2	\dots	f_r
rf	$\frac{f_1}{n}$	$\frac{f_2}{n}$	\dots	$\frac{f_r}{n}$

U našem primjeru imamo:

x	1	2	3	4	5	6
f	1	4	1	6	4	4
rf	$\frac{1}{20}$	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{6}{20}$	$\frac{4}{20}$	$\frac{4}{20}$

x	1	2	3	4	5	6
f	1	4	1	6	4	4
rf	0.05	0.20	0.05	0.30	0.20	0.20

Tabeli se najčešće dodaje još jedan redak koji opisuje *kumulativnu funkciju* (u tabeli označen s cf), a sastoji se od parcijalnih sumi relativnih frekvencija: Općenito to izgleda ovako:

x	x_1	x_2	\dots	x_r
f	f_1	f_2	\dots	f_r
rf	$\frac{f_1}{n}$	$\frac{f_2}{n}$	\dots	$\frac{f_r}{n}$
cf	$\frac{f_1}{n}$	$\frac{f_1}{n} + \frac{f_2}{n}$	\dots	$\frac{f_1}{n} + \frac{f_2}{n} + \dots + \frac{f_r}{n}$

Primijetimo da zadnji broj u retku kumulativne funkcije mora biti točno¹ 1 jer je to suma svih relativnih frekvencija:

$$\frac{f_1}{n} + \frac{f_2}{n} + \dots + \frac{f_r}{n} = \frac{f_1 + f_2 + \dots + f_r}{n} = \frac{n}{n} = 1.$$

Označimo k -ti broj u retku kumulativne funkcije s φ_k , tj. $\varphi_k = \frac{f_1}{n} + \frac{f_2}{n} + \dots + \frac{f_k}{n}$. Kumulativna funkcija je funkcija $cf : \mathbb{R} \rightarrow \mathbb{R}$ koja je određena vrijednostima u zadnjem retku tabele frekvencija ovako:

$$cf(x) = \begin{cases} 0, & \text{ako je } x < x_1 \\ \varphi_k, & \text{ako je } x_k \leq x < x_{k+1} \\ 1, & \text{ako je } x \geq x_r \end{cases}$$

U našem primjeru imamo:

x	1	2	3	4	5	6
f	1	4	1	6	4	4
rf	0.05	0.20	0.05	0.30	0.20	0.20
cf	0.05	0.25	0.3	0.6	0.8	1

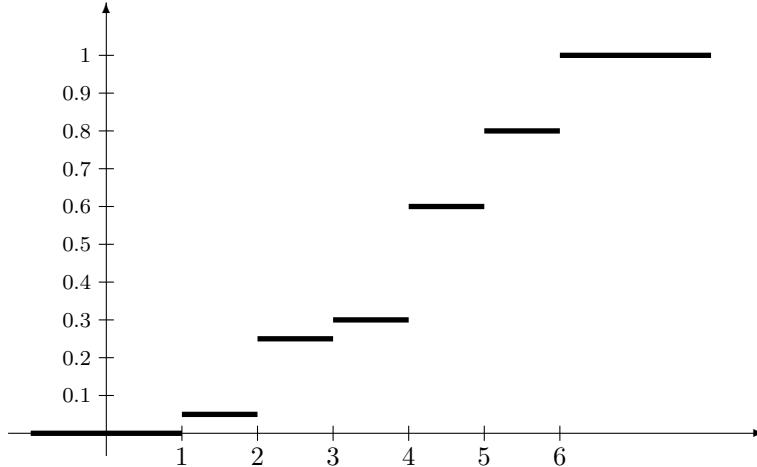
$$cf(-100) = cf(-2) = cf(0) = cf(0.9) = 0, \text{ zbog } -100, -2, 0, 0.9 < x_1 = 1$$

$$cf(1) = cf(1.5) = 0.05, \text{ jer je } 1 = x_1 \leq 1.5 < x_2 = 2$$

$$cf(4) = cf(4.99) = 0.6, \text{ jer je } 4 = x_4 \leq 4, 4.99 < x_5 = 5$$

$$cf(6) = cf(6.5) = cf(10) = cf(100) = 1, \text{ zbog } 6, 6.5, 10, 100 \geq x_6 = 6$$

Graf kumulativne funkcije u našem primjeru izgleda ovako:



¹Prilikom računanja se zbog grešaka zaokruživanja može dogoditi da umjesto 1 dobijemo neki njemu blizak broj poput 0.998 ili 1.0003, u ovisnosti o točnosti kojom provodimo računanje.

Zadaci

Zadatak 1.

Dani su brojevi: 4, 6, 6, 2, 1, 3, 2, 7, 3, 6, 3, 4, 1, 1, 2, 2, 7, 6, 6, 6.

Odredite razrede, sastavite tabelu frekvencija s kumulativnom funkcijom, skicirajte histogram i graf kumulativne funkcije. Izračunajte $cf(-1)$, $cf(2)$, $cf(5)$ i $cf(8)$.

Rješenje.

zadanih brojeva ima: 20

razredi: 1, 2, 3, 4, 6, 7

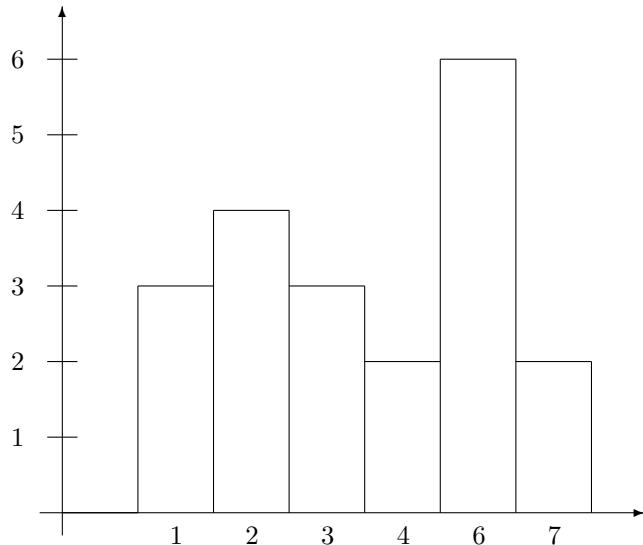
broj razreda: 6

tabela frekvencija i kumulativna funkcija:

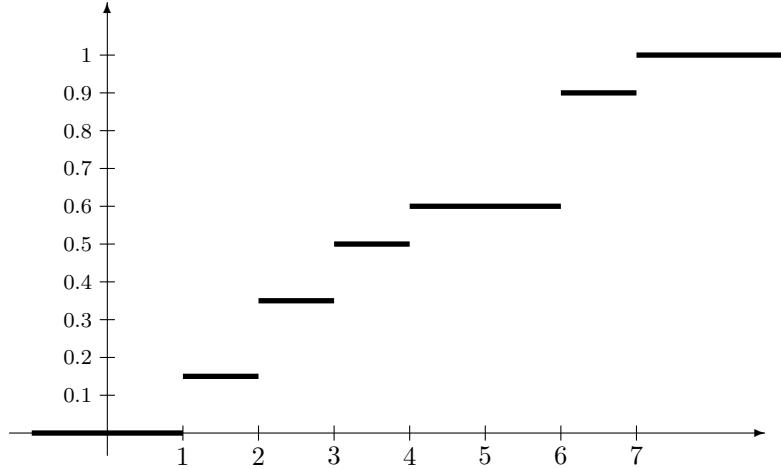
x	1	2	3	4	6	7
f	3	4	3	2	6	2
rf	0.15	0.2	0.15	0.1	0.3	0.1
cf	0.15	0.35	0.5	0.6	0.9	1

$$cf(-1) = 0, cf(2) = 0.35, cf(5) = 0.6, cf(8) = 1$$

histogram:



graf kumulativne funkcije relativnih frekvencija:



Zadatak 2. Dani su brojevi: 2, 2, 2, 4, 4, 5, 6, 5, 2, 6, 3, 1, 5, 5, 5, 2, 3, 4, 1, 3.

Odredite razrede, sastavite tabelu frekvencija s kumulativnom funkcijom, skicirajte histogram i graf kumulativne funkcije. Izračunajte $cf(0)$, $cf(1)$, $cf(4)$ i $cf(7)$.

Zadatak 3. Dani su brojevi: 4, 0, 5, 4, 8, 7, 2, 2, 3, 3, 5, 2, -1, 4, 8, 3, 1, 3, -1, 4, 2, 8, 2, 3, 8, 0, -1, 0, 1, 3, 2, 8, 7, 4, 2, 0, -1, 8, 1, 8.

Odredite razrede, sastavite tabelu frekvencija s kumulativnom funkcijom, skicirajte histogram i graf kumulativne funkcije. Izračunajte $cf(-2)$, $cf(0)$, $cf(4)$ i $cf(10)$.

Uz pomoć tabele frekvencija možemo brzo saznati ukupan postotak zadanih brojeva koji su manji od nekog broja, kao i njihov ukupan broj.

x	x_1	x_2	x_3	\dots	x_{r-1}	x_r
f	f_1	f_2	f_3	\dots	f_{r-1}	f_r
rf	$\frac{f_1}{n}$	$\frac{f_2}{n}$	$\frac{f_3}{n}$	\dots	$\frac{f_{r-1}}{n}$	$\frac{f_r}{n}$
cf	φ_1	φ_2	φ_3	\dots	φ_{r-1}	$\varphi_r (= 1)$

Sada možemo lagano, bez prebrojavanja, dobiti odgovore na sljedeće osnovne tipove pitanja:
Koliko ukupno podatka i koliki postotak podataka ima vrijednost

1. manju od ili jednaku b , $X \leq b$

2. veću od a , $X > a$

pri čemu se složenija pitanja dobivaju kombiniranjem ova dva.

Budući je $\varphi_k = \frac{f_1 + f_2 + \dots + f_k}{n}$, to znači da je $n\varphi_k$ broj podataka koji imaju vrijednost najviše x_k . Dakle odgovor na pitanje "Koliko ukupno podataka ima vrijednost najviše x_k ?" je naprsto² $n \cdot \varphi_k$ ili $n \cdot cf(x_k)$. Ako nas zanima postotak, a ne ukupan broj, onda je to $cf(x_k) \cdot 100\%$.

Primjer 2. Sljedeća tabela prikazuje srednje dnevne temperature ($^{\circ}\text{C}$) tijekom 28 dana.

temp. ($^{\circ}\text{C}$)	-4	-2	1	2	3	5	7	8	9	10	13
f	1	2	2	3	1	1	3	6	2	5	2
rf	0.036	0.071	0.071	0.107	0.036	0.036	0.107	0.214	0.071	0.179	0.071
cf	0.036	0.107	0.179	0.286	0.321	0.357	0.464	0.679	0.75	0.929	1.000

- (a) Koliki postotak promatralih dana i koji je njihov ukupan broj koji su imali srednju dnevnu temperaturu od najviše 5°C ?

Među promatranim danima njih $28 \cdot cf(5) = 28 \cdot 0.357 = 9.996 \approx 10$, odnosno $cf(5) \cdot 100\% = 35.7\%$ je imalo srednju dnevnu temperaturu od najviše 5°C .

Općenito, odgovor na pitanje "Koliko ukupno podataka ima vrijednost najviše b ?" ($X \leq b$) je $n \cdot cf(b)$. Slično, ako nas zanima postotak, onda je to $cf(b) \cdot 100\%$.

- (b) Koliki postotak promatralih dana i koji je njihov ukupan broj koji su imali srednju dnevnu temperaturu od najviše 0°C ?

Među promatranim danima njih $28 \cdot cf(0) = 28 \cdot 0.107 = 2.996 \approx 3$, odnosno $cf(0) \cdot 100\% = 10.7\%$ je imalo srednju dnevnu temperaturu od najviše 0°C .

Ako nas zanima koliko podataka je imalo veću vrijednost od nekog broja a ($X > a$), onda prvo nađemo koliko podataka je imalo vrijednost najviše a i pomoću tog podatka dobijemo traženi (oduznemo ga od broja svih podataka). Dakle odgovor na pitanje "Koliko ukupno podataka ima vrijednost veću od a ?" je $n - n \cdot cf(a) = n(1 - cf(a))$. Ako nas zanima postotak, a ne ukupan broj, onda je to $(1 - cf(a)) \cdot 100\%$.

- (c) Koliki postotak promatralih dana i koji je njihov ukupan broj koji su imali srednju dnevnu temperaturu veću od 6°C ?

Među promatranim danima njih $28(1 - cf(6)) = 28 \cdot (1 - 0.357) = 28 \cdot 0.643 = 18.004 \approx 18$, odnosno $(1 - cf(6)) \cdot 100\% = 64.3\%$ je imalo srednju dnevnu temperaturu veću od 6°C .

Ako nas zanima odgovor na pitanje "Koliko ukupno podataka ima veću vrijednost od a i manju od ili jednaku b ?" ($a < X \leq b$), onda je odgovor $n(cf(b) - cf(a))$. Pripadni postotak je $(cf(b) - cf(a)) \cdot 100\%$.

- (d) Koliki postotak promatralih dana i koji je njihov ukupan broj koji su imali srednju dnevnu temperaturu veću od 1°C , ali ne veću od 11°C ?

Takvih dana ima 21, jer je $28(cf(11) - cf(1)) = 28(0.929 - 0.179) = 28 \cdot 0.75 = 21$, što čini 75% svih promatralih dana.

²Zbog grešaka zaokruživanja često nećemo dobiti cijeli broj, no on će se jako malo razlikovati od prave vrijednosti koju onda neće biti problem odrediti.

- (e) Koliki postotak promatranih dana i koji je njihov ukupan broj koji su imali srednju dnevnu temperaturu jednaku 7°C ?

Ovo je specijalan slučaj prethodnog, što znači da nas zanima broj dana (i njihov postotak) koji su imali temperaturu veću od 5°C (to je razred koji prethodi zadanim) i ne veću od 7°C . Takvih dana ima tri, jer je $28(cf(7) - cf(5)) = 28(0.464 - 0.357) = 28 \cdot 0.107 = 2.996$, što čini 10.7% svih promatranih dana.

Naravno, traženi podatak se može direktno očitati iz retka s relativnim frekvencijama (ako ga imamo na raspolaganju).

Malo treba pripaziti u slučaju kada je riječ o najmanjem razredu (pa od njega ne postoji manji), u našem primjeru je to -4°C . Da bismo izračunali broj dana i postotak dana koji su imali srednju dnevnu temperaturu -4°C , primijetimo da je to isto kao da tražimo postotak dana koji su imali srednju temperaturu od najviše -4°C , čime smo problem sveli na slučaj (b), dakle broj takvih dana je $28cf(-4) = 28 \cdot 0.036 = 1.008 \approx 1$, a postotak je $cf(-4) \cdot 100\% = 3.6\%$.

upit	ukupan broj	postotak
$X \leq b$	$n cf(b)$	$100 cf(b)$
$X > a$	$n(1 - cf(a))$	$100(1 - cf(a))$
$a < X \leq b$	$n(cf(b) - cf(a))$	$100(cf(b) - cf(a))$

Zadaci

Zadatak 4. Zadana je sljedeća tabela frekvencija:

x	-3	-2	-1	0	1	2	3
f	5	1	3	3	3	2	3
rf	0.25	0.05	0.15	0.15	0.15	0.1	0.15
cf	0.25	0.3	0.45	0.6	0.75	0.85	1

Koristeći se vrijednostima kumulativne funkcije relativnih frekvencija izračunajte koliko ukupno podatka i koliko posto podataka ima vrijednost

- (a) veću od -2 , $X > -2$
- (b) manju od ili jednaku 2 , $X \leq 2$
- (c) veću od -3 i manju od ili jednaku 1 , $-3 < X \leq 1$
- (d) jednaku 1 , $X = 1$

Zadatak 5. Zadana je sljedeća tabela frekvencija:

x	-2	-1	1	2	3	4	5
f	2	4	2	2	5	4	1
rf	0.1	0.2	0.1	0.1	0.25	0.2	0.05
cf	0.1	0.3	0.4	0.5	0.75	0.95	1

Koristeći se vrijednostima kumulativne funkcije relativnih frekvencija izračunajte koliko ukupno podataka i koliko posto podataka ima vrijednost

- (a) veću od -2 , $X > -2$
- (b) manju od ili jednaku 3 , $X \leq 3$
- (c) veću od -1 i manju od ili jednaku 5 , $-1 < X \leq 5$
- (d) jednaku 0 , $X = 0$

3 Aritmetička sredina, mod i medijan

Umjesto mnogo brojeva x_1, x_2, \dots, x_n (n može biti jako velik) mi bismo željeli izračunati nekoliko brojeva, što manje to bolje, koji bi nam dali neke korisne informacije o njima. Neki od takvih brojeva su aritmetička sredina, mod i medijan.

Aritmetička sredina

Neka je dano n brojeva x_1, x_2, \dots, x_n . Njihova *aritmetička sredina*, u oznaci \bar{x} , se definira formulom:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

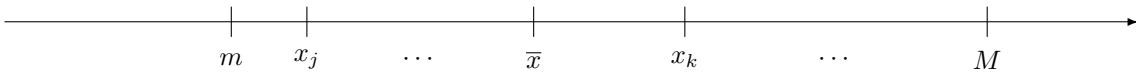
Aritmetička sredina je veća od ili jednaka najmanjem od danih brojeva. Neka je $m = \min\{x_1, x_2, \dots, x_n\}$.

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \geq \frac{m + m + \dots + m}{n} = \frac{n \cdot m}{n} = m = \min\{x_1, x_2, \dots, x_n\}.$$

Aritmetička sredina je manja od ili jednaka najvećem od danih brojeva. Neka je $M = \max\{x_1, x_2, \dots, x_n\}$.

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \leq \frac{M + M + \dots + M}{n} = \frac{n \cdot M}{n} = M = \max\{x_1, x_2, \dots, x_n\}.$$

Dakle vrijedi $\min\{x_1, x_2, \dots, x_n\} \leq \bar{x} \leq \max\{x_1, x_2, \dots, x_n\}$. Na brojevnom pravcu bi pripadajuća slika izgledala ovako:



Primjer 1. Aritmetička sredina brojeva 1, 1, 1, 1, 1, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6 je jednaka:

$$\bar{x} = \frac{1 + 1 + 1 + 1 + 1 + 2 + 2 + 3 + 3 + 3 + 4 + 4 + 4 + 5 + 5 + 5 + 5 + 6 + 6 + 6}{20} = \frac{68}{20} = 3.4$$

Ako su poznate frekvencije danih brojeva,

x	x_1	x_2	\dots	x_n
f	f_1	f_2	\dots	f_n

onda se aritmetička sredina može računati formulom:

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_n x_n}{f_1 + f_2 + \dots + f_n}.$$

Primjer 2. Aritmetičku sredinu brojeva danim u tabeli frekvencija

x	1	2	3	4	5	6
f	5	2	3	3	4	3

računamo ovako:

$$\bar{x} = \frac{5 \cdot 1 + 2 \cdot 2 + 3 \cdot 3 + 3 \cdot 4 + 4 \cdot 5 + 3 \cdot 6}{5 + 2 + 3 + 3 + 4 + 3} = \frac{68}{20} = 3.4.$$

Mod

Mod (ili *dominantna vrijednost*) je onaj podatak u danom nizu brojeva x_1, x_2, \dots, x_n koji se najčešće pojavljuje i to barem dvaput. Niz brojeva može imati jedan mod, više modova ili niti jedan mod.

Niz $2, 3, 3, 5, 7, 7, 7, 8, 9$ ima jedinstveni mod 7.

Niz $2, 3, 4, 5, 7, 8, 9, 11, 13$ nema moda.

Niz $2, 3, 3, 3, 5, 7, 7, 7, 8$ ima dva moda: 3 i 7.

U histogramu je mod određen najvišim pravokutnikom jer visina pravokutnika je proporcionalna s brojem pojavljivanja pripadajućeg podatka.

Primjetimo da je mod podatak koji nam daje više informacija o zadanim podacima od aritmetičke sredine. Dok aritmetička sredina intuitivno predstavlja broj "oko kojeg se gomila najviše podataka" (što čak ne mora biti točno), za mod znamo da je to broj koji se najviše puta pojavljuje među podacima.

Pogledajmo primjer s dva niza brojeva koji imaju istu aritmetičku sredinu i nađimo njihove pripadne mode.

Primjer 1. Zadani su brojevi $x_1 = 16000, x_2 = 4000, x_3 = 4000, x_4 = 4000$. Njihova aritmetička sredina je 7000, a mod je 4000.

Primjer 2. Zadani su brojevi $x_1 = 10000, x_2 = 6000, x_3 = 6000, x_4 = 6000$. Njihova aritmetička sredina je također 7000, a mod je 6000.

Medijan

Neka su dani podaci x_1, x_2, \dots, x_n sortirani uzlazno. *Medijan* (ili *centralna vrijednost*), u oznaci \tilde{x} ("iks tilda") se ugrubo definira kao "srednji broj" u nizu. Ako podataka ima neparan broj, tj. $n = 2k + 1$, onda je $\tilde{x} = x_{k+1}$, a to je upravo srednji broj u nizu. Ako podataka ima paran broj, $n = 2k$, onda nema srednjeg broja pa se \tilde{x} definira kao srednja vrijednost srednjih dva broja u nizu, tj. $\tilde{x} = \frac{x_k + x_{k+1}}{2}$.

$$\tilde{x} = \begin{cases} \text{srednji broj} & , n \text{ neparan} \\ \frac{\text{zbroj srednjih dva broja}}{2} & , n \text{ paran} \end{cases} = \begin{cases} x_{k+1} & , n = 2k + 1 \\ \frac{x_k + x_{k+1}}{2} & , n = 2k \end{cases}$$

$11, 11, \boxed{16}, 17, 25 \quad \tilde{x} = x_3 = 16$

$1, 4, 8, \boxed{8,10}, 16, 16, 19 \quad \tilde{x} = \frac{x_4 + x_5}{2} = \frac{8+10}{2} = 9$

Ako brojeva ima puno, onda je traženje medijana traženjem srednjeg broja ili srednjih brojeva u nizu nespretno, no ako imamo tablicu frekvencija s kumulativnom funkcijom na raspolaganju, onda ga možemo naći jako brzo. Moguća su dva slučaja.

- U retku s kumulativnom funkcijom postoji vrijednost koja je točno jednaka 0.5, kao u sljedećoj tablici:

x	1	2	$\boxed{3}$	$\boxed{4}$	5	6
f	5	2	3	3	4	3
rf	0.25	0.1	0.15	0.15	0.2	0.15
cf	0.25	0.35	$\boxed{0.5}$	0.65	0.85	1

Medijan je u ovom slučaju jednak srednjoj vrijednosti razreda kojoj pripada vrijednost kumulativne funkcije 0.5 i sljedećeg po redu razreda, u ovom slučaju $\tilde{x} = \frac{3+4}{2} = 3.5$.

- U retku s kumulativnom funkcijom nema vrijednosti 0.5, kao u sljedećoj tablici:

x	1	2	3	4	5	6
f	6	2	4	2	3	3
rf	0.3	0.1	0.2	0.1	0.15	0.15
cf	0.3	0.4	0.6	0.7	0.85	1

Medijan je u ovom slučaju jednak onom razredu kojem pripada najmanja vrijednost kumulativne funkcije koja je veća od 0.5, u ovom slučaju $\tilde{x} = 3$.

Zadaci

Zadatak 1.

Dani su brojevi: 5, 1, 5, 3, 2, 3, 6, 5, 6, 1, 2, 5, 1, 6, 3, 1, 3, 4, 1, 6. Odredite razrede, sastavite tabelu frekvencija s kumulativnom funkcijom te pronađite mod(ove) i medijan.

Rješenje.

Zadanih brojeva ima 20, razredi su 1, 2, 3, 4, 5, 6, ima ukupno 6 razreda.
tabela frekvencija i kumulativna funkcija:

x	1	2	3	4	5	6
f	5	2	4	1	4	4
rf	0.25	0.1	0.2	0.05	0.2	0.2
cf	0.25	0.35	0.55	0.6	0.8	1

Među danim brojevima se broj 1 najviše puta pojavljuje (pripadajuća vrijednost u retku s frekvencijama je najveća) pa je on jedinstveni mod.

Broja 0.5 nema među vrijednostima kumulativne funkcije pa tražimo najmanju od 0.5 veću vrijednost. To je 0.55 koja pripada razredu 3 pa je medijan: $\tilde{x} = 3$.

Zadatak 2.

Dani su brojevi: 3, 4, 1, 5, 6, 2, 4, 5, 3, 5, 1, 1, 2, 1, 4, 6, 1, 5, 4, 6. Odredite razrede, sastavite tabelu frekvencija s kumulativnom funkcijom te pronađite mod(ove) i medijan.

Zadatak 3.

Dani su brojevi: 56, 65, 66, 60, 61, 66, 61, 63, 66, 61, 63, 57, 57, 64, 58, 57, 57, 65, 58, 67, 66, 58, 59, 57, 67, 57, 62, 66, 56, 62, 61. Odredite razrede, sastavite tabelu frekvencija s kumulativnom funkcijom te pronađite mod(ove) i medijan.

Zadatak 4.

Kad dodamo broj 15 jednom popisu brojeva, aritmetička sredina poraste za 2. Kada dodamo broj 1 ovako povećanom popisu brojeva, aritmetička sredina se smanji za 1. Koliko je brojeva bilo u početnom popisu? (Rješenje je na stranici 123.)

Zadatak 5.

Aritmetička sredina pet nenegativnih cijelih brojeva je 15. Njihov medijan je 18. Koja je najveća moguća vrijednost najvećeg od tih brojeva? (Rješenje je na stranici 123.)

Zadatak 6.

Helena je pet dana svaki dan trčala jedan sat i petnaest minuta i tri dana svaki dan jedan sat i trideset minuta. Koliko vremena bi trebala trčati deveti dan da bi joj prosjek trčanja po danu iznosio 85 minuta? (Rješenje je na stranici 123.)

4 Kvartil. Percentil. Kvantil

Medijan se može opisati kao "srednji broj" u nizu, što znači da je to broj sa svojstvom da je pola zadanih brojeva, ili 50%, od njega manje. U nekim slučajevima mogu biti zanimljivi brojevi sa svojstvom da je od njih manje 25% ili 75% posto danih brojeva. Ta dva broja, u oznakama Q_1 i Q_3 redom, zajedno s medijanom (kojeg u ovom kontekstu označavamo Q_2) se zovu *kvartili*, pri čemu je Q_1 prvi, Q_2 drugi, a Q_3 treći kvartil.

Broj koji ima svojstvo da je neki proizvoljan postotak r zadanih brojeva od njega manje označavamo P_r – takve brojeve zovemo *percentilima*. Vrijedi: $Q_1 = P_{25}$, ("dvadesetpeti percentil") $\tilde{x} = Q_2 = P_{50}$ ("pedeseti percentil") i $Q_3 = P_{75}$ ("sedamdesetpeti percentil").

Kvantil q_s se definira preko percentila izrazom: $q_s = P_{100s}$, dakle imamo npr. $q_{0.5} = P_{50}$, $q_{0.23} = P_{23}$.

1. kvartil: $Q_1 = P_{25} = q_{0.25}$ medijan prve polovice podataka	2. kvartil: $Q_2 = P_{50} = q_{0.5} = \tilde{x}$ medijan	3. kvartil: $Q_3 = P_{75} = q_{0.75}$ medijan druge polovice podataka
---	---	--

Da bismo mogli brzo računati percentile, potrebno je prvo sastaviti tabelu frekvencija.

x	x_1	x_2	x_3	\dots	x_{m-1}	x_m
f	f_1	f_2	f_3	\dots	f_{m-1}	f_m
rf	$\frac{f_1}{n}$	$\frac{f_2}{n}$	$\frac{f_3}{n}$	\dots	$\frac{f_{m-1}}{n}$	$\frac{f_m}{n}$
cf	$\frac{f_1}{n}$	$\frac{f_1+f_2}{n}$	$\frac{f_1+f_2+f_3}{n}$	\dots	$\frac{f_1+f_2+\dots+f_{m-1}}{n}$	$\frac{f_1+f_2+\dots+f_{m-1}+f_m}{n} = 1$

Postupak nalaženja proizvoljnog percentila (ili kvantila) je u biti jednak postupku nalaženju medijana (opisanog na stranici 13), tj. pedesetog percentilea. Trebamo naći percentil P_r (tj. broj sa svojstvom da $r\%$ svih podataka nisu od njega veći). Potražimo gdje se u retku kumulativne funkcije nalazi broj $r/100$ i neka radi jednostavnosti redak kumulativne funkcije u tabeli frekvencija izgleda ovako:

$$\text{cf } | \varphi_1 (= \frac{f_1}{n}) \quad \varphi_2 \quad \varphi_3 \quad \dots \quad \varphi_{m-1} \quad \varphi_m (= 1)$$

Dakle, broj φ_k se u tabeli nalazi u retku kumulativne funkcije ispod razreda x_k iz prvog retka. Razlikujemo dva slučaja.

- Broj $r/100$ je jedan od brojeva u retku kumulativne funkcije, tj. $r/100 = \varphi_k$. Tada je

$$P_r = \frac{x_k + x_{k+1}}{2}.$$

- Broj $r/100$ se nalazi između dvije susjedne vrijednosti φ_k i φ_{k+1} u retku kumulativne funkcije, tj. $\varphi_k < r/100 < \varphi_{k+1}$. Tada je

$$P_r = x_{k+1}.$$

Primjer 1.

Potražimo 35. percentil P_{35} (tj. kvantil $q_{0.35}$) uz pomoć sljedeće tabele.

x	1	2	3	4	5	6
f	5	2	3	3	4	3
rf	0.25	0.1	0.15	0.15	0.2	0.15
cf	0.25	0.35	0.5	0.65	0.85	1

Broj $\frac{35}{100} = 0.35$ se nalazi u retku kumulativne funkcije pa je traženi percentil P_{35} u ovom slučaju jednak srednjoj vrijednosti razreda kojоj pripada vrijednost kumulativne funkcije 0.35 i sljedećeg po redu razreda, u ovom slučaju $P_{35} = q_{0.35} = \frac{2+3}{2} = 2.5$.

Primjer 2.

Potražimo 75. percentil P_{75} (tj. kvantil $q_{0.75}$) uz pomoć sljedeće tabele.

x	1	2	3	4	5	6
f	6	2	4	2	3	3
rf	0.3	0.1	0.2	0.1	0.15	0.15
cf	0.3	0.4	0.6	0.7	0.85	1

Broj $\frac{75}{100} = 0.75$ se ne nalazi u retku kumulativne funkcije pa je traženi percentil P_{75} jednak onom razredu kojem pripada najmanja vrijednost kumulativne funkcije koja je veća od 0.75 (a to je 0.85), u ovom slučaju $P_{75} = q_{0.75} = 5$.

Iz definicije slijedi $P_0 = q_0 = x_1$, a to je najmanji od zadanih brojeva. Stoti percentil $P_{100} = q_1$ nije definiran, no možemo posebno definirati da je on jednak najvećem od zadanih brojeva (x_m) uvećan za 1, tj. $P_{100} = q_1 = x_m + 1$.

Primjer 3.

Sljedeća tabela prikazuje srednje dnevne temperature ($^{\circ}\text{C}$) tijekom 28 dana.

temp. ($^{\circ}\text{C}$)	-4	-2	1	2	3	5	7	8	9	10	13
f	1	2	2	3	1	1	3	6	2	5	2
rf	0.036	0.071	0.071	0.107	0.036	0.036	0.107	0.214	0.071	0.179	0.071
cf	0.036	0.107	0.179	0.286	0.321	0.357	0.464	0.679	0.75	0.929	1.000

Pronadimo sve kvartile i percentil P_{10} .

Prvi kvartil Q_1 je jednak dvadesetpetom percentilu. Broja $\frac{25}{100} = 0.25$ nema u retku kumulativne funkcije pa tražimo prvi od njega veći, a to je 0.286. Toj vrijednosti kumulativne funkcije odgovara razred 2 i to je tražena vrijednost, tj. imamo $Q_1 = P_{25} = 2$.

Slično dobivamo da je medijan, odnosno drugi kvartil jednak $P_{50} = 8$.

Broj $\frac{75}{100} = 0.75$ se nalazi u retku kumulativne funkcije pa je tražena vrijednost trećeg kvartila jednaka aritmetičkoj sredini pripadajućeg i sljedećeg po vrijednosti razreda: $Q_3 = P_{75} = \frac{9+10}{2} = 9.5$.

Broja $\frac{10}{100} = 0.1$ nema u retku kumulativne funkcije pa tražimo prvi od njega veći, a to je 0.107. Toj vrijednosti kumulativne funkcije odgovara razred -2 i to je tražena vrijednost, tj. imamo $P_{10} = -2$.

Zadaci

Zadatak 1.

Zadana je sljedeća tabela frekvencija:

x	-3	-2	0	1	2	3	4
f	1	2	5	3	2	3	4
rf	0.05	0.1	0.25	0.15	0.1	0.15	0.2
cf	0.05	0.15	0.4	0.55	0.65	0.8	1

Odredite

1. medijan \tilde{x}
2. prvi kvartil Q_1
3. percentil P_{75}
4. kvantil $q_{0.15}$

Zadatak 2.

Zadana je sljedeća tabela frekvencija:

x	-1	0	1	2	3	4	6
f	1	3	3	4	3	2	4
rf	0.05	0.15	0.15	0.2	0.15	0.1	0.2
cf	0.05	0.2	0.35	0.55	0.7	0.8	1

Odredite

1. medijan \tilde{x}
2. treći kvartil Q_3
3. percentil P_{35}
4. kvantil $q_{0.75}$

5 Varijanca. Standardna devijacija. Čebiševljev teorem

Jedan od brojeva koji nam može dati neku informaciju o zadanim brojevima x_1, x_2, \dots, x_n je, kako smo vidjeli, njihova aritmetička sredina \bar{x} :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

Ono što ne znamo je koliko zadani brojevi *odstupaju* od \bar{x} . Kvadratno odstupanje v_j podatka x_j od aritmetičke sredine \bar{x} se definira sa $v_j = (x_j - \bar{x})^2$. Sada opet dobivamo mnogo brojeva i zato računamo aritmetičku sredinu svih kvadratnih odstupanja koju označavamo sa σ^2 i zovemo *varijanca* (ili *srednje kvadratno odstupanje*):

$$\sigma^2 = \frac{v_1 + v_2 + \dots + v_n}{n} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}.$$

Taj broj predstavlja srednje kvadratno odstupanje (ili *raspršenost*) zadanih brojeva oko njihove aritmetičke sredine.

Standardna devijacija σ je jednaka drugom korijenu varijance:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}.$$

Primjer 1.

Zadani su brojevi $x_1 = 16, x_2 = 4, x_3 = 4, x_4 = 4$. Njihova aritmetička sredina je

$$\bar{x} = \frac{16 + 4 + 4 + 4}{4} = 7,$$

pripadna kvadratna odstupanja su redom $v_1 = (x_1 - \bar{x})^2 = (16 - 7)^2 = 81, v_2 = (x_2 - \bar{x})^2 = (4 - 7)^2 = 9 = v_3 = v_4$ i varijanca je

$$\sigma^2 = \frac{(16 - 7)^2 + (4 - 7)^2 + (4 - 7)^2 + (4 - 7)^2}{4} = \frac{81 + 9 + 9 + 9}{4} = 27.$$

Standardna devijacija je jednaka: $\sigma = \sqrt{27} \approx 5.20$.

Primjer 2.

Zadani su brojevi $x_1 = 10, x_2 = 6, x_3 = 6, x_4 = 6$. Njihova aritmetička sredina je

$$\bar{x} = \frac{10 + 6 + 6 + 6}{4} = 7,$$

pripadna kvadratna odstupanja su redom $v_1 = (x_1 - \bar{x})^2 = (10 - 7)^2 = 9, v_2 = (x_2 - \bar{x})^2 = (6 - 7)^2 = 1 = v_3 = v_4$ i varijanca je

$$\sigma^2 = \frac{(10 - 7)^2 + (6 - 7)^2 + (6 - 7)^2 + (6 - 7)^2}{4} = \frac{9 + 1 + 1 + 1}{4} = 3.$$

Standardna devijacija je jednaka: $\sigma = \sqrt{3} \approx 1.73$.

U oba primjera smo dobili iste aritmetičke sredine (dakle samo pomoću njih ne možemo uočiti po čemu se ta dva niza zadanih brojeva razlikuju), ali u prvom su zadani brojevi udaljeniji od njihove aritmetičke sredine (raspršeniji su) od onih u drugom, što je vidljivo iz varijance – veća varijanca znači da je raspršenost veća. Pogledajmo kako možemo dobiti jednostavniju formulu za računanje varijance.

$$\begin{aligned}
\sigma^2 &= \frac{1}{n}((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2) \\
&= \frac{1}{n}(x_1^2 - 2x_1\bar{x} + \bar{x}^2 + x_2^2 - 2x_2\bar{x} + \bar{x}^2 + \cdots + x_n^2 - 2x_n\bar{x} + \bar{x}^2) \\
&= \frac{1}{n}(x_1^2 + x_2^2 + \cdots + x_n^2 - 2x_1\bar{x} - 2x_2\bar{x} - \cdots - 2x_n\bar{x} + \underbrace{\bar{x}^2 + \bar{x}^2 + \cdots + \bar{x}^2}_n) \\
&= \frac{1}{n}(x_1^2 + x_2^2 + \cdots + x_n^2) - \frac{2\bar{x}}{n}(x_1 + x_2 + \cdots + x_n) + \frac{1}{n}n\bar{x}^2 \\
&= \frac{1}{n}(x_1^2 + x_2^2 + \cdots + x_n^2) - 2\bar{x}\bar{x} + \bar{x}^2 \\
&= \bar{x}^2 - \bar{x}^2,
\end{aligned}$$

gdje je $\bar{x}^2 = \frac{1}{n}(x_1^2 + x_2^2 + \cdots + x_n^2)$. Jednostavnije formule za računanje varijance i standardne devijacije glase:

$$\sigma^2 = \bar{x}^2 - \bar{x}^2, \quad \sigma = \sqrt{\bar{x}^2 - \bar{x}^2}. \quad (1)$$

Primjer 3.

Nadite standardnu devijaciju brojeva 1, 2, 3, 6, 7 pomoću formule (1).

$$\begin{aligned}
\bar{x} &= \frac{1+2+3+6+7}{5} = 3.8 \\
\bar{x}^2 &= \frac{1^2+2^2+3^2+6^2+7^2}{5} = \frac{1+4+9+36+49}{5} = 19.8 \\
\sigma^2 &= \bar{x}^2 - \bar{x}^2 = 19.8 - 14.44 = 5.36 \\
\sigma &= \sqrt{\sigma^2} = \sqrt{5.36} \approx 2.32
\end{aligned}$$

Primjer 4.

Zadani su brojevi 2.3, 1.0, 0.2, 3.2, 2.5, 2.4, 1.3, 0.6, 0.5, 3.0, 1.6, 2.0, 1.4, 0.9, 2.2, 1.2, 0.7, 1.8, 0.6, 2.0, 2.0, 2.7, 2.0, 2.5, 1.0, 1.8, 2.0, 1.2, 1.3. Predstavite zadane podatke tabelom frekvencija, napravite histogram za grupiranje podataka u šest grupa koje odgovaraju podjeli intervala koji sadrži podatke u šest jednakih dijelova, izračunajte aritmetičku sredinu, varijancu i standardnu devijaciju zadanih podataka.

Najmanji broj je 0.2, a najveći 3.2, dakle svi podaci se nalaze u intervalu [0.2, 3.2]. Duljina tog intervala je $3.2 - 0.2 = 3$ pa je duljina svakog od šest jednakih dijelova tog intervala $\frac{3}{6} = 0.5$. Tih šest jednakih dijelova su intervali [0.2, 0.7], [0.7, 1.2], [1.2, 1.7], [1.7, 2.2], [2.2, 2.7], [2.7, 3.2].

x	0.2	0.5	0.6	0.7	0.9	1.0	1.2	1.3	1.4	1.6	1.8	2.0	2.2	2.3	2.4	2.5	2.7	3.0	3.2
f	1	1	2	1	1	2	2	2	1	1	2	6	1	1	1	2	1	1	1

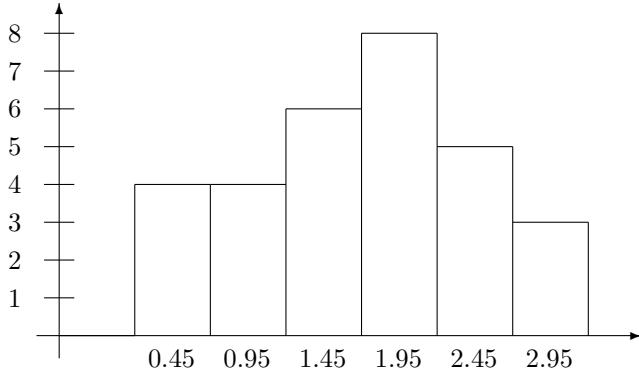
$$\bar{x} \approx 1.663, \quad \sigma^2 \approx 0.583, \quad \sigma \approx 0.764$$

U slučajevima kada ima puno različitih podataka s malim frekvencijama, nerijetko se podaci grupiraju po intervalima. Ovdje ćemo cijeli interval u kojem se nalaze podaci ([0.2, 3.2]) podijeliti na 6 jednakih

podintervala. Za aproksimaciju vrijednosti koje pripadaju nekom podintervalu čemo uzeti polovište podintervala, a za frekvencije čemo uzeti broj podataka koji se nalaze u pojedinom podintervalu. Tako dobivamo manju i čitljiviju tabelu, a aritmetička sredina, varijanca i standardna devijacija ovako grupiranih podataka se obično ne razlikuju mnogo od pravih vrijednosti.

x	0.45	0.95	1.45	1.95	2.45	2.95
f	4	4	6	8	5	3

$$\bar{x} \approx 1.7, \quad \sigma^2 \approx 0.5625, \quad \sigma \approx 0.75$$



Ako o nekim brojčanim podacima ne znamo ništa osim njihove aritmetičke sredine i standardne devijacije, onda na osnovu ta dva podatka ipak možemo još nešto saznati pomoću *Čebiševljevog teorema*.

Čebiševljev teorem.

Neka su \bar{x} aritmetička sredina i $\sigma > 0$ standardna devijacija nekog niza brojeva x_1, x_2, \dots, x_n . Neka je $r \leq n$ broj brojeva iz zadanog niza x_1, x_2, \dots, x_n tako da je $|x_i - \bar{x}| > k\sigma$ za neki broj k . Tada vrijedi:

$$r \leq \frac{n}{k^2}. \quad (2)$$

Dokaz: Neka je y_1, y_2, \dots, y_n niz nenegativnih realnih brojeva, \bar{y} njihova aritmetička sredina i $a \in \mathbb{R}$ neki realan broj. Tada vrijedi:

$$\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j = \frac{1}{n} \sum_{y_j \leq a} y_j + \frac{1}{n} \sum_{y_j > a} y_j \geq \frac{1}{n} \sum_{y_j > a} y_j \geq \frac{1}{n} \sum_{y_j > a} a = \frac{a}{n} \sum_{y_j > a} 1 = \frac{a}{n} r,$$

gdje je $r = \sum_{y_j > a} 1$ = broj brojeva y_j koji su veći od a . Dobivamo:

$$r \leq \frac{n\bar{y}}{a}. \quad (3)$$

Neka je sada x_1, x_2, \dots, x_n niz proizvoljnih realnih brojeva, \bar{x} njihova aritmetička sredina i σ^2 njihova varijanca. Definirajmo $y_j = (x_j - \bar{x})^2$, $j = 1, \dots, n$. Nadalje, neka je $a = k^2\sigma^2$, gdje je $k \in \mathbb{R}^+$ neki pozitivan broj. Sada imamo:

$$\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2 = \sigma^2.$$

Primjetimo da vrijedi $y_j \geq 0, j = 1, \dots, n$ pa možemo primijeniti nejednakost (3):

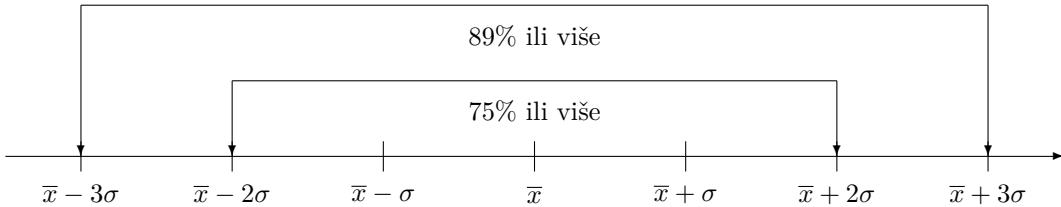
$$r \leq \frac{n\bar{y}}{a} = \frac{n\sigma^2}{k^2\sigma^2} = \frac{n}{k^2},$$

gdje je r broj brojeva $y_j = (x_j - \bar{x})^2$ koji su veći od $a = k^2\sigma^2$, tj. broj brojeva x_j za koje vrijedi $|x_j - \bar{x}| > k\sigma$. \square

Ako je $k = 3$ onda imamo $r \leq \frac{n}{9}$, tj. najviše jedna devetina (oko 11%) od svih n podataka se nalazi izvan intervala $[\bar{x} - 3\sigma, \bar{x} + 3\sigma]$, tj. barem se 89% od svih n podataka nalazu *unutar* tog intervala.

Ako je $k = 2$ onda imamo $r \leq \frac{n}{4}$, tj. najviše jedna četvrtina (25%) od svih n podataka se nalazi izvan intervala $[\bar{x} - 2\sigma, \bar{x} + 2\sigma]$, tj. barem se 75% od svih n podataka nalazu *unutar* tog intervala.

Ako je $k = 1$ onda imamo $r \leq n$, tj. moguće je sve – da svi brojevi budu izvan intervala ($l = n$) ili da svi brojevi budu unutar njega ($l = 0$).



Dakle aritmetička sredina \bar{x} i standardna devijacija σ *zajedno* nam mogu dati dobar uvid u raspršenost podataka oko njihove aritmetičke sredine (ako je $\sigma > 1$).

Koliko podataka se, u standardnim jedinicama, nalazi u intervalima $[-1, 1]$, $[-2, 2]$ i $[-3, 3]$?

$$z = \frac{x - \bar{x}}{\sigma} \quad \Rightarrow \quad x = \bar{x} + z \cdot \sigma$$

$[-1, 1]_{s.j.} = [\bar{x} - \sigma, \bar{x} + \sigma]$	ne može se ništa konkretno reći
$[-2, 2]_{s.j.} = [\bar{x} - 2\sigma, \bar{x} + 2\sigma]$	barem 75% brojeva je unutar intervala
$[-3, 3]_{s.j.} = [\bar{x} - 3\sigma, \bar{x} + 3\sigma]$	barem 89% brojeva je unutar intervala

Primjer 5.

Koji je najmanji interval oko aritmetičke sredine $\bar{x} = 10$ za koji sa sigurnošću možemo tvrditi da sadrži barem 50% zadanih podataka, ako je $\sigma = 1.75$?

Rješenje.

Prema nejednakosti (2) iz Čebiševljevog teorema i prema zahtjevu zadatka mora vrijediti

$$r \leq n/k^2 = n/2,$$

dakle $k^2 = 2$, tj. $k = \sqrt{2} \approx 1.414$. Traženi interval je $[\bar{x} - k\sigma, \bar{x} + k\sigma] = [\bar{x} - \sqrt{2}\sigma, \bar{x} + \sqrt{2}\sigma] = [7.525, 12.475]$.

Zadaci**Zadatak 1.**

Dani su brojevi: 4, 3, 1, 5, 5, 1, 6, 4, 6, 1, 4, 6, 2, 3, 3, 6, 6, 4, 3, 3.

Odredite razrede, sastavite tabelu frekvencija s kumulativnom funkcijom te izračunajte aritmetičku sredinu, varijancu i standardnu devijaciju.

Rješenje.

zadanih brojeva ima: 20

razredi: 1, 2, 3, 4, 5, 6

broj razreda: 6

tabela frekvencija i kumulativna funkcija:

x	1	2	3	4	5	6
f	3	1	5	4	2	5
rf	0.15	0.05	0.25	0.2	0.1	0.25
cf	0.15	0.2	0.45	0.65	0.75	1

aritmetička sredina: $\bar{x} = 3.8$

varijanca: $\sigma^2 = \bar{x}^2 - \bar{x}^2 = 17.3 - 14.44 = 2.86$

standardna devijacija: $\sigma = 1.69$

Zadatak 2.

Dani su brojevi: 6, 3, 2, 3, 2, 4, 6, 5, 6, 5, 4, 1, 4, 1, 4, 5, 3, 4, 2, 4.

Odredite razrede, sastavite tabelu frekvencija s kumulativnom funkcijom te izračunajte aritmetičku sredinu, varijancu i standardnu devijaciju.

Zadatak 3.

Dani su brojevi: 94, 97, 103, 103, 96, 104, 99, 100, 97, 102, 95, 102, 103, 103, 95, 101, 99, 102, 103, 102, 100, 97, 96, 99, 98, 98, 102, 103, 96, 94, 97.

Odredite razrede, sastavite tabelu frekvencija s kumulativnom funkcijom te izračunajte aritmetičku sredinu, varijancu i standardnu devijaciju.

Zadatak 4.

Koji je najmanji interval oko aritmetičke sredine $\bar{x} = 12$ za koji sa sigurnošću možemo tvrditi da sadrži barem 60% zadanih podataka, ako je $\sigma = 1.3$?

6 Usporedivanje različitih mjerena. Usporedivanje različitih rezultata

Usporedivanje različitih mjerena

Mjerenja iste veličine se mogu izvoditi različitim načinima ili instrumentima i pojavljuje se potreba za jednostavnim postupkom uspoređivanja rezultata različitih mjerena. U takvoj situaciji imamo dva niza brojeva x_1, x_2, \dots, x_n i y_1, y_2, \dots, y_m i njima pripadajuće aritmetičke sredine i standardne devijacije \bar{x}, σ_x i \bar{y}, σ_y .

Željeli bismo ocijeniti koje mjerjenje je bilo točnije. To općenito nije moguće odrediti, ali je za očekivati da je točnije ono mjerjenje s manjim odstupanjima od aritmetičke sredine (s manjim *rasipanjem* oko aritmetičke sredine), tj. s manjom standardnom devijacijom. U primjeru koji slijedi to ne možemo odmah napraviti, jer su vrijednosti u različitim mjernim jedinicama. Jedan način bi bio da jarde pretvorimo u metre, a drugi je da usporedimo omjere standardne devijacije i aritmetičke sredine. Mjerjenje kod kojeg je taj omjer manji je preciznije.

Dakle uspoređujemo brojeve $\frac{\sigma_x}{\bar{x}}$ i $\frac{\sigma_y}{\bar{y}}$ i manji omjer određuje preciznije mjerjenje.

Primjer 1.

Dva niza mjerena iste udaljenosti su dala sljedeće podatke:

mjerjenje A: $105m, 98m, 92m, 107m, 101m$ (u metrima)

mjerjenje B: $109yd, 113yd, 107yd, 115yd, 111yd$ (u jardima; $1yd = 0.914m$)

Aritmetička sredina i standardna devijacija tih dvaju mjerena su redom:

mjerjenje A: $\bar{x} = 100.6m, \sigma \approx 5.3m, \frac{\sigma}{\bar{x}} \approx \frac{5.3m}{100.6m} \approx 0.053$

mjerjenje B: $\bar{x} = 111yd, \sigma \approx 2.82yd, \frac{\sigma}{\bar{x}} \approx \frac{2.82yd}{111yd} \approx 0.025$

Mjerjenje **B** ima manji omjer i smatramo ga preciznijim.

Zadatak 1.

Dva niza mjerena istog volumena su dala sljedeće podatke:

mjerjenje A: $101.5l, 101.3l, 98.7l, 99.4l, 101.2l$ (u litrama)

mjerjenje B: $26.1gal, 26.1gal, 25.9gal, 26.7gal, 26.6gal$ (u galonima; $1gal = 3.7854l$)

Odredite koje je od ova dva mjerena preciznije.

Uspoređivanje različitih rezultata

Tipična situacija u kojoj treba uspoređivati različite rezultate su ispitni. Naime, veći broj dobivenih bodova ne znači odmah i bolji uspjeh. Pogledajmo primjer.

Marko je dobio 85 bodova na testu na kojem je prosječna vrijednost bodova svih kandidata bila $\bar{x} = 79$, a $\sigma = 8$. Alen je dobio 75 bodova na testu, koji ne mora provjeravati isto gradivo, a na kojem je $\bar{x} = 70$, a $\sigma = 5$. Tko je postigao bolji rezultat od njih dvojice? Ne možemo direktno uspoređivati njihove bodove zbog različitih \bar{x} i σ , što znači da njihove rezultate moramo prevesti u standardne jedinice. Ako je x jedan od danih brojeva iz uzorka sa aritmetičkom sredinom \bar{x} i standardnom devijacijom σ , onda se x prevodi u standardnu jedinicu z formulom

$$z = \frac{x - \bar{x}}{\sigma}.$$

Ovom formulom se niz brojeva x_1, x_2, \dots, x_n s aritmetičkom sredinom \bar{x} i standardnom devijacijom σ prevodi u niz brojeva z_1, z_2, \dots, z_n s aritmetičkom sredinom 0 i standardnom devijacijom 1:

$$\begin{aligned}\bar{z} &= \frac{1}{n}(z_1 + z_2 + \dots + z_n) = \frac{1}{n} \left(\frac{x_1 - \bar{x}}{\sigma} + \frac{x_2 - \bar{x}}{\sigma} + \dots + \frac{x_n - \bar{x}}{\sigma} \right) = \frac{1}{n\sigma} (x_1 + x_2 + \dots + x_n - n \cdot \bar{x}) = \\ &= \frac{1}{\sigma} \left(\frac{1}{n} (x_1 + x_2 + \dots + x_n) - \bar{x} \right) = 0 \\ \sigma_z^2 &= \bar{z}^2 - \bar{z}^2 = \bar{z}^2 = \frac{1}{n} (z_1^2 + z_2^2 + \dots + z_n^2) = \frac{1}{n} \left(\left(\frac{x_1 - \bar{x}}{\sigma} \right)^2 + \left(\frac{x_2 - \bar{x}}{\sigma} \right)^2 + \dots + \left(\frac{x_n - \bar{x}}{\sigma} \right)^2 \right) = \\ &= \frac{1}{n} \left(\frac{(x_1 - \bar{x})^2}{\sigma^2} + \frac{(x_2 - \bar{x})^2}{\sigma^2} + \dots + \frac{(x_n - \bar{x})^2}{\sigma^2} \right) = \frac{1}{\sigma^2} \left(\frac{(x_1 - \bar{x})^2}{n} + \frac{(x_2 - \bar{x})^2}{n} + \dots + \frac{(x_n - \bar{x})^2}{n} \right) = 1\end{aligned}$$

U primjeru imamo redom:

$$\text{Marko } z = \frac{85 - 79}{8} = 0.75$$

$$\text{Alen } z = \frac{75 - 70}{5} = 1$$

U standardnim jedinicama veća vrijednost označava bolji rezultat pa zaključujemo da je Alen bolje riješio test.

Zadatak 1.

Na jednom drugom testiranju je Marko dobio 95 bodova na testu na kojem je prosječna vrijednost bodova svih kandidata bila $\bar{x} = 90$, a $\sigma = 6$. Alen je dobio 86 bodova na testu, koji ne mora provjeravati isto gradivo, a na kojem je $\bar{x} = 80$, a $\sigma = 8$. Tko je postigao bolji rezultat od njih dvojice?

7 Linearna regresija

Zadani su parovi brojeva: $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Postoji mogućnost da su te točke grupirane oko nekog pravca, što bi značilo da postoji linearna povezanost (korelacija) između nizova brojeva x_1, x_2, \dots, x_n i y_1, y_2, \dots, y_n . Jednadžba tog pravca, kojeg zovemo *pravac regresije* je

$$y = ax + b$$

gdje su a i b nepoznati brojevi koje dobivamo *metodom najmanjih kvadrata*. Navedenom metodom dobivamo sljedeće formule za tražene brojeve a i b :

$$a = \frac{\sum_{j=1}^n x_j y_j - n\bar{x} \cdot \bar{y}}{\sum_{j=1}^n x_j^2 - n\bar{x}^2} = \frac{\frac{1}{n} \sum_{j=1}^n x_j y_j - \bar{x} \cdot \bar{y}}{\frac{1}{n} \sum_{j=1}^n x_j^2 - \bar{x}^2} = \frac{\bar{xy} - \bar{x} \cdot \bar{y}}{\bar{x}^2 - \bar{x}^2} = \frac{\bar{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2},$$

$$b = \bar{y} - a\bar{x},$$

gdje je $\bar{xy} = \frac{1}{n} \sum_{j=1}^n x_j y_j$. Primijetimo da iz formule za b slijedi da se točka (\bar{x}, \bar{y}) nalazi na traženom pravcu.

Odstupanje zadanih podataka od pravca zadanog jednadžbom $y = cx + d$ računamo po formuli

$$\frac{1}{n} \sum_{k=1}^n (y_k - (cx_k + d))^2.$$

Broj izračunat tom formulom zovemo *srednje kvadratno odstupanje* zadanih podataka od tog pravca. *Pravac regresije* je jedinstveni pravac kojem je srednje kvadratno odstupanje, (tj., $\frac{1}{n} \sum_{k=1}^n (y_k - (ax_k + b))^2$) najmanje moguće.

Koeficijent korelacije r je mjera linearne povezanosti nizova brojeva x_1, \dots, x_n i y_1, \dots, y_n , a računamo ga prema formuli:

$$r = \frac{\sum_{j=1}^n x_j y_j - n\bar{x} \cdot \bar{y}}{\sqrt{\sum_{j=1}^n x_j^2 - n\bar{x}^2} \sqrt{\sum_{j=1}^n y_j^2 - n\bar{y}^2}} = \frac{\bar{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\bar{x}^2 - \bar{x}^2} \sqrt{\bar{y}^2 - \bar{y}^2}} = \frac{\bar{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y}$$

Uvijek vrijedi: $-1 \leq r \leq 1$ i što je $|r|$ bliži broju 1, to su zadani nizovi brojeva bolje linearne povezani. Ako je r blizak nuli, onda je linearne povezanost slaba ili nikakva. Zgodno je znati da je predznak broja r isti kao i predznak koeficijenta smjera pravca regresije.

Da bismo izračunali a i b , prvo sastavljamo sljedeću tabelu:

j	x	y	$x \cdot y$	x^2	y^2
1	x_1	y_1	$x_1 y_1$	x_1^2	y_1^2
2	x_2	y_2	$x_2 y_2$	x_2^2	y_2^2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n	x_n	y_n	$x_n y_n$	x_n^2	y_n^2
\sum	$\sum_{j=1}^n x_j$	$\sum_{j=1}^n y_j$	$\sum_{j=1}^n x_j y_j$	$\sum_{j=1}^n x_j^2$	$\sum_{j=1}^n y_j^2$
	\bar{x}	\bar{y}	$\bar{x}\bar{y}$	\bar{x}^2	\bar{y}^2

Vrijednosti iz zadnjeg retka uvrštavamo u pripadajuće formule.

Pomoću jednadžbe pravca regresije možemo procijeniti vrijednosti koje nisu zadane, tj. za $x \notin \{x_1, x_2, \dots, x_n\}$ možemo, uvrstivši ga u jednadžbu dobiti aproksimaciju njemu pripadajućeg $y = ax + b$. Slično, za $y \notin \{y_1, y_2, \dots, y_n\}$ pripadajuću aproksimaciju za x dobivamo uvrštanjem y u jednadžbu pravca regresije $y = ax + b$ i rješavanjem po x .

Primjer 1.

Mjerenjem statističkih obilježja x i y su dobiveni sljedeći brojevi:

x	3	5	6	8	10
y	3	4	2	2	2

1. Prikažite točke (x_j, y_j) u pravokutnom koordinatnom sustavu.
2. Odredite jednadžbu pravca regresije i skicirajte ga zajedno s točkama.
3. Izračunajte koeficijent korelacijske.
4. Izračunajte srednje kvadratno odstupanje podataka iz tabele od odgovarajućih vrijednosti za pravac regresije.

Rješenje.

j	x_j	y_j	$x_j \cdot y_j$	x_j^2	y_j^2
1	3	3	9	9	9
2	5	4	20	25	16
3	6	2	12	36	4
4	8	2	16	64	4
5	10	2	20	100	4
Σ	32	13	77	234	37
	6.4	2.6	15.4	46.8	7.4

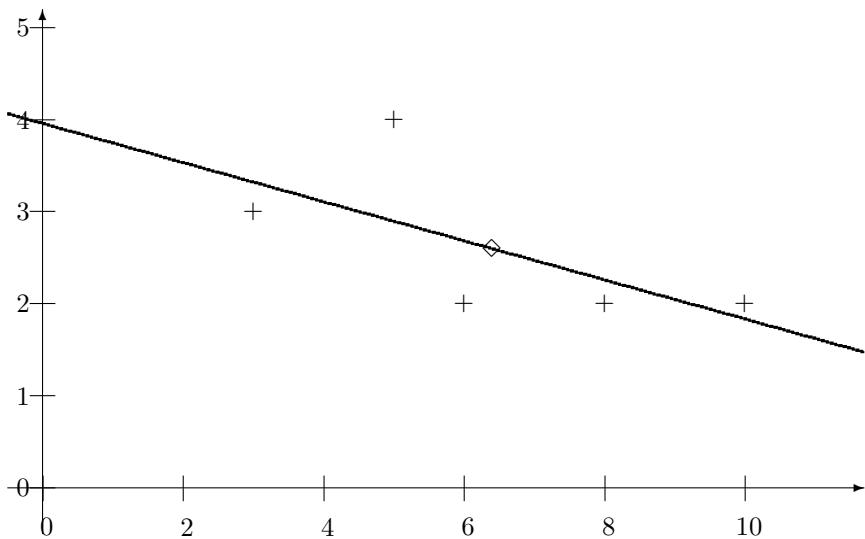
$$\bar{x} = \frac{32}{5} = 6.4, \bar{y} = \frac{13}{5} = 2.6, \bar{x} \cdot \bar{y} = 16.64, \bar{x}\bar{y} = 15.4, \bar{x}^2 = 46.8, \bar{y}^2 = 7.4$$

$$y = ax + b; \quad a = \frac{\bar{xy} - \bar{x} \cdot \bar{y}}{\bar{x^2} - \bar{x}^2}; \quad b = \bar{y} - a\bar{x}$$

$$y = -0.212 \cdot x + 3.959$$

$$r = \frac{\bar{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\bar{x^2} - \bar{x}^2} \sqrt{\bar{y^2} - \bar{y}^2}} = -0.641, \quad r^2 = 0.411$$

Srednje kvadratno odstupanje je: $0.3767 = \frac{1}{n} \sum_{k=1}^n (y_k - (ax_k + b))^2$.



Posebno označena točka na pravcu ima koordinate $(\bar{x}, \bar{y}) = (6.4, 2.6)$.

Primjer 2.

Dvije standardne igraće kockice, plavu i crvenu, bacamo 11 puta. Neka je x_k broj koji je "pao" na plavoj kockici u k -tom bacanju, a y_k suma brojeva koji su "pali" na plavoj i crvenoj kockici u k -tom bacanju.

Na plavoj kockici su pali redom sljedeći brojevi: 3, 3, 2, 3, 4, 4, 2, 1, 1, 4, 3,
a na crvenoj: 2, 3, 1, 4, 3, 2, 4, 1, 2, 4, 4,

Nadite jednadžbu pravca regresije za točke $(x_k, y_k), k = 1, \dots, 11$ i skicirajte ga u koordinatnoj ravni zajedno sa točkama. Izračunajte koeficijent korelacije.

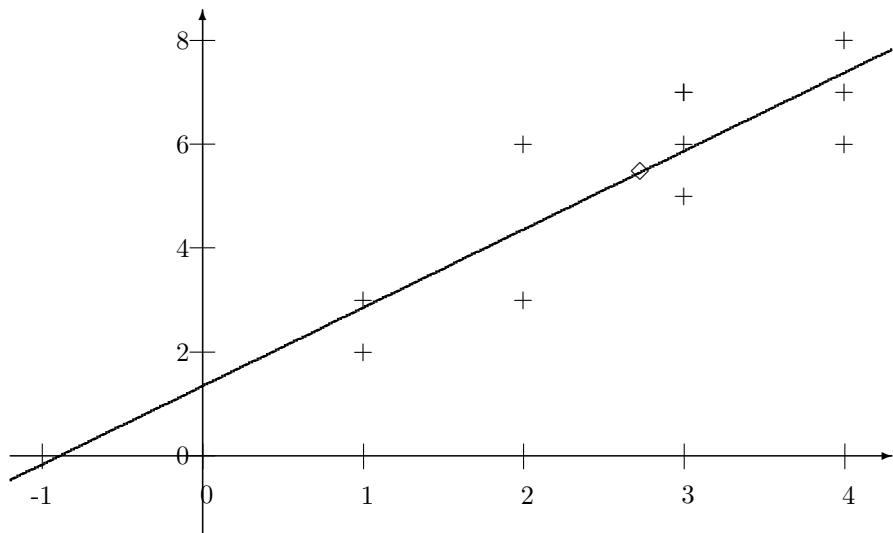
Rješenje.

j	x_j	y_j	$x_j \cdot y_j$	x_j^2	y_j^2
1	3	5	15	9	25
2	3	6	18	9	36
3	2	3	6	4	9
4	3	7	21	9	49
5	4	7	28	16	49
6	4	6	24	16	36
7	2	6	12	4	36
8	1	2	2	1	4
9	1	3	3	1	9
10	4	8	32	16	64
11	3	7	21	9	49
Σ	30	60	182	94	366
	2.73	5.45	16.55	8.55	33.27

$$\bar{x} = \frac{30}{11} = 2.73, \quad \bar{y} = \frac{60}{11} = 5.45$$

$$y = 1.507 \cdot x + 1.343, \quad r = 0.85, \quad r^2 = 0.71$$

srednje kvadratno odstupanje je 1.0041



Posebno označena točka na pravcu ima koordinate $(\bar{x}, \bar{y}) = (2.73, 5.45)$.

Primjer 3.

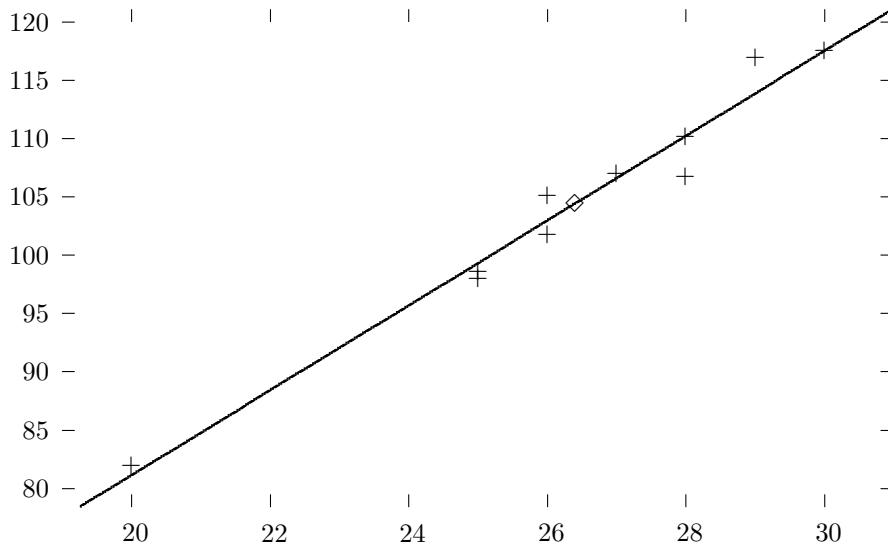
Zadane su sljedeće točke: $(20, 82), (25, 98), (25, 98.6), (26, 101.8), (28, 106.8), (27, 107), (26, 105.2), (28, 110.2), (30, 117.6), (29, 117)$. Nađite jednadžbu pravca regresije i skicirajte ga u koordinatnoj ravnini zajedno sa zadanim točkama. Izračunajte koeficijent korelacijske funkcije.

Rješenje.

j	x_j	y_j	$x_j \cdot y_j$	x_j^2	y_j^2
1	20	82	1640	400	6724
2	25	98	2450	625	9604
3	25	98.6	2465	625	9721.96
4	26	101.8	2646.8	676	10363.24
5	28	106.8	2990.4	784	11406.24
6	27	107	2889	729	11449
7	26	105.2	2735.2	676	11067.04
8	28	110.2	3085.6	784	12144.04
9	30	117.6	3528	900	13829.76
10	29	117	3393	841	13689
Σ	264	1044.2	27823	7040	109998.28
	26.4	104.42	2782.3	704	10999.83

$$\bar{x} = \frac{264}{10} = 26.4, \quad \bar{y} = \frac{1044.2}{10} = 104.42$$

Jednadžba pravca regresije je $y = 3.638 \cdot x + 8.375$. Koeficijent korelacijske funkcije je $r = 0.98$, $r^2 = 0.97$. Srednje kvadratno odstupanje je 3.1134.



Posebno označena točka na pravcu ima koordinate $(\bar{x}, \bar{y}) = (26.4, 104.42)$.

Primjer 4.

Broj stanovništva u milijunima jednog grada u godinama kada se provodilo popisivanje je dano sljedećem tabelom:

godina	1857	1869	1880	1890	1900	1910	1921	1931	1948	1953	1961	1971	1981	1991
stanovnika (mil.)	1.09	1.2	1.25	1.43	1.58	1.73	1.72	1.89	1.89	1.97	2.08	2.21	2.3	2.39

Nadite jednadžbu pravca regresije i skicirajte ga u koordinatnoj ravnini zajedno sa zadanim točkama. Procijenite koliko je stanovnika bilo 1940. godine. Procijenite koje godine je bilo dva milijuna stanovnika.

Rješenje.

j	x_j	y_j	$x_j \cdot y_j$	x_j^2	y_j^2
1	1857	1.09	2024.13	3448449	1.1881
2	1869	1.2	2242.8	3493161	1.44
3	1880	1.25	2350	3534400	1.5625
4	1890	1.43	2702.7	3572100	2.0449
5	1900	1.58	3002	3610000	2.4964
6	1910	1.73	3304.3	3648100	2.9929
7	1921	1.72	3304.12	3690241	2.9584
8	1931	1.89	3649.59	3728761	3.5721
9	1948	1.89	3681.72	3794704	3.5721
10	1953	1.97	3847.41	3814209	3.8809
11	1961	2.08	4078.88	3845521	4.3264
12	1971	2.21	4355.91	3884841	4.8841
13	1981	2.3	4556.3	3924361	5.29
14	1991	2.39	4758.49	3964081	5.7121
Σ	26963	24.73	47858.35	51952929	45.9209
	1925.9286	1.7664	3418.4536	3710923.5	3.2801

$$a \approx \frac{3418.4536 - 1925.9286 \cdot 1.7664}{3710923.5 - 1925.9286^2} \approx 0.00958$$

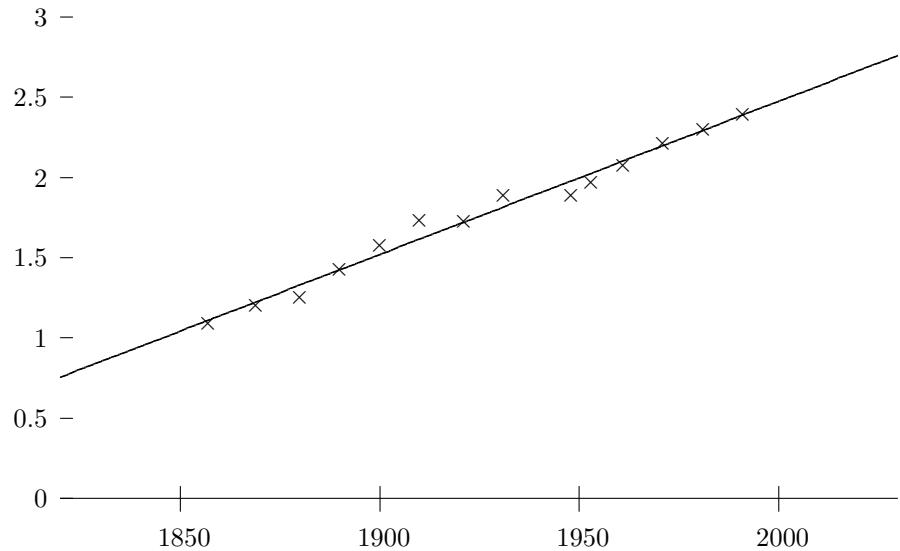
$$b \approx 1.7664 - 1925.9286 \cdot 0.00958 \approx -16.68$$

Napomenimo da je ovo računanje osjetljivo na greške zaokruživanja. Npr. da su vrijednosti iz zadnjeg retka tabele zaokružene na dvije decimale, dobili bismo netočne vrijednosti za a i b .

Srednje kvadratno odstupanje je 0.00295.

Godine 1940. je bilo približno $a \cdot 1940 + b \approx 1.91$ milijuna stanovnika.

Dva milijuna stanovnika je bilo otprilike $(2 - b)/a \approx 1950$. godine.



Zadaci

Zadatak 1. Zadane su sljedeće točke: $(14, 24.8), (12, 31.6), (14, 28), (17, 20.8), (18, 20.6), (17, 22.2), (20, 17.2), (17, 25.6), (22, 15.2), (21, 18.2)$ Nađite jednadžbu pravca regresije i skicirajte ga u koordinatnoj ravnini zajedno sa zadanim točkama. Izračunajte koeficijent korelacije.

Zadatak 2. Dvije standardne igrače kockice, plavu i crvenu, bacamo 11 puta. Neka je x_k broj koji je "pao" na plavoj kockici u k -tom bacanju, a y_k suma brojeva koji su "pali" na plavoj i crvenoj kockici u k -tom bacanju.

Na plavoj kockici su pali redom sljedeći brojevi: 4, 5, 4, 1, 2, 4, 2, 2, 6, 3, 3,
a na crvenoj: 4, 5, 1, 3, 1, 6, 1, 5, 5, 1, 2,

Nađite jednadžbu pravca regresije za točke $(x_k, y_k), k = 1, \dots, 11$ i skicirajte ga u koordinatnoj ravnini zajedno sa točkama. Izračunajte koeficijent korelacije.

VJEROJATNOST I STATISTIKA

PRIMJER 1. KOLOKVIJA

Napomena: sve decimalne brojeve izračunajte na četiri decimale točno, uz zaokruživanje, ako je to potrebno.

1. Zadani su brojevi:

0	5	2	3	-1
0	4	4	2	5
1	-1	3	2	2
4	-1	5	0	4

- (a) Odredite razrede i sastavite tabelu frekvencija, relativnih frekvencija i kumulativne funkcije relativnih frekvencija.
 (b) Izračunajte aritmetičku sredinu \bar{x} , varijancu σ^2 , standardnu devijaciju σ i mod(ove).
 (c) Skicirajte histogram frekvencija.

2. Zadana je sljedeća tabela frekvencija:

x	-1	0	1	2	3	4	5	7
f	4	4	1	3	1	3	1	3
rf	0.2	0.2	0.05	0.15	0.05	0.15	0.05	0.15
cf	0.2	0.4	0.45	0.6	0.65	0.8	0.85	1

- (a) Izračunajte $cf(3)$, $cf(-0.69)$ i $cf(-1.6)$.
 (b) Skicirajte graf kumulativne funkcije relativnih frekvencija.
 (c) Koristeći se vrijednostima kumulativne funkcije relativnih frekvencija izračunajte koliko posto podataka ima vrijednost
 i. veću od 0, $X > 0$
 ii. manju od 3, $X < 3$
 iii. veću od 0 i manju od ili jednaku 5, $0 < X \leq 5$
 iv. veću od 1 i manju od 4, $1 < X < 4$
 (d) Koristeći se tabelom relativnih frekvencija odredite
 i. medijan \tilde{x}
 ii. prvi kvartil Q_1
 iii. 65. percentil (P_{65})
 iv. kvantil $q_{0.53}$

3. Mjerenjem statističkih obilježja x i y su dobiveni sljedeći brojevi:

x	2	3	4	5	7
y	4	2	5	1	6

- (a) Prikažite točke (x_j, y_j) u pravokutnom koordinatnom sustavu.
 (b) Odredite jednadžbu pravca regresije i skicirajte ga zajedno s točkama.
 (c) Izračunajte koeficijent korelacije.

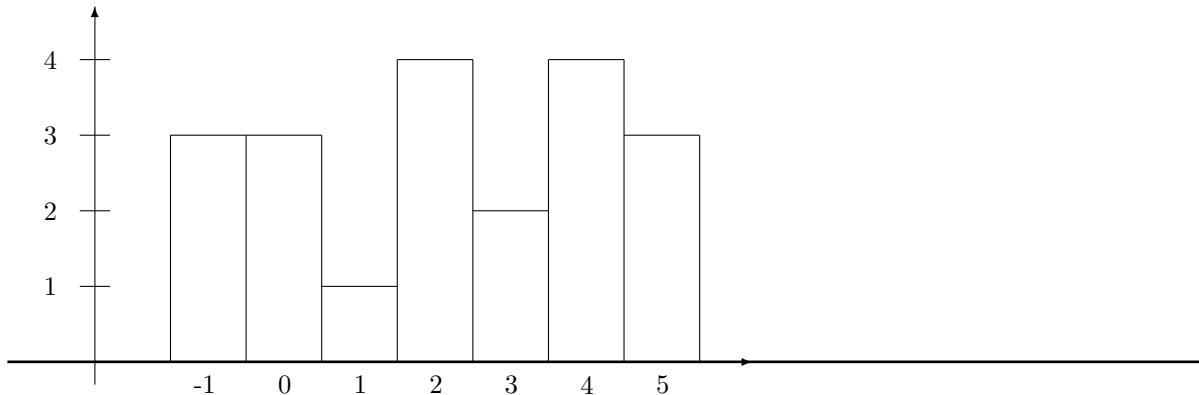
SKICE RJEŠENJA

Zadano je $n = 20$ brojeva. Razredi su: -1, 0, 1, 2, 3, 4, 5 i ih m = 7.

x	-1	0	1	2	3	4	5
f	3	3	1	4	2	4	3
rf	0.15	0.15	0.05	0.2	0.1	0.2	0.15
cf	0.15	0.3	0.35	0.55	0.65	0.85	1

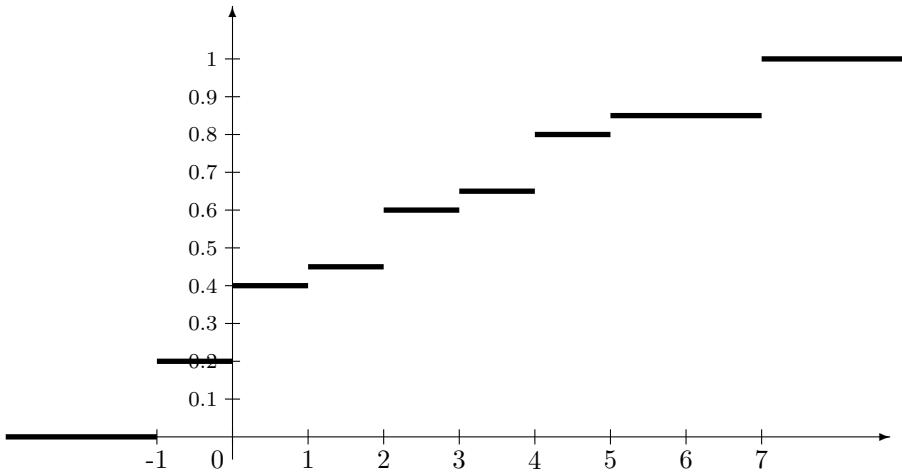
modovi su: 2, 4, $\bar{x} = 2.15$, $\sigma^2 = \bar{x}^2 - \bar{x}^2 = 8.85 - 4.6225 = 4.2275$, $\sigma = 2.0561$.

histogram:



$$cf(3) = 0.65, cf(-0.69) = 0.2, cf(-1.6) = 0$$

graf kumulativne funkcije relativnih frekvencija:



$$[X > 0 : 60\%], [X < 3 : 60\%], [0 < X \leq 5 : 45\%], [1 < X < 4 : 20\%]$$

$$\tilde{x} = Q_2 = P_{50} = 2, \quad Q_1 = P_{25} = 0, \quad P_{65} = 3.5, \quad q_{0.53} = 2$$

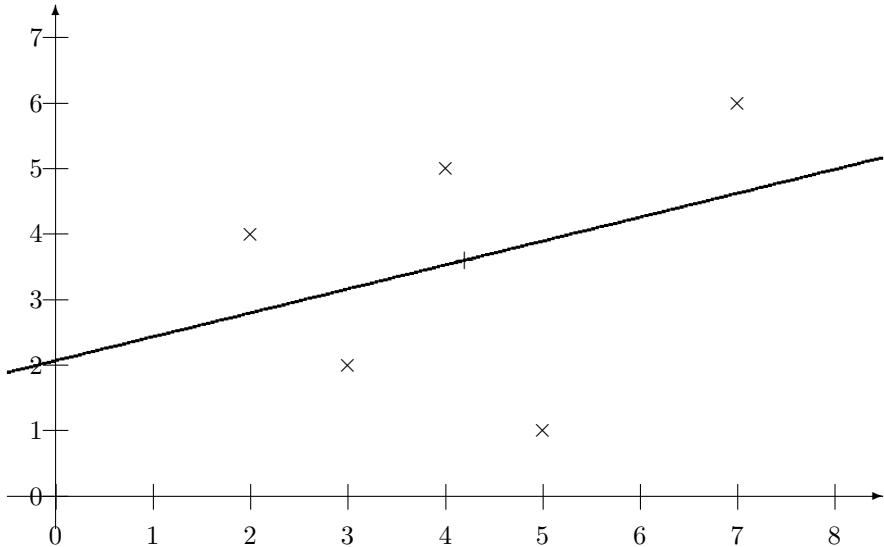
j	x_j	y_j	$x_j \cdot y_j$	x_j^2	y_j^2
1	2	4	8	4	16
2	3	2	6	9	4
3	4	5	20	16	25
4	5	1	5	25	1
5	7	6	42	49	36
Σ	21	18	81	103	82
	4.2	3.6	16.2	20.6	16.4

$$\bar{x} = \frac{21}{5} = 4.2, \bar{y} = \frac{18}{5} = 3.6, \bar{x} \cdot \bar{y} = 15.12, \bar{x}\bar{y} = 16.2, \bar{x}^2 = 20.6, \bar{y}^2 = 16.4$$

$$y = a \cdot x + b; \quad a = \frac{\bar{x}\bar{y} - \bar{x} \cdot \bar{y}}{\bar{x}^2 - \bar{x}^2}; \quad b = \bar{y} - a \cdot \bar{x}$$

$$y = 0.3649 \cdot x + 2.0676$$

$$r = \frac{\bar{x}\bar{y} - \bar{x} \cdot \bar{y}}{\sqrt{\bar{x}^2 - \bar{x}^2} \sqrt{\bar{y}^2 - \bar{y}^2}} = 0.3385, \quad r^2 = 0.1146$$



Posebno označena točka na pravcu ima koordinate $(\bar{x}, \bar{y}) = (4.2, 3.6)$.

Dio II

Vjerojatnost

8 Događaj. Vjerojatnost

Začeci teorije vjerojatnosti su stvoren i sredinom 17. stoljeća, kada je francuski plemić Chevalier de Méré postavio nekoliko pitanja o vjerojatnosti nekih događaja prilikom kockanja Blaisu Pascalu. Na osnovu rasprave koja je uslijedila, Pascal i Pierre de Fermat su razvili osnove teorije vjerojatnosti.

De Méré se kladio na sljedeće događaje:

- U četiri bacanja simetrične kockice će pasti barem jedna šestica.
- U dvadesetčetiri bacanja dviju simetričnih kockica će barem jednom pasti šestica na obje kockice.

Vidjet ćemo kasnije kako možemo izračunati vjerojatnosti navedenih događaja.

Teorija vjerojatnosti se bavi modeliranjem procesa kojima nije moguće unaprijed predvidjeti ishod, kao što je to npr. bacanje novčića na slučajan način - znamo da će pasti ili pismo ili glava, ali ne znamo sa sigurnošću što će točno pasti u jednom bacanju. Mogli bismo reći da je bacanje novčića *slučajni pokus* s dva moguća *ishoda* koji su elementi skupa $\{p, g\}$.

Drugi primjer bi bio bacanje standardne igrače kockice sa šest strana s jednim od brojeva 1, 2, 3, 4, 5, 6 (ili odgovarajućim brojem točkica) na svakoj strani. Nakon što bacimo kockicu i nakon što ona mirno stane na horizontalnoj podlozi, njena gornja strana će uvijek određivati jedan od brojeva iz skupa $\{1, 2, 3, 4, 5, 6\}$ (kažemo da je npr. "pao" broj 5), ali ne možemo unaprijed znati koji od njih. Dakle na bacanje kockice također možemo gledati kao na slučajni pokus sa šest različitih ishoda. Neka je $A = \{2, 3, 5\}$ i ako bacanjem kockice padne broj 2, reći ćemo da je nastupio događaj A .

Svaki od brojeva koji mogu pasti na kockici se smatra *elementarnim događajem* (intuitivno, elementarni događaj se ne prikazuje preko jednostavnijih događaja, bilo zato što se to ne može ili zato što nas to ne zanima), a skup svih elementarnih događaja Ω zovemo *prostor elementarnih događaja*.

Događaj se definira kao bilo koji podskup prostora elementarnih događaja Ω , tj. ako je $A \subseteq \Omega$, onda je skup A događaj. Broj elemenata skupa A označavamo s $|A|$. Za događaj A ćemo reći da je nastupio ako je ishod nekog slučajnog pokusa jedan od elementarnih događaja iz A , npr. u primjeru s bacanjem kockice možemo promatrati događaje

$A = \text{pao je prost broj} = \{2, 3, 5\}$ i

$B = \text{pao je neparan broj} = \{1, 3, 5\}$

pa ako je bacanjem kockice pao broj 2, onda je događaj A nastupio, a događaj B nije.

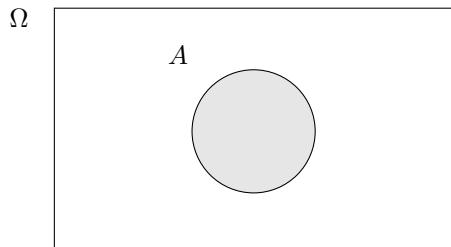
Kako su događaji skupovi, onda je s njima moguće obavljati skupovne operacije kao što su komplement, presjek, unija i razlika. Tako u gornjem primjeru možemo npr. promotriti događaje

$\bar{A} = \Omega \setminus A = \text{nije pao prost broj} = \{1, 4, 6\}$,

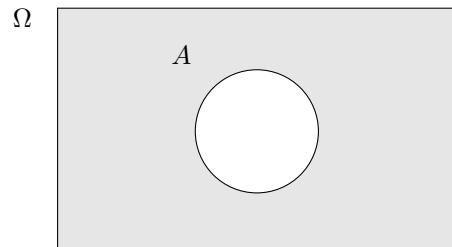
$A \cap B = \text{pao je neparan i prost broj} = \{3, 5\}$,

$A \cup B = \text{pao je neparan ili prost broj} = \{1, 2, 3, 5\}$ i

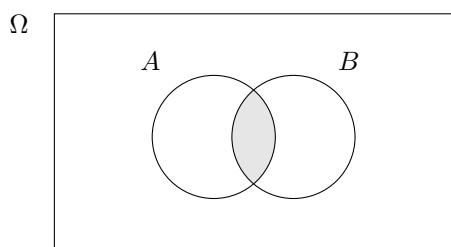
$A \setminus B = \text{pao je prost broj koji nije neparan} = \{2\}$.



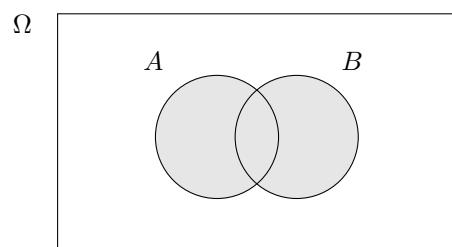
događaj A



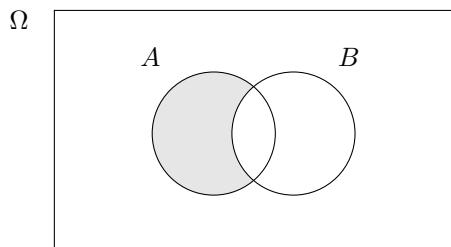
događaj \bar{A}



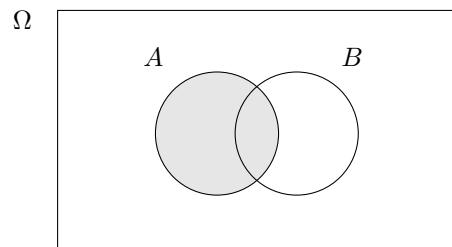
presjek događaja A i B : $A \cap B$



unija događaja A i B : $A \cup B$



razlika događaja A i B : $A \setminus B$



$A = (A \setminus B) \cup (A \cap B)$

Primjer 1.

Promotrimo sve moguće ishode bacanja dvaju novčića. Na svakom od njih može pasti ili pismo ili glava pa je prostor elementarnih događaja jednak $\Omega = \{pp, pg, gp, gg\}$. (Radi jednostavnosti zamislimo da su novčići različite veličine i da prvo zapisujemo ishod bacanja većeg od njih.)

Primjer 2.

Promotrimo sve moguće ishode bacanja dviju standardnih kockica, crvene i plave. Jeden ishod je predstavljen uređenim parom brojeva (c, p) gdje prva komponenta para predstavlja broj koji je pao na crvenoj kockici, a druga komponenta predstavlja broj koji je pao na plavoj kockici. Prostor elementarnih događaja Ω je skup svih takvih parova brojeva $\Omega = \{(1, 1), (1, 2), \dots, (6, 6)\}$ kojeg se jednostavno može prikazati sljedećom tabelom:

$c \setminus p$	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Sljedeći događaji su označeni u gornjoj tabeli:

$A =$ manji od dva broja koji su pali je 3 = $\{(3, 3), (3, 4), (3, 5), (3, 6), (4, 3), (5, 3), (6, 3)\}$ (označeni pravokutnicima)

$B =$ razlika većeg i manjeg broja je 2 = $\{(3, 1), (4, 2), (5, 3), (6, 4), (1, 3), (2, 4), (3, 5), (4, 6)\}$ (označeni ovalima)

$A \cup B =$ elementarni događaji označeni pravokutnicima ili ovalima

$A \cap B =$ elementarni događaji označeni pravokutnicima i ovalima

Primjer 3.

Jedan šahovski klub ima tri igrača K_1, K_2, K_3 među kojima treba slučajno izabrati dvojicu koji će predstavljati klub na turniru. Slučajni pokus se sastoji od biranja dva igrača od njih tri. Svaki ishod je predstavljen dvočlanim podskupom skupa $\{K_1, K_2, K_3\}$ pa je prostor elementarnih događaja skup $\Omega = \{\{K_1, K_2\}, \{K_1, K_3\}, \{K_2, K_3\}\}$ koji ima tri elemenata.

Primjer 4.

Zamislimo slučajni pokus koji za ishod ima broj zahtjeva za prikazivanje određene web stranice tijekom nekog određenog vremenskog perioda (jedan minut, jedan sat, 24 sata, jedan mjesec...). Budući je nepoznat maksimalni broj zahtjeva, najzgodnije je za prostor elementarnih događaja uzeti skup $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$ koji je beskonačan. Iako će broj zahtjeva biti konačan, teško je, a i nepotrebno, ograničiti taj broj. Naravno, uvjek možemo za gornju granicu uzeti neki jako veliki broj za kojeg možemo biti sigurni da nikad neće biti dosegnut, ali pokazuje se da je to nepotrebno komplikiranje, a iz nekih teoretskih razloga je pogodnije za prostor elementarnih događaja uzeti navedeni beskonačni skup.

Zbog toga se koristi malo drugačija definicija prostora elementarnih događaja od navedene.

Definicija.

Prostor elementarnih događaja nekog slučajnog pokusa je skup Ω sa svojstvom da svakom ishodu pokusa odgovara točno jedan element iz tog skupa i da različitim ishodima odgovaraju različiti elementi tog skupa. *Događaj* je podskup prostora elementarnih događaja.

Ako nije navedeno drugačije smatrat ćemo da je skup svih elementarnih događaja Ω konačan skup.

Zadaci

Zadatak 1.

Napišite prostor elementarnih događaja Ω koji sadrži sve ishode bacanja triju novčića. Odredite sljedeće događaje:

A = pala su barem dva pisma

B = pale su najviše dvije glave

$A \cap B = ?$

$A \cup B = ?$

(Rješenje je na stranici 122.)

Zadatak 2.

Promotrimo sve moguće ishode bacanja dviju standardnih kockica, crvene i plave. Jedan ishod je predstavljen uređenim parom brojeva (c, p) gdje prva komponenta para predstavlja broj koji je pao na crvenoj kockici, a druga komponenta predstavlja broj koji je pao na plavoj kockici. Odredite sljedeće događaje:

P_p = produkt brojeva koji su pali na kockicama je prost broj

P_s = suma brojeva koji su pali na kockicama je prost broj

A_6 = suma brojeva koji su pali na kockicama je 6

Zadatak 3.

U posudi se nalazi pet zelenih i dvije žute kuglice. Slučajno se, bez vraćanja, jedna po jedna izvlače dvije kuglice. Odredite prostor elementarnih događaja za navedeni slučajni pokus. Kako bi izgledao prostor elementarnih događaja ako bismo svaku kuglicu vratili natrag u posudu odmah nakon njenog izvlačenja? (Rješenje je na stranici 122.)

Klasična definicija vjerojatnosti 'a priori'

Neka je $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$ prostor elementarnih događaja vezan uz neki slučajni pokus tako da su svi elementarni događaji, kojih mora biti konačan broj, jednako mogući. Neka je A događaj, tj. $A \subseteq \Omega$. Tada se klasična definicija vjerojatnosti '*a priori*' događaja A definira formulom

$$P(A) = \frac{m_A}{m} = \frac{|A|}{|\Omega|} = \frac{\text{broj povoljnih elementarnih događaja}}{\text{broj svih elementarnih događaja}} \quad (4)$$

gdje je $m_A = |A|$ broj elementarnih događaja od kojih se sastoji događaj A , a $m = |\Omega|$ broj svih elementarnih događaja iz prostora elementarnih događaja Ω ($|\Omega|$ je broj elemenata skupa Ω). Naravno, da bi formula imala smisla, skup Ω mora biti konačan.

Svojstva vjerojatnosti

1. Direktno iz definicije se dobiva

$$P(\Omega) = \frac{|\Omega|}{|\Omega|} = 1 \quad \text{i} \quad P(\emptyset) = \frac{|\emptyset|}{|\Omega|} = 0$$

2. Za svaki događaj A vrijedi $\emptyset \subseteq A \subseteq \Omega$ iz čega zbog $0 = |\emptyset| \leq |A| \leq |\Omega|$ dobivamo (dijeljenjem s $|\Omega|$) sljedeće:

$$0 \leq P(A) \leq 1.$$

3. Ako su A i B događaji takvi da vrijedi $A \subseteq B$, onda vrijedi $|A| \leq |B|$ odakle (dijeljenjem s $|\Omega|$) dobivamo

$$P(A) \leq P(B).$$

4. Ako je $A \subseteq \Omega$ događaj, onda je $\bar{A} = \Omega \setminus A$ događaj suprotan događaju A i zbog $|\bar{A}| = |\Omega| - |A|$ vrijedi

$$P(\bar{A}) = \frac{|\bar{A}|}{|\Omega|} = \frac{|\Omega| - |A|}{|\Omega|} = 1 - P(A).$$

5. Ako su A i B disjunktni događaji ($A \cap B = \emptyset$), $A, B \subseteq \Omega$ onda vrijedi $|A \cup B| = |A| + |B|$ pa imamo

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

6. Neka su A i B događaji, $A, B \subseteq \Omega$. Sada se zbog $A \cup B = A \cup (B \setminus A)$ i činjenice da je ta unija disjunktna (pa se može primijeniti prethodna jednakost) dobiva

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

Zbog $B = (B \setminus A) \cup (A \cap B)$, što je također disjunktna unija, dobivamo:

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

što zajedno s prethodnom relacijom daje sljedeću formulu za vjerojatnost unije dva proizvoljna događaja (aditivna formula):

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

Primjer 1.

Kolika je vjerojatnost da će bacanjem standardne simetrične kockice pasti broj veći od 2?

Rješenje.

Prostor elementarnih događaja je skup $\Omega = \{1, 2, 3, 4, 5, 6\}$, a događaj čija vjerojatnost nas zanima je $A = \{3, 4, 5, 6\}$. Broj povoljnih elementarnih događaja je $m_A = 4$, broj svih događaja je $m = 6$ pa je vjerojatnost događaja A prema klasičnoj formuli 'a priori' jednaka

$$P(A) = \frac{4}{6} = \frac{2}{3}.$$

Primjer 2.

U posudi se nalazi sedam zelenih i tri plave kuglice. Iz posude se izvlači jedna kuglica. Kolika je vjerojatnost da će izvučena kuglica biti zelene boje?

Rješenje.

Prostor elementarnih događaja je skup $\Omega = \{z_1, z_2, z_3, z_4, z_5, z_6, z_7, p_1, p_2, p_3\}$, a događaj čija vjerojatnost nas zanima je $A = \{z_1, z_2, z_3, z_4, z_5, z_6, z_7\}$. Broj povoljnih elementarnih događaja je $m_A = 7$, broj svih događaja je $m = 10$ pa je vjerojatnost događaja A prema klasičnoj formuli 'a priori' jednaka

$$P(A) = \frac{7}{10}.$$

Primjer 3.

Marko je bacio jedan, a Petar dva novčića. Kolika je vjerojatnost da je obojici pao isti broj pisama?

Rješenje.

Označimo s A događaj "Marku i Petru je pao isti broj pisama." Prostor elementarnih događaja je skup $\Omega = \{(m, p_1, p_2) | m, p_1, p_2 \in \{P, G\}\}$ (gdje je m ishod bacanja Markovog novčića, a p_1, p_2 su ishodi bacanja Petrovih novčića). Popišimo sve elementarne događaje i pogledajmo koliko ih ima povoljnih:

<i>Marko</i>	<i>Petar</i>	
P	P	P
P	P	G
P	G	P
P	G	G
G	P	P
G	P	G
G	G	P
G	G	G

Iz tablice vidimo da je $A = \{(P, P, G), (P, G, P), (G, G, G)\}$ i s obzirom da elementarnih događaja ima ukupno osam, tražena vjerojatnost je jednaka $P(A) = \frac{3}{8}$.

Principi prebrojavanja

Klasična formula vjerojatnosti 'a priori' je jednostavna, ali često nije lagano pronaći brojeve koje treba u nju uvrstiti. U tome nam pomažu principi prebrojavanja u kombinatorici. Ima ih dva: princip sume i princip produkta.

princip sume

Ako su A i B događaji koji ne mogu nastupiti istovremeno (tj. $A \cap B = \emptyset$) onda događaj $A \cup B$ može nastupiti na $|A| + |B|$ načina. (Princip sume vrijedi i ako je riječ o konačno mnogo događaja koji u parovima ne mogu nastupiti istovremeno.)

Primjer 1.

U tri posude se nalazi redom 19, 23 i 21 bombon. Na koliko načina se može uzeti jedan bombon iz svih tih posuda?

Jedan bombon iz svih posuda se može uzeti na $19 + 23 + 21 = 63$ načina.

Primjer 2.

Promatramo dva događaja prilikom bacanja simetrične kockice:

A = pao je broj veći od 3 = {4, 5, 6},

B = pao je paran prost broj = {2}.

Kako niti jedan prost broj veći od 3 nije paran, događaji A i B ne mogu nastupiti istovremeno ($A \cap B = \emptyset$) pa vrijedi $|A \cup B| = |A| + |B|$, tj. događaj $A \cup B$ može nastupiti na 4 načina.

princip produkta

Ako su događaji A i B nezavisni, tj. mogu nastupiti neovisno jedan o drugom, onda kombinacije događaja A i B mogu nastupiti na $|A| \cdot |B|$ načina. (Princip produkta vrijedi i ako je riječ o konačno mnogo događaja koji su u parovima nezavisni.)

Primjer 3.

U tri posude se nalazi redom 19, 23 i 21 bombon. Na koliko načina se mogu uzeti tri bombona, po jedan iz svake od tih posuda?

Tri bombona, po jedan iz svake od tih posuda, se može uzeti na $19 \cdot 23 \cdot 21 = 9177$ načina.

Primjer 4.

Promatramo dva događaja prilikom bacanja dviju simetričnih kockica, plave i crvene:

A = na plavoj kockici je pao paran broj = {2, 4, 6}

B = na crvenoj kockici je pao neparan prost broj = {3, 5}

Događaji A i B su očito nezavisni (rezultat bacanja plave kockice ne može utjecati na rezultat bacanja crvene i obrnuto). Sve kombinacije događaja A i B se mogu prikazati preko skupa uređenih parova (p, c) , gdje je p elementaran događaj iz skupa A , a c elementaran događaj iz skupa B . Skup svih takvih kombiniranih događaja je kartezijev produkt skupova A i B , $A \times B = \{(2, 3), (2, 5), (4, 3), (4, 5), (6, 3), (6, 5)\}$ i broj njegovih elemenata je $|A \times B| = |A| \cdot |B| = 6 \cdot 6 = 36$.

Formula uključivanja isključivanja (FUI formula)

Formula uključivanja-isključivanja se koristi za izračunavanje broja elemenata unije konačno mnogo konačnih skupova (ili komplementa takve unije) i za slučaj unije dva skupa glasi ovako:

$$|A \cup B| = |A| + |B| - |A \cap B|$$

Formula se dokazuje tako da se uoči da se u izrazu $|A| + |B|$ dva puta uračunavaju elementi koji se nalaze u oba skupa, tj. u njihovom presjeku pa se broj tih elemenata mora oduzeti.

Kako je komplement unije jednak presjeku komplementa (tj. $\overline{A \cup B} = \overline{A} \cap \overline{B}$), onda imamo

$$|\overline{A \cup B}| = |\overline{A} \cap \overline{B}| = |S| - |A \cup B| = |S| - |A| - |B| + |A \cap B|$$

gdje je S neki skup koji sadrži oba skupa A i B , a s obzirom na koji se promatra komplement. Za slučaj tri skupa formula glasi ovako:

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$

odnosno

$$|\overline{A \cup B \cup C}| = |\overline{A} \cap \overline{B} \cap \overline{C}| = |S| - |A| - |B| - |C| + |A \cap B| + |A \cap C| + |B \cap C| - |A \cap B \cap C|$$

i sada ne bi trebalo biti problem napisati formulu uključivanja-isključivanja za četiri skupa.

Općenita formula uključivanja-isključivanja:

$$\begin{aligned} \left| \bigcup_{j=1}^n A_j \right| &= |A_1 \cup A_2 \cup \dots \cup A_n| = \\ &= \sum_{j=1}^n |A_j| - \sum_{j \neq k} |A_j \cap A_k| + \sum |A_j \cap A_k \cap A_l| + \dots + (-1)^{n+1} |A_1 \cap A_2 \cap \dots \cap A_n| \end{aligned}$$

Odnosno, ako je S konačan skup koji sadrži sve skupove A_j , $j = 1, \dots, n$ ($A_1, A_2, \dots, A_n \subseteq S$), onda imamo

$$\begin{aligned} \left| \bigcap_{j=1}^n \overline{A_j} \right| &= |\overline{A_1} \cap \overline{A_2} \cap \dots \cap \overline{A_n}| = |S| - |A_1 \cup A_2 \cup \dots \cup A_n| = \\ &= |S| - \sum_{j=1}^n |A_j| + \sum_{j \neq k} |A_j \cap A_k| - \sum |A_j \cap A_k \cap A_l| + \dots + (-1)^n |A_1 \cap A_2 \cap \dots \cap A_n| \end{aligned}$$

Primjer. U jednoj skupini ljudi se njih 50 bavi košarkom, njih 40 nogometom i 30 šahom. Košarkom i nogometom se bavi njih 17, nogometom i šahom 8, a šahom i košarkom 15. Sa sva ta tri sporta se bavi njih 5.

1. Koliko njih se bavi barem jednim od tih sportova?
2. Ako je u grupi bilo ukupno 105 osoba, koliko njih se nije bavilo niti jendim od ta tri sporta?

Rješenje. Označimo s K , N i \check{S} skupove ljudi iz promatrane skupine koji se bavi odgovarajućim sportom (K - nogomet, N - košarka i \check{S} - šah). Imamo redom: $|K| = 50$, $|N| = 40$, $|\check{S}| = 30$, $|K \cap N| = 17$, $|N \cap \check{S}| = 8$, $|\check{S} \cap K| = 15$ i $|K \cap N \cap \check{S}| = 5$. Formula uključivanja isključivanja nam daje da je broj ljudi koji se bave barem jednim sportom jednak

$$|K \cup N \cup \check{S}| = |K| + |N| + |\check{S}| - |K \cap N| - |\check{S} \cap K| - |N \cap \check{S}| + |K \cap N \cap \check{S}| = 50 + 40 + 30 - 17 - 15 - 8 + 5 = 85.$$

Broj ljudi iz promatrane skupine ljudi (označimo je skupom S) koji se ne bave niti jednim od navedenih sportova je

$$|\overline{K \cup N \cup \check{S}}| = |S| - |K \cup N \cup \check{S}| = 105 - 85 = 20.$$

Formula uključivanja isključivanja (FUI formula) – zadaci

1. Koliko ima prirodnih brojeva manjih od 1001 koji nisu djeljivi ni brojem 5 ni brojem 7?

$$\begin{aligned} G &= \{1, 2, 3, 4, \dots, 999, 1000\} \\ P &= \{5, 10, 15, 20, 25, 30, 35, 40, \dots, 995, 1000\} \subseteq G \\ S &= \{7, 14, 21, 28, 35, 42, \dots, 987, 994\} \subseteq G \end{aligned}$$

Nas zanima $|G| - |P \cup S|$, a prema FUI slijedi $|P \cup S| = |P| + |S| - |P \cap S|$, gdje je $P \cap S$ skup svih brojeva manjih od 1001 djeljivih i brojem 5 i brojem 7, tj. djeljivih brojem 35:

$$P \cap S = \{35, 70, 105, \dots, 980\}$$

Sada imamo: $|P| = \left\lfloor \frac{1000}{5} \right\rfloor = 200$, $|S| = \left\lfloor \frac{1000}{7} \right\rfloor = 142$, $|P \cap S| = \left\lfloor \frac{1000}{35} \right\rfloor = 28$ pa je rješenje:

$$|G| - |P \cup S| = |\overline{P \cap S}| = |\overline{P \cup S}| = |P| + |S| - |P \cap S| = 1000 - 200 - 142 + 28 = 686$$

Korišteno je rješenje sljedećeg zadatka (i to tri puta): koliko ima brojeva manjih od 1001 i djeljivih brojem 7?

To su brojevi 7, 14, 21, 28, 35, ..., 994, gdje je 994 najveći broj koji nije veći od 1000 i koji je djeljiv brojem 7 i vrijedi $994 = 7 \cdot \left\lfloor \frac{1000}{7} \right\rfloor = 7 \cdot 142$. Ako te brojeve podijelimo brojem 7, dobivamo niz 1, 2, 3, 4, 5, ..., 142 pa je odmah vidljivo da ih ima 142.

2. Nakon provedene ankete u nekom razredu od 32 učenika ustanovljeno je da njih 11 ih voli punjenu papriku, 12 sarmu, 10 grah, 4 učenika voli i punjenu papriku i sarmu, 6 i sarmu i grah, 5 i punjenu papriku i grah, a 2 vole sva tri jela.

- (a) Koliko učenika voli barem jedno od tih jela?
- (b) Koliko učenika ne voli niti jedno od tih jela?

Neka je R skup svih učenika tog razreda, F skup učenika koji vole punjenu papriku, S skup učenika koji vole sarmu i G skup učenika koji vole grah.

- (a) Treba izračunati broj $|F \cup S \cup G| = |F| + |S| + |G| - |F \cap S| - |F \cap G| - |S \cap G| + |F \cap S \cap G| = 11 + 12 + 10 - 4 - 5 - 6 + 2 = 20$.
- (b) Treba izračunati broj $|\overline{F \cup S \cup G}| = |R| - |F \cup S \cup G| = 32 - 20 = 12$.

Faktorijeli. Binomni koeficijenti. Permutacije. Kombinacije

Uvedimo neke pojmove iz kombinatorike.

Faktorijeli

Umnožak svih prirodnih brojeva od 1 do n se označava $n!$ i čita se "en faktorijela". Dakle imamo: $n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot (n-1) \cdot n$. Posebno se definira $0! = 1$. Za $n \geq 0$ vrijedi: $(n+1)! = (n+1) \cdot n!$

Primjer 1.

$$1! = 1, 2! = 1 \cdot 2 = 2, 3! = 1 \cdot 2 \cdot 3 = 6, 4! = 1 \cdot 2 \cdot 3 \cdot 4 = 3! \cdot 4 = 24$$

Binomni koeficijenti

Za prirodne brojeve n i k , $0 \leq k \leq n$, binomni koeficijent $\binom{n}{k}$ ("en povrh ka") definiramo sa

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

Posebno se definira $\binom{n}{0} = 1$, za $n \geq 0$.

Primjer 2.

$$\binom{5}{2} = \frac{5 \cdot 4}{1 \cdot 2} = 10, \quad \binom{9}{4} = \frac{9 \cdot 8 \cdot 7 \cdot 6}{1 \cdot 2 \cdot 3 \cdot 4} = 126$$

Vrijedi sljedeća relacija

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

koja nam ubrzava računanje binomnih koeficijanata $\binom{n}{k}$ kod kojih je $n > \frac{n}{2}$:

$$\binom{9}{7} = \binom{9}{9-7} = \binom{9}{2} = \frac{9 \cdot 8}{1 \cdot 2} = 36, \quad \binom{16}{14} = \binom{16}{2} = \frac{16 \cdot 15}{1 \cdot 2} = 120.$$

Za $n \geq 1$ i $0 < k < n$ rekurzivna relacija za binomne koeficijente glasi

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k},$$

uz početne uvjete: za svaki $n \geq 0$ $\binom{n}{0} = \binom{n}{n} = 1$.

Permutacije

Niz od n objekata se zove *permutacija* tih objekata. Npr. *ABCDE*, *ACBED*, *DBCAE* su tri permutacije od pet slova.

Da bi izračunali na koliko načina možemo n različitih objekata poredati u niz (tj. koliko ima permutacija od n različitih elemenata) uočimo da prvi element niza možemo izabrati na n načina. Drugi element niza možemo izabrati na $n-1$ način. Treći element niza možemo izabrati na $n-2$ načina itd. Predzadnji element niza možemo izabrati na dva načina (jer su ostali elementi već izabrani), a zadnji element niza je jednoznačno određen. Slijedi da n različitih elemenata možemo poredati u niz na

$$n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 2 \cdot 1 = n!$$

različitih načina.

Kombinacije

Kombinacija od k elemenata n -teročlanog skupa je bilo koji njegov k -člani podskup. Broj svih kombinacija od k elemenata n -teročlanog skupa je $\binom{n}{k}$.

Među deset igrača jednog šahovskog kluba treba izabrati tim od njih četiri koji će predstavljati klub na turniru. Na koliko načina se može izabrati četveročlani tim ako se na raspolaganju ima deset igrača?

Odgovor: četveročlani tim se može izabrati na $\binom{10}{4} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{1 \cdot 2 \cdot 3 \cdot 4} = 210$ načina.

Primjer 3.

U standardnom špilu od 32 karte se nalaze karte od "sedmice" do "asa" u sve četiri boje (pik, tref, herc i karo). Iz ovog špila karata izvlači se, bez vraćanja, na slučajan način 5 karata. Izračunajte vjerojatnost da su od pet izvučenih karata

- (a) sve karte iste boje,
- (b) dva "dečka", dvije karte vrijednosti manje od "desetke" i pik "as",

Rješenja.

(a) Karata iste boje ima osam, a pet karata iste boje od njih osam možemo dobiti na $\binom{8}{5} = \binom{8}{3} = \frac{8 \cdot 7 \cdot 6}{1 \cdot 2 \cdot 3} = 56$ različitih načina. Boja imamo četiri pa je broj povoljnih ishoda jednak $4 \cdot 56 = 224$. Pet karata iz špila od 32 karte možemo izabrati na $\binom{32}{5} = \frac{32 \cdot 31 \cdot 30 \cdot 29 \cdot 28}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 201376$ različitih načina.

Tražena vjerojatnost je jednaka: $\frac{224}{201376} \approx 0.00111$.

(b) Dva "dečka" od njih četiri možemo izabrati na $\binom{4}{2} = \frac{4 \cdot 3}{1 \cdot 2} = 6$ različitih načina.

Karte manje vrijednosti od "desetke" su "sedmice", "osmice" i "devetke" kojih ukupno ima $4+4+4=12$, dakle dvije karte vrijednosti manje od "desetke" možemo izabrati na $\binom{12}{2} = \frac{12 \cdot 11}{1 \cdot 2} = 66$ načina. Pik "as" je jedna konkretna karta pa je možemo izabrati na jedan način. Pet karata iz špila od 32 karte možemo izabrati na 201376 različitih načina. Tražena vjerojatnost je: $\frac{6 \cdot 66 \cdot 1}{201376} \approx 0.0019665$.

Zadaci

Zadatak 1. Dječak u lijevom džepu ima 7, a u desnem 5 različitih staklenih kuglica.

1. Na koliko načina može izvući jednu kuglicu iz svojih džepova?
Jednu kuglicu iz svojih džepova može izvući na $7 + 5 = 12$ načina.

2. Na koliko načina može izvući dvije kuglice, tako da jednu kuglicu izvuče iz lijevog, a drugu iz desnog džepa?

Iz lijevog džepa jednu kuglicu može izvući na 7, a iz desnog na 5 načina. Izvlačenja su međusobno nezavisna pa dvije kuglica na opisani način može izvući na $7 \cdot 5 = 35$ načina.

3. Na koliko načina može izvući dvije kuglice iz svojih džepova?

Dvije kuglice od njih 12 može izvući na $\binom{12}{2} = \frac{12 \cdot 11}{1 \cdot 2} = 66$ načina.

4. Na koliko načina može izvući po dvije kuglice iz svakog od svojih džepova?

Dvije kuglice iz lijevog džepa može izvući na $\binom{7}{2} = \frac{7 \cdot 6}{1 \cdot 2} = 21$ način.

Dvije kuglice iz desnog džepa može izvući na $\binom{5}{2} = \frac{5 \cdot 4}{1 \cdot 2} = 10$ načina.

Po dvije kuglice iz svakog od svojih džepova može izvući na $21 \cdot 10 = 210$ načina.

Zadatak 2. U jednoj ladici se nalazi 19 košulja, a u drugoj 7 kravata. Na koliko načina se mogu kombinirati jedna košulja i jedna kravata?

Zadatak 3. Koliko nizova od četiri slova možemo dobiti koristeći mala slova hrvatske abecede?

Koristimo princip produkta jer izbor slova na jednom mjestu ne ovisi o izboru slova na drugom. Na svakom od četiri mesta se može naći trideset slova pa je traženi broj jednak

$$30 \cdot 30 \cdot 30 \cdot 30 = 30^4 = 810000.$$

Zadatak 4. Na automobilskim registarskim tablicama se nalazi niz od četiri znamenke i dva velika slova engleske abecede (engleska abeceda ima 26 slova). Koliko je ukupno moguće napraviti takvih registarskih tablica?

Zadatak 5. Šifra za pristup nekoj bazi podataka mora imati barem 6 znakova, ali ne više od 8 (znamenke i mala slova engleske abecede), pri čemu barem jedan znak mora biti znamenka. Koliko je različitih šifri na raspolaganju za pristup toj bazi podataka?

Da bi izračunali traženi broj, primijenit ćemo princip sume na broj šifri od 6, 7 i 8 znakova.

Promotrimo šifre duljine 6. Primjetimo da je broj šifri duljine 6 znakova koje imaju barem jednu znamenku jednak broju svih šifri duljine 6 koje možemo dobiti od slova i znamenaka umanjen za broj svih šifri duljine 6 koje možemo dobiti samo od slova. Broj svih šifri duljine 6 koje možemo dobiti od slova (ima ih 26) i znamenaka (ima ih 10) dobivamo principom produkta: 36^6 .

Broj svih šifri duljine 6 koje možemo dobiti samo od slova je 26^6 .

Dakle, broj šifri duljine 6 koje se mogu koristiti za pristup je: $P_6 = 36^6 - 26^6$.

Analogno dobivamo brojeve šifri duljine 7 i 8: $P_7 = 36^7 - 26^7$, $P_8 = 36^8 - 26^8$.

Traženi broj šifri je: $P_6 + P_7 + P_8$.

Zadatak 6. Izračunajte $6!$, $8!$ i $11!$.

Zadatak 7. Izračunajte $\binom{4}{0}$, $\binom{6}{3}$, $\binom{8}{3}$, $\binom{11}{4}$, $\binom{32}{4}$ i $\binom{52}{5}$.

Zadatak 8. Izračunajte $\binom{4}{3}$, $\binom{7}{5}$, $\binom{8}{6}$, $\binom{11}{11}$, $\binom{30}{25}$ i $\binom{52}{45}$.

Zadatak 9. Ispišite sve permutacije slova X, Y, Z .

Zadatak 10. Na koliko se načina mogu bez ponavljanja poredati znamenke od 1 do 6 u niz?

Zadatak 11. Koliko ima permutacija slova $A, B, C, D, E, F, G, H, I, J$?

Zadatak 12. Koliko ima nizova od osam bitova (bajtova) koji u sebi imaju točno tri jedinice?

Zadatak 13. Imate po jednu kovanicu vrijednosti pet kuna, dvije kune, jedne kune, pedeset lipa, dvadeset lipa i deset lipa. Koliko različitih iznosa možete dobiti pomoću tri kovanice?

Zadatak 14. U špilu od 32 karte se nalaze karte od "sedmice" do "asa" (vrijednosti) u sve četiri boje (pik, tref, herc i karo). Iz ovog špila karata se izvlači na slučajan način, bez vraćanja, 5 karata. Izračunajte vjerojatnost da su od pet izvučenih karata

- (a) točno dvije "dame" i točno tri "sedmice"

Dvije dame od njih četiri možemo izabrati na $\binom{4}{2}$ različita načina. Neovisno o tome, tri sedmice se, od njih četiri, mogu izabrati na $\binom{4}{3} = \binom{4}{1}$ različita načina. Dakle dvije dame i tri sedmice se, prema principu produkta, mogu izabrati na $\binom{4}{2} \binom{4}{1}$ različitih načina.

Ukupan broj načina na koje možemo odabratiti pet karata od njih 32 je $\binom{32}{5}$ pa je tražena vjerojatnost jednaka

$$\frac{\binom{4}{2} \binom{4}{3}}{\binom{32}{5}} = \frac{\binom{4}{2} \binom{4}{1}}{\binom{32}{5}} = \frac{\frac{4 \cdot 3}{1 \cdot 2} \cdot \frac{4}{1}}{\frac{32 \cdot 31 \cdot 30 \cdot 29 \cdot 28}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5}} = \frac{24}{201376} \approx 0.000119$$

- (b) točno tri "asa" i točno tri herca

Ako imam pet karata od kojih su tri "asa" i tri herca, onda je to moguće jedino ako je jedan od "aseva" u hercu. Dakle u tih pet karata imamo: jedan "as" u hercu (jedna mogućnost), još dva asa (od preostala tri - $\binom{3}{2} = 3$ mogućnosti) i još dva herca (od preostalih sedam - $\binom{7}{2} = 21$ mogućnost). Dakle broj načina na koje možemo dobiti takvih pet karata je $1 \cdot 3 \cdot 21 = 63$. Tražena vjerojatnost je $\frac{63}{\binom{32}{5}} = \frac{63}{201376} \approx 0.0003$.

- (c) točno tri iste boje i točno četiri "asa"

Tražena vjerojatnost je nula, jer prisustvo četiriju "aseva" znači da su prisutne sve četiri boje pa je uvjet nemoguće ispuniti.

- (d) točno dva herca i točno tri "desetke" Razlikujemo dva slučaja:

niti jedna "desetka" nije u hercu: Tri "desetke" od kojih niti jedna nije u hercu možemo izabrati na $\binom{3}{3} = 1$ način. Preostale dvije karte moraju biti u hercu, ali niti jedna ne smije biti "desetka" pa ih možemo izabrati na $\binom{7}{2} = \frac{7 \cdot 6}{1 \cdot 2} = 21$ način. Ovakvih pet karata možemo dakle izvući na ukupno 21 način.

jedna "desetka" je u hercu: Ako je jedna "desetka" u hercu, onda preostale dvije možemo izabrati na $\binom{3}{2} = \binom{3}{1} = 3$ načina. Od preostalih dviju karata, jedna mora, a druga ne smije biti u hercu i niti jedna od njih ne smije biti "desetka". Kartu u hercu možemo izabrati na $\binom{7}{1} = 7$ načina, a onu drugu na $32 - 8 - 4 + 1 = 21$ način; od ukupno 32 karte njih 8 su u hercu (-8), njih 4 su "desetke" (-4), a kako ima jedna "desetka" u hercu, nju smo dva puta oduzeli zbog čega je potreban +1 na kraju. Dakle ovakvih pet karata možemo izvući na $3 \cdot 7 \cdot 21 = 441$ načina.

Pomoću principa sume dobivamo da je broj načina na koje možemo dobiti pet karata iz zadatka jednak $21 + 441 = 462$. Tražena vjerojatnost je $\frac{462}{201376} \approx 0.002294$.

(e) četiri karte iste vrijednosti (poker)

Označimo s N_v broj načina da od pet izvučenih karata bude njih četiri iste vrijednosti $v \in \{7, 8, 9, 10, J, Q, K, A\}$. Prema principu sume, broj načina na koji možemo dobiti poker je suma tih vrijednosti: $N_7 + N_8 + N_9 + N_{10} + N_J + N_Q + N_K + N_A$. Izračunajmo N_v za proizvoljnu vrijednost v . Broj načina da izvučemo četiri karte iste vrijednosti je $\binom{4}{4}$, dakle 1. Peta karta može biti bilo koja od preostalih $32 - 4 = 28$ pa nju možemo izvući na $\binom{28}{1} = 28$ načina. Dakle imamo: $N_v = 1 \cdot 28 = 28$. Ukupan broj načina da dobijemo poker je $8 \cdot 28 = 224$ pa je tražena vjerojatnost $\frac{224}{\binom{32}{5}} = \frac{224}{201376} \approx 0.00111$.

(f) pet karata iste boje

Pet karata dane boje (od njih osam) se može izvući na $\binom{8}{5} = \binom{8}{3} = \frac{8 \cdot 7 \cdot 6}{1 \cdot 2 \cdot 3} = 56$ načina. Boja ima četiri, pa se pet karata iste boje može izvući na $4 \cdot 56 = 224$ načina. Tražena vjerojatnost je $\frac{224}{\binom{32}{5}} = \frac{224}{201376} \approx 0.00111$.

Zadatak 15. U šipu karata se nalaze karte od "sedmice" do "asa" (vrijednosti) u sve četiri boje (pik, tref, herc i karo). Iz ovog špila karata izvlačite na slučajan način, bez vraćanja, 5 karata. Izračunajte vjerojatnost da su od pet izvučenih karata

- (a) točno tri "devetke" i točno tri pika ;
- (b) točno dvije "devetke" i točno tri "sedmice" ;
- (c) točno četiri iste boje i pik "as",
- (d) točno dvije "sedmice" i točno dva pika,

- (e) točno dvije iste boje i točno četiri "devetke".
- (f) točno tri "sedmice" i točno tri pika.

Zadatak 16. Imamo dvije kocke (na svakoj stranici po jedan broj od 1 do 6): jednu plavu i jednu crvenu.
Izračunajte vjerojatnost da će u dva bacanja (istovremeno bacamo obje kocke)

- (a) zbroj biti paran oba puta,
- (b) zbroj biti manji od 8 oba puta,
- (c) umnožak biti neparan broj točno jednom,
- (d) umnožak biti prost broj barem jednom.

Zadatak 17. U posudi imamo sedam plavih, tri zelene i pet crnih kuglica. Slučajno izvlačimo četiri kuglice bez vraćanja. Izračunajte vjerojatnost

- (a) da će sve kuglice biti iste boje,
- (b) da će barem dvije kuglice biti plave boje,
- (c) da će biti izvučene kuglice točno dviju boja,
- (d) da će sve kuglice biti različitih boja.

Izračunajte tražene vjerojatnosti ako se nakon svakog izvlačenja kuglica vraća u posudu.

Aksiomi vjerojatnosti. Osnovna svojstva

Neka je Ω neprazan konačan skup i P funkcija koja svakom skupu $A \subseteq \Omega$ pridružuje realan broj. Funkciju P zovemo vjerojatnost ako vrijede sljedeća tri aksioma:

P_1 Za svaki $A \subseteq \Omega$ vrijedi $P(A) \geq 0$. (P je nenegativna funkcija)

P_2 Za svaka dva disjunktna skupa $A, B \subseteq \Omega$ vrijedi $P(A \cup B) = P(A) + P(B)$. (P je aditivna funkcija)

P_3 $P(\Omega) = 1$ (P je normirana funkcija.)

Pogledajmo neke posljedice ovih aksioma i dokažimo neke ih.

1. $P(\emptyset) = 0$

2. Ako je $A \subseteq B \subseteq \Omega$ onda vrijedi

$$P(A) \leq P(B) \quad \text{i} \quad P(B \setminus A) = P(B) - P(A).$$

(monotonost vjerojatnosti)

3. Ako je $A \subseteq \Omega$, onda vrijedi

$$P(A) \leq 1.$$

4. za proizvoljne $A, B \subseteq \Omega$ vrijedi

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

5. Ako je $A \subseteq \Omega$, onda vrijedi

$$P(\overline{A}) = 1 - P(A)$$

Dokaz.

1. Ako uvrstimo $A = \Omega$ i $B = \emptyset$ u aksiom P_2 , dobivamo $P(\Omega) = P(\Omega) + P(\emptyset)$, iz čega direktno slijedi tražena relacija $P(\emptyset) = 0$.
2. Ako je $A \subseteq B$, onda vrijedi $B = A \cup (B \setminus A)$. Događaji A i $B \setminus A$ su disjunktni pa prema aksiomu P_2 vrijedi $P(B) = P(A) + P(B \setminus A)$ iz čega slijedi jedna od traženih relacija. Prema aksiomu P_1 vrijedi $P(B \setminus A) \geq 0$ pa vrijedi i druga tražena relacija.
3. Prema upravo dokazanom i zbog $A \subseteq \Omega$ dobivamo $P(A) \leq P(\Omega)$. Primjena aksioma P_1 dovršava dokaz tražene relacije.
4. Vrijedi $A \cup B = A \cup (B \setminus (A \cap B))$. Sada zbog $A \cap B \subseteq B$ i već dokazanog dobivamo

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

5. Skupovi A i \overline{A} su disjunktni i vrijedi $A \cup \overline{A} = \Omega$ pa imamo

$$1 = P(\Omega) = P(A \cup \overline{A}) = P(A) + P(\overline{A}), \quad \text{tj.} \quad P(\overline{A}) = 1 - P(A).$$

Računanje vjerojatnosti događaja preko vjerojatnosti elementarnih događaja

Problem izračunavanja vjerojatnosti $P(A)$ nekog događaja A (koji je podskup nekog prostora elementarnih događaja Ω) se sada svodi na definiranje funkcije koja zadovoljava aksiome P_1 , P_2 i P_3 (i koju onda nazivamo vjerojatnost). Jedan jednostavan način da se to uradi je da se odrede vjerojatnosti svih elementarnih događaja $\omega \in \Omega$ prema prirodi zadatka. To znači da svakom elementarnom događaju pridružujemo neki broj, kojeg smatramo njegovom vjerojatnosti, pazeći da budu ispunjeni svi aksiomi.

Promotrimo primjer bacanja simetrične kockice. Elementarni događaji su predstavljeni padanjem nekog od brojeva od jedan do šest i vjerojatnosti svih tih elementarnih događaja su međusobne jednakе.

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\})$$

Kako suma svih tih vjerojatnosti odgovara vjerojatnosti $P(\Omega)$ događaja Ω koja prema aksiomu P_3 mora biti jednaka 1, onda vrijedi

$$P(\{1\}) + P(\{2\}) + P(\{3\}) + P(\{4\}) + P(\{5\}) + P(\{6\}) = 1,$$

odnosno

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}.$$

Primijetimo da vjerojatnosti elementarnih događaja moraju biti brojevi veći od ili jednaki 0 i manji od ili jednaki 1, inače ne bi vrijedio aksiom P_1 .

Proizvoljan događaj A je podskup skupa koji se dobije kao unija svih elementarnih događaja:

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

Neka je događaj A *pao je broj veći od dva*. Njegova vjerojatnost $P(A)$ se lagano računa klasičnom formulom a priori: broj povoljnih ishoda je četiri (pao je broj veći od dva, dakle 3 ili 4 ili 5 ili 6), a broj svih ishoda je šest. Dakle

$$P(A) = \frac{4}{6} = \frac{2}{3}.$$

Uočimo da vrijedi

$$P(A) = \frac{4}{6} = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = P(\{3\}) + P(\{4\}) + P(\{5\}) + P(\{6\}).$$

Dakle vjerojatnost događaja je jednaka sumi vjerojatnosti svih elementarnih događaja od kojih se on sastoji, tj. preciznije:

$$P(A) = P(\{e_1\}) + P(\{e_2\}) + \cdots + P(\{e_k\}),$$

pri čemu je $A = \{e_1, e_2, \dots, e_k\}$, tj. općenito vrijedi formula

$$P(A) = \sum_{e \in A} P(\{e\}). \tag{5}$$

Zbog nenegativnosti vjerojatnosti događaja vrijedi $P(A) \geq 0$, što znači da vrijedi aksiom P_1 .

Primijetite da vrijedi i $P(\Omega) = 1$, što znači da je zadovoljen aksiom P_3 , što znači da suma svih vjerojatnosti pridruženih elementarnim događajima mora biti 1.

Preostalo je još da pokažemo da vrijedi i aksiom P_2 , što je napravljeno u sljedećoj cjelini koja se bavi aditivnom formulom.

Aditivna formula

Neka su događaji A i B disjunktni podskupovi prostora elementarnih događaja Ω . Tada se vjerojatnost događaja $A \cup B$ može izračunati pomoću vjerojatnosti događaja A i B , kao što se to može vidjeti iz sljedećeg primjera.

Primjer 1.

Neka je $\Omega = \{1, 2, 3, 4, 5, 6\}$, tj. promatramo rezultate bacanja standardne kockice i

A = pao je broj manji od 3 = $\{1, 2\}$,

B = pao je broj veći od 4 = $\{5, 6\}$.

Zanima nas vjerojatnost događaja $A \cup B = \{1, 2, 5, 6\}$ izražena preko vjerojatnosti $P(A)$ i $P(B)$. Sada imamo:

$$P(A \cup B) = \underbrace{P(\{1\}) + P(\{2\})}_{P(A)} + \underbrace{P(\{5\}) + P(\{6\})}_{P(B)} = P(A) + P(B).$$

Dakle u ovom primjeru smo vidjeli da se vjerojatnost unije dvaju događaja može izračunati preko vjerojatnosti tih dvaju događaja.

Općenito bismo to pokazali ovako: neka su $A \subseteq \Omega$ i $B \subseteq \Omega$ disjunktni događaji (tj. $A \cap B = \emptyset$). Želimo izraziti $P(A \cup B)$ pomoću $P(A)$ i $P(B)$. Iz formule (5) i zbog činjenice da su događaji A i B disjunktni redom slijedi

$$P(A \cup B) = \sum_{e \in A \cup B} P(\{e\}) = \sum_{e \in A} P(\{e\}) + \sum_{e \in B} P(\{e\}) = P(A) + P(B).$$

Dokazali smo formulu

$$P(A \cup B) = P(A) + P(B), \text{ uz uvjet } A \cap B = \emptyset \quad (6)$$

i pokazali da vrijedi aksiom P_2 . Ovime smo ujedno pokazali da klasična formula računanja vjerojatnosti 'a priori' predstavlja funkciju vjerojatnosti i da za nju vrijede sva svojstva koja smo dobili direktno iz aksioma. To znači da u situacijama kada su vjerojatnosti svih elementarnih događaja međusobno jednake možemo koristiti tu klasičnu formulu.

No, željeli bismo formulu za $P(A \cup B)$ koja bi uvijek vrijedila, bez dodatnih uvjeta. Pogledajmo sljedeći primjer.

Primjer 2.

Neka je, kao i u prethodnom primjeru, $\Omega = \{1, 2, 3, 4, 5, 6\}$ i

C = pao je broj veći od 1 i manji od 5 = $\{2, 3, 4\}$,

D = pao je broj veći od 2 i manji od 6 = $\{3, 4, 5\}$.

Zanima nas vjerojatnost događaja $C \cup D = \{2, 3, 4, 5\}$. Sada imamo:

$$P(C \cup D) = \underbrace{P(\{2\}) + P(\{3\}) + P(\{4\})}_{P(C)} + P(\{5\}),$$

iz čega vidimo da traženu vjerojatnost ne možemo dobiti samo preko vjerojatnosti događaja C i D pa ćemo dobiveni izraz malo preuređiti:

$$P(C \cup D) = \underbrace{P(\{2\}) + P(\{3\}) + P(\{4\})}_{P(C)} + \underbrace{P(\{3\}) + P(\{4\}) + P(\{5\})}_{P(D)} - P(\{3\}) - P(\{4\})$$

pri čemu smo $P(C \cup D)$ izrazili preko $P(C)$ i $P(D)$, ali i preko $-P(\{3\}) - P(\{4\})$. Budući je $\{3, 4\} = C \cap D$, slijedi:

$$P(C \cup D) = P(C) + P(D) - P(C \cap D).$$

Ovdje smo dobili da se $P(C \cup D)$ može izračunati pomoću $P(C)$, $P(D)$ i $P(C \cap D)$, tj. vjerojatnost unije dvaju događaja se može izračunati preko vjerojatnosti tih dvaju događaja i vjerojatnosti njihova presjeka (što je također događaj).

Pokažimo da to uvijek vrijedi. Prikažimo $A \cup B$ kao uniju dvaju disjunktnih događaja i primijenimo upravo dobivenu formulu (6). Ako događaji A i B nisu disjunktni, onda B sadrži neke elemente koji se nalaze u A . Događaj $B \setminus A$ se sastoji od svih elemenata iz B koji nisu u A . To znači da su događaji A i $B \setminus A$ disjunktni. Vrijedi $A \cup (B \setminus A) = A \cup B$. Sada možemo primijeniti gornju formulu (6):

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

S druge strane je $B = (A \cap B) \cup (B \setminus A)$ i ta unija je disjunktna pa opet primjenom formule (6) imamo:

$$P(B) = P(A \cap B) + P(B \setminus A) \Rightarrow P(B \setminus A) = P(B) - P(A \cap B).$$

Sada korištenjem zadnjih dviju dobivenih formula dobivamo važnu *aditivnu formulu*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (7)$$

Aditivna formula vrijedi uvijek i to bez ikakvih uvjeta.

Analogna aditivna formula za vjerojatnost unije triju događaja se dobije uz pomoć aditivne formule za uniju dvaju događaja ovako:

$$\begin{aligned} P(A \cup B \cup C) &= P((A \cup B) \cup C) = P(A \cup B) + P(C) - P((A \cup B) \cap C) = \\ &= P(A) + P(B) - P(A \cap B) + P(C) - P((A \cap C) \cup (B \cap C)) = \\ &= P(A) + P(B) + P(C) - P(A \cap B) - (P(A \cap C) + P(B \cap C) - P((A \cap C) \cap (B \cap C))) = \\ &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned}$$

Zadaci

Zadatak 1.

Ante i Branko gađaju metu. Vjerojatnost da Ante pogodi metu u jednom gađanju je $\frac{1}{2}$, a da je Branko pogodi u jednom gađanju je $\frac{2}{3}$. Vjerojatnost da obojica pogode metu je $\frac{1}{3}$. Kolika je vjerojatnost da će barem jedan od njih pogoditi metu u jednom gađanju?

Promotrimo sljedeće događaje:

$A =$ Ante je pogodio metu,

$B =$ Branko je pogodio metu.

Tada je:

$A \cap B =$ i Ante i Branko su pogodili metu,

$A \cup B =$ barem jedan od njih dvojice je pogodio metu.

Prema uvjetima zadatka imamo redom: $P(A) = \frac{1}{2}$, $P(B) = \frac{2}{3}$, $P(A \cap B) = \frac{1}{3}$. Tražimo $P(A \cup B)$.

Primjenom aditivne formule dobivamo:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{2} + \frac{2}{3} - \frac{1}{3} = \frac{5}{6}.$$

Dakle vjerojatnost da će barem jedan od njih pogoditi metu u jednom gađanju je $\frac{5}{6}$.

Zadatak 2.

Baca se standardna igrača kockica sa šest strana i promatraju se sljedeći događaji:

$A =$ pao je prost broj $= \{2, 3, 5\}$ i

$B =$ pao je neparan broj $= \{1, 3, 5\}$.

Izračunajte vjerojatnosti $P(A)$, $P(B)$ i $P(A \cup B)$ klasičnom formulom (4), a nakon toga izračunajte $P(A \cup B)$ pomoću aditivne formule.

Zadatak 3.

Na drugoj godini veleučilišta je provedena anketa o sportovima kojima se bave studenti. Ustanovljeno je da 50% studenata igra košarku, 40% studenata igra nogomet i 30% studenata igra šah. Košarku i nogomet igra 17% studenata, nogomet i šah igra 8% studenata, a šah i košarku igraju 15% studenata. Sva tri sporta igra 5% studenata. Pomoću formule za vjerojatnost unije tri događaja odredite kolika je vjerojatnost da se slučajno odabrani student druge godine veleučilišta bavi barem jednim od ta tri sporta. Koliki postotak studenata se ne bavi niti jednim od ta tri sporta?

Uvjetna vjerojatnost

Vjerojatnost da na standardnoj kockici padne prost broj je, kako znamo, jedna polovina ($P(\{2, 3, 5\}) = \frac{3}{6} = \frac{1}{2}$). Zamislimo sada da smo bacili kockicu i da nam netko kaže da je pao broj veći od tri. Situacija se sada mijenja jer je očito da je vjerojatnost da je pao prost broj jednaka jednoj trećini: broj povoljnih ishoda je jedan (5), a broj svih ishoda je tri (4 ili 5 ili 6).

Općenito će nas zanimati kako izračunati vjerojatnost nekog događaja A uz pretpostavku da je nastupio neki drugi događaj B . Taj drugi događaj "eliminira" neke elementarne događaje iz igre (u našem primjeru su to 1, 2 i 3). Govori se o *uvjetnoj vjerojatnosti* događaja u oznaci: $P(A|B)$ (vjerojatnost događaja A uz uvjet da je nastupio događaj B). Klasičnom formulom za računanje vjerojatnosti lagano računamo uvjetnu vjerojatnost iz gornjeg primjera:

$$P(\{2, 3, 5\}|\{4, 5, 6\}) = P(\{5\}|\{4, 5, 6\}) = \frac{1}{3}.$$

Općenita formula za računanje uvjetne vjerojatnosti je

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad (8)$$

pri čemu, naravno, mora vrijediti $P(B) > 0$. Primjenimo li je na naš primjer za $A = \{2, 3, 5\}$ i $B = \{4, 5, 6\}$, imamo:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{5\})}{P(\{4, 5, 6\})} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}.$$

Primjer 1.

Bacamo dvije kockice – crvenu i plavu. Nadite vjerojatnost da je na jednoj kockici pao broj dva, ako je suma brojeva koji su pali na obje kockice jednaka šest. Također nadite vjerojatnost da je na jednoj kockici pao broj dva.

Rješenje. Neka je

$A =$ na jednoj kockici je pao broj dva

$B =$ suma brojeva koji su pali na obje kockice jednaka je šest

Tražimo $P(A|B)$ i $P(A)$. Rezultat bacanja kockica možemo predstaviti uređenim parom (broj na plavoj kockici, broj na crvenoj kockici). Vjerojatnost svakog elementarnog događaja je $1/36$. Sada imamo: $A = \{(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (1, 2), (3, 2), (4, 2), (5, 2), (6, 2)\}$, $B = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$ i $A \cap B = \{(2, 4), (4, 2)\}$. Tražene vjerojatnosti su:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{36}}{\frac{5}{36}} = \frac{2}{5} = 0.4, \quad P(A) = \frac{11}{36} \approx 0.306.$$

Zadaci

Zadatak 1.

Stranice igraće kocke zalipljene su neprozirnim papirom i to brojevi 1, 2, 3 bijelim, a 4, 5, 6 crvenim. Kocka je bačena i kad se zaustavila na gornjoj stranici bio je crveni papir. Uz pomoć formule za uvjetnu vjerojatnost izračunajte vjerojatnost da je na gornjoj strani bio paran broj.

$$A = \text{pao je paran broj} = \{2, 4, 6\},$$

$$B = \text{na gornjoj strani je bio crveni papir} = \{4, 5, 6\}, P(B) = \frac{3}{6}$$

$$P(A \cap B) = \frac{2}{6}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{3}.$$

Zadatak 2.

Promotrimo sve obitelji s dvoje djece i pretpostavimo da su vjerojatnosti rođenja djevojčice (g) i dječaka (b) međusobno jednake. Takve obitelji možemo opisati parovima iz skupa

$$\Omega = \{(b, b), (b, g), (g, b), (g, g)\},$$

gdje npr. par (b, g) označava obitelj s dječakom i djevojčicom pri čemu je dječak stariji od djevojčice (pa je oznaka dječaka prvi element para). Slučajno odabiranje obitelji s dvoje djece se može predstaviti slučajnim izborom jednog od elemenata skupa Ω pri čemu su svi izbori jednakoj vjerojatni. Slučajno izaberemo obitelj i ustanovljavamo da je u njoj dječak. Kolika je vjerojatnost da je u njoj još jedan dječak (tj. ukupno dva dječaka)?

$$A = \{(b, b), (b, g), (g, b)\} = \text{u obitelji ima jedan dječak}$$

$$B = \{(b, b)\} = \text{u obitelji su dva dječaka}$$

Tražimo $P(B|A)$.

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B)}{P(A)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

Zadatak 3.

Slova riječi "lotos" napisana su na pet kartica koje su stavljene u kutiju. Slučajno se izvlače tri kartice jedna za drugom bez vraćanja. Kolika je vjerojatnost da će biti izvučena riječ "sto"?

$$A_1 = \text{prvo izvučeno slovo je } 's'$$

$$A_2 = \text{drugo izvučeno slovo je } 't'$$

$$A_3 = \text{treće izvučeno slovo je } 'o'$$

$$A = A_1 \cap A_2 \cap A_3$$

$$P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) = \frac{1}{5} \cdot \frac{1}{4} \cdot \frac{2}{3} = \frac{1}{30}$$

Rezultat slijedi iz dvostrukе primjene formule za uvjetnu vjerojatnost (8) prvi put na vjerojatnost $P((A \cap B) \cap C)$ i drugi put na vjerojatnost $P(A \cap B)$:

$$P(A \cap B \cap C) = P((A \cap B) \cap C) = P(A \cap B)P(C|A \cap B) = P(A)P(B|A)P(C|A \cap B).$$

Zadatak 4.

U metu prvo gađa Ante i potom Branko. Vjerojatnost da Ante pogodi metu je 0.6. Ako Ante promaši, Branko pogaća metu s vjerojatnosti 0.3, no ako Ante pogodi, Branko pogaća metu s vjerojatnosti 0.2.

1. Kolika je vjerojatnost da će barem jedan od njih dvojice pogoditi metu?
2. Kolika je vjerojatnost da će točno jedan od njih dvojice pogoditi metu?

Zadatak 5.

Pretpostavite simetričnu varijantu za sve vjerojatnosti koje nisu određene, npr. vjerojatnosti rođenja djevojčice i dječaka su međusobno jednake, tj. 0.5.

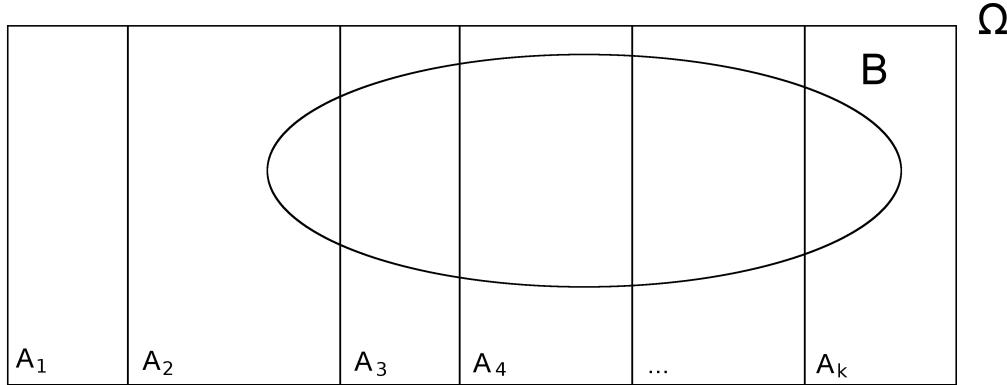
1. Obitelj ima dvoje djece, od kojih je stariji dječak. Izračunajte vjerojatnost da oni imaju dva dječaka.
2. Obitelj ima dvoje djece, od kojih je jedno dječak. Izračunajte vjerojatnost da oni imaju dva dječaka.
3. Obitelj ima dvoje djece, od kojih je jedno dječak koji je rođen u utorak. Izračunajte vjerojatnost da oni imaju dva dječaka.

(Rješenje je na stranici 124.)

Potpuna vjerojatnost

Neka je $\Omega = A_1 \cup A_2 \cup \dots \cup A_k$ disjunktna unija skupova $A_j, j = 1, \dots, k$ (kažemo da je skupovima A_1, A_2, \dots, A_k dana particija skupa Ω) i neka je $B \subseteq \Omega$ podskup skupa Ω (neki događaj). Tada je

$$B = B \cap \Omega = B \cap (A_1 \cup A_2 \cup \dots \cup A_k) = (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_k)$$



Skupovi $B \cap A_j, j = 1, \dots, k$ su međusobno disjunktni pa vrijedi

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \dots + P(B \cap A_k).$$

Upotrijebimo li formulu (8), dobivamo:

$$P(B \cap A_j) = P(A_j \cap B) = P(A_j)P(B|A_j)$$

iz čega slijedi formula potpune (ili totalne) vjerojatnosti:

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_k)P(B|A_k). \quad (9)$$

Primjer 1.

U tvornici se nalaze tri stroja S_1, S_2 i S_3 koji proizvode neku robu. Znamo sljedeće:

- stroj S_1 proizvede 50% robe, od čega je 3% s greškom,
- stroj S_2 proizvede 30% robe, od čega je 4% s greškom i
- stroj S_3 proizvede 20% robe, od čega je 5% s greškom

Nadite vjerojatnost da slučajno izabrani predmet među robom koju su proizveli ta tri stroja ima grešku.

Rješenje.

B = predmet ima grešku, A_j = predmet je proizведен na stroju S_j , $j = 1, 2, 3$. Tražimo $P(B)$. Uvrštavanjem u formulu potpune (ili totalne) vjerojatnosti (9) dobivamo:

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) = 0.5 \cdot 0.03 + 0.3 \cdot 0.04 + 0.2 \cdot 0.05 = 0.037.$$

Bayesova formula

Neka je A_1, A_2, \dots, A_k particija skupa Ω (= unija svih elementarnih događaja) i $B \subseteq \Omega$ neki događaj. Želimo izračunati $P(A_j|B)$, tj. vjerojatnost da je nastupio događaj A_j , ako je nastupio događaj B . (Ako zamislimo da su A_1, A_2, \dots, A_k mogući uzroci događaja B , onda $P(A_j|B)$ možemo shvatiti kao vjerojatnost da je događaj A_j bio uzrok događaja B). Formula (8) nam daje

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)}.$$

Opet iz formule (8) slijedi $P(A_j \cap B) = P(B \cap A_j) = P(A_j)P(B|A_j)$ pa imamo:

$$\frac{P(A_j \cap B)}{P(B)} = \frac{P(A_j)P(B|A_j)}{P(B)}.$$

Sada iskoristimo formulu potpune vjerojatnosti (9) da izračunamo nazivnik i dobivamo Bayesovu formulu:

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_k)P(B|A_k)}. \quad (10)$$

Događaje A_1, A_2, \dots, A_k često zovemo hipoteze (kao moguće uzroke događaja B).

Primjer 1.

Uzmimo iste podatke kao u primjeru na stranici 60. Provjerom je među proizvedenom robom pronađen neispravan primjerak. Za svaki stroj nađite vjerojatnost da ga je on proizveo, tj. nađite vjerojatnosti $P(A_1|B), P(A_2|B)$ i $P(A_3|B)$.

Rješenje.

U prethodnom primjeru smo formulom potpune vjerojatnosti izračunali $P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) = 0.037$. Imamo redom:

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{0.5 \cdot 0.003}{0.037} = \frac{15}{37} \approx 0.405 = 40.5\%$$

$$P(A_2|B) = \frac{P(A_2 \cap B)}{P(B)} = \frac{0.3 \cdot 0.004}{0.037} = \frac{12}{37} \approx 0.325 = 32.5\%$$

$$P(A_3|B) = \frac{P(A_3 \cap B)}{P(B)} = \frac{0.2 \cdot 0.005}{0.037} = \frac{10}{37} \approx 0.27 = 27\%$$

Zadaci

Zadatak 1.

Za studente na trogodišnjem studiju su jedne godine dobiveni sljedeći podaci:

- 30% studenata su na 1. godini i 10% njih ima automobil
- 40% studenata su na 2. godini i 20% njih ima automobil
- 20% studenata su na 3. godini i 40% njih ima automobil
- 10% studenata su apsolventi i 60% njih ima automobil

Slučajno biramo studenta i zanima nas sljedeće:

1. Koja je vjerojatnost da on ima automobil?
2. Ako on ima automobil, koja je vjerojatnost da je on student 3. godine?

Rješenje.

Prvo treba odrediti događaje. Ω je skup svih studenata. $A_1 \subseteq \Omega$ sadrži sve studente s prve godine (dakle $P(A_1)$ je vjerojatnost da je slučajno izabrani student na prvoj godini). Slično se definiraju A_2 , i A_3 . A_4 sadrži sve apsolvente. $B \subseteq \Omega$ sadrži sve studente koji imaju automobil.

1. Trebamo naći $P(B)$. Rješenje nalazimo primjenom formule potpune vjerojatnosti (9).

$$\begin{aligned} P(B) &= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) + P(A_4)P(B|A_4) = \\ &= 0.3 \cdot 0.1 + 0.4 \cdot 0.2 + 0.2 \cdot 0.4 + 0.1 \cdot 0.6 = \\ &= 0.03 + 0.08 + 0.08 + 0.06 = 0.25. \end{aligned}$$

Dakle vjerojatnost da slučajno izabrani student ima automobil je jedna četvrtina, odnosno jedna četvrtina studenata ima automobil.

2. Trebamo naći $P(A_3|B)$. Rješenje nalazimo primjenom Bayesove formule (10). Nazivnik smo već izračunali.

$$P(A_3|B) = \frac{P(A_3)P(B|A_3)}{P(B)} = \frac{0.2 \cdot 0.4}{0.25} = \frac{0.08}{0.25} = \frac{8}{25} = 0.32.$$

Zadatak 2.

Od studenata na jednom veleučilištu 4% muškaraca i 1% žena imaju visinu veću od 190 cm. Nadalje, 60% studenata su žene. Slučajno odabrani student je viši od 190 cm. Kolika je vjerojatnost da je taj student žena?

Rješenje.

Prvo moramo odrediti događaje:

$$A = \{\text{studenti i studentice viši od 190 cm}\}$$

$$M = \{\text{studenti}\}$$

$$W = \{\text{studentice}\}$$

Tražena vjerojatnost je $P(W|A)$ i dobivamo je primjenom Bayesove formule (10):

$$P(W|A) = \frac{P(W)P(A|W)}{P(W)P(A|W) + P(M)P(A|M)} = \frac{0.01 \cdot 0.6}{0.01 \cdot 0.6 + 0.04 \cdot 0.4} = \frac{0.006}{0.022} = \frac{3}{11} \approx 0.27.$$

Zadatak 3.

Tvornice A i B proizvode mobilne telefone bez opcije "foto-slikanja" i s ugrađenom opcijom "foto-slikanja". Tvornica A proizvodi 15% mobilnih telefona bez opcije "foto-slikanja", dok tvornica B

proizvodi 77% mobilnih telefona s opcijom "foto-slikanja". Tvornica B proizvodi tri puta više mobilnih telefona od tvornice A . Mobiteli se u istom omjeru isporučuju istom zastupniku koji isključivo zastupa tvornice A i B . (Taj zastupnik prodaje mobilne telefone samo tih tvornica, a te tvornice ne prodaju svoje mobilne telefone nigdje drugdje.) Izračunajte vjerojatnost da je slučajno izabran kupac mobitela s opcijom "foto-slikanja" kupio mobitel proizveden u tvornici A .

Zadatak 4.

Tvornice E i F proizvode gume za osobna i teretna vozila. Tvornica E proizvodi 45% guma za osobna vozila, dok tvornica F proizvodi 59% guma za osobna vozila. Tvornica E proizvodi tri puta više guma od tvornice F . Gume se u istom omjeru isporučuju istom zastupniku koji isključivo zastupa tvornice E i F . (Taj zastupnik prodaje gume samo tih tvornica, a te tvornice ne prodaju svoje gume nigdje drugdje.) Izračunajte vjerojatnost da je slučajno izabran kupac guma za teretna vozila kupio gume proizvedene u tvornici F .

Nezavisni događaji

Pojednostavljeno (i neprecizno) rečeno, za događaje A i B kažemo da su *nezavisni* ako događanje niti jednoga od njih ne utječe na vjerojatnost događanja drugoga. Preciznije: događaj B ne ovisi o događaju A ako vrijedi

$$P(B) = P(B|A). \quad (11)$$

Iz formule (8) sada slijedi: $P(A \cap B) = P(A)P(B|A) = P(A)P(B)$. Kažemo da su događaji A i B nezavisni ako vrijedi

$$P(A \cap B) = P(A)P(B). \quad (12)$$

Inače kažemo da su zavisni.

Napomena.

Ako vrijedi $P(A \cap B) = P(A)P(B)$ i ako je $P(A)P(B) \neq 0$, onda vrijedi i $P(B) = P(B|A)$ i $P(A) = P(A|B)$:

$$P(A \cap B) = P(B \cap A) = P(A)P(B) \Rightarrow \begin{cases} P(A)P(B) = P(A \cap B) = P(B)P(A|B) \Rightarrow P(A) = P(A|B) \\ P(A)P(B) = P(B \cap A) = P(A)P(B|A) \Rightarrow P(B) = P(B|A) \end{cases} .$$

Napomena.

- Dva disjunktna događaja ne moraju biti nezavisni.

Bacamo standardnu kockicu. A = pao je paran broj, B = pao je neparan broj.

$$P(A) = \frac{1}{2} = P(B), A \cap B = \emptyset \Rightarrow P(A \cap B) = 0 \neq \frac{1}{4} = P(A)P(B)$$

- Dva disjunktna događaja su nezavisna ako jedan od njih ima vjerojatnost nula.

$$0 = P(\emptyset) = P(A \cap B) = P(A)P(B) \Rightarrow P(A) = 0 \text{ ili } P(B) = 0.$$

Primjer 1.

Novčić bacamo tri puta za redom.

$$A = \text{prvi put je pala glava} = \{GGG, GGP, GPG, GPP\}$$

$$B = \text{drugi put je pala glava} = \{GGG, GGP, PGG, PGP\}$$

$$C = \text{točno dva puta za redom je pala glava} = \{GGP, PGG\}$$

Provjerimo nezavisnost po dva događaja od gore navedenih. Prvo pronađimo vjerojatnosti zadanih događaja klasičnom formulom:

$$P(A) = \frac{4}{8} = \frac{1}{2}$$

$$P(B) = \frac{4}{8} = \frac{1}{2}$$

$$P(C) = \frac{2}{8} = \frac{1}{4}$$

Zatim nađimo presjeke po dva događaja i vjerojatnosti novodobivenih događaja:

$$A \cap B = \{GGG, GGP\} \Rightarrow P(A \cap B) = \frac{2}{8} = \frac{1}{4}$$

$$A \cap C = \{GGP\} \Rightarrow P(A \cap C) = \frac{1}{8}$$

$$B \cap C = \{GGP, PGG\} = C \Rightarrow P(B \cap C) = P(C) = \frac{1}{4}$$

Sada nam samo preostaje da provjerimo uvjet iz definicije nezavisnosti:

$$P(A \cap B) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A)P(B) \Rightarrow \text{događaji } A \text{ i } B \text{ su nezavisni}$$

$$P(A \cap C) = \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{4} = P(A)P(C) \Rightarrow \text{događaji } A \text{ i } C \text{ su nezavisni}$$

$$P(B \cap C) = \frac{1}{4} \neq \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{4} = P(B)P(C) \Rightarrow \text{događaji } B \text{ i } C \text{ nisu nezavisni}$$

Zadaci

Zadatak 1.

Dva strijelca, Alen i Branko, gađaju metu. Poznato je da Alen pogađa metu s vjerovatnošću $\frac{1}{2}$, a Branko s vjerovatnošću $\frac{2}{5}$. Obojica gađaju metu. Nađite vjerovatnost da će barem jedan od njih pogoditi metu, ako su njihova gađanja mete nezavisni događaji.

Rješenje.

$A = \text{Alen je pogodio metu}, P(A) = \frac{1}{2}$,

$B = \text{Branko je pogodio metu}, P(B) = \frac{2}{5}$.

Tražimo $P(A \cup B)$ i zbog nezavisnosti znamo da vrijedi $P(A \cap B) = P(A)P(B) = \frac{1}{2} \cdot \frac{2}{5} = \frac{1}{5}$. Korištenjem aditivnog teorema dobivamo:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{2} + \frac{2}{5} - \frac{1}{5} = \frac{7}{10}.$$

Nezavisnost triju događaja se definira ovako:

Tri događaja A , B i C su nezavisna ako vrijedi:

- $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, $P(B \cap C) = P(B)P(C)$ i
- $P(A \cap B \cap C) = P(A)P(B)P(C)$

Zadatak 2.

U uvjetima zadatka 8. na stranici 58 s obiteljima s dvoje djece treba ustanoviti jesu li nezavisni sljedeći događaji:

$A = \text{obitelj ima barem jednu djevojčicu} = \{(g, g), (g, b), (b, g)\}$

$B = \text{obitelj ima djevojčicu i dječaka} = \{(g, b), (b, g)\}$

Zadatak 3.

Dva strijelca nezavisno jedan od drugoga gađaju metu. Vjerovatnost pogotka za jednog je 0.9, a za drugog 0.8. Kolika je vjerovatnost da će meta biti pogodena (tj. da će barem jedan od njih pogoditi metu)?

Zadatak 4.

Neka su A i B nezavisni događaji. Pokažite da su tada nezavisni i događaji A i \bar{B} .

(Rješenje je na stranici 125.)

9 Diskretna slučajna varijabla. Distribucija slučajne varijable.

Slučajna varijabla je formalizacija intuitivnog pojma događaja sa slučajnim ishodima. *Diskretna slučajna varijabla* dakle modelira događaj s konačno ili prebrojivo mnogo ishoda koji imaju svoje vjerojatnosti događanja. Npr. bacamo novčić deset puta. Slučajna varijabla X povezana s ovom situacijom je broj pisama koji su pali. Vrijednosti koje X može poprimiti su iz skupa $\{0, 1, \dots, 10\}$ pa je X diskretna slučajna varijabla.

Primjeri diskretnih slučajnih varijabli:

- broj djece u obitelji,
- broj posjetilaca u kinima petkom navečer,
- broj pacijenata u ambulanti,
- broj neispravnih žarulja u pakiranjima od deset komada,
- broj sekundi proteklo između uzastopnih ulazaka kupaca u trgovinu.

Ishod pokusa ne mora biti broj, npr. kada bacamo novčić, ishodi koje možemo dobiti su "pismo" ili "glava". Mi, međutim, često želimo ishode predstaviti brojevima da bi olakšali analizu pokusa. Slučajna varijabla je funkcija koja svakom ishodu pokusa pridružuje neki broj.

Nas zanimaju samo oni prostori elementarnih događaja koji opisuju slučajne pokuse s konačno ili najviše prebrojivo mnogo ishoda. Neka je Ω konačan ili prebrojiv prostor elementarnih događaja.

Formalno, *diskretna slučajna varijabla* je funkcija sa prostora elementarnih događaja Ω u skup realnih brojeva, tj. funkcija X oblika³

$$X : \Omega \rightarrow \mathbb{R}.$$

Argumenti te funkcije su elementarni događaji prostora elementarnih događaja, a njene vrijednosti su realni brojevi⁴. Tim vrijednostima slučajne varijable su pridružene vjerojatnosti u obliku distribucije (razdiobe).

Primjeri slučajnih varijabli i distribucija:

- Bacanje pravilnog novčića se može opisati slučajnom varijablom $X : \{pismo, glava\} \rightarrow \{0, 1\}$ tako da je

$$X(\omega) = \begin{cases} 0, & \omega = pismo \\ 1, & \omega = glava \end{cases}$$

Pripadajuća distribucija je:

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

³Ako je skup Ω prebrojiv, onda se na funkciju X zadaju dodatni uvjeti, jer nisu sve funkcije oblika $X : \Omega \rightarrow \mathbb{R}$ slučajne varijable. Analiza takvih slučajnih varijabli nadilazi okvire ovog materijala i neće biti provedena. Primjeri koji se obrađuju u ovom materijalu ne zahtijevaju analizu slučajnih varijabli tog tipa.

⁴Intuitivno, slučajna varijabla poprima vrijednosti na slučajan način.

- Gađanje mete u jednom pokušaju se može opisati slučajnom varijablom $X : \{uspjeh, neuspjeh\} \rightarrow \{0, 1\}$ tako da je

$$X(\omega) = \begin{cases} 1, & \omega = uspjeh \\ 0, & \omega = neuspjeh \end{cases}$$

Pripadajuća distribucija bi bila:

$$X \sim \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$$

gdje je p vjerojatnost uspjeha (meta pogodjena), a q vjerojatnost neuspjeha (meta promašena).

- Slučajnom varijablom $X : \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\} \rightarrow \mathbb{R}$ se može opisati (modelirati) rezultat bacanja simetrične kockice sa šest strana pri čemu svaki broj ima istu vjerojatnost pojavljivanja. Možemo definirati $X(\omega_k) = k$. Pripadajuća distribucija je:

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

- Vrijeme (u sekundama ili u minutama) između dolaska dva uzastopna kupca u trgovinu se može opisati slučajnom varijablom $X : \mathbb{R} \rightarrow \mathbb{R}$ tako da definiramo $X(k) = k$. Pripadajuća distribucija bi bila:

$$X \sim \begin{pmatrix} 0 & 1 & 2 & 3 & \dots & k & \dots \\ p_0 & p_1 & p_2 & p_3 & \dots & p_k & \dots \end{pmatrix}$$

gdje je p_k vjerojatnost da je između dva uzastopna dolaska kupaca u trgovinu proteklo k sekundi (minuta,...).

Za općeniti (konačan) prostor događaja $\Omega = \{\omega_1, \omega_2, \omega_3, \dots, \omega_n\}$ i slučajnu varijablu $X : \Omega \rightarrow \mathbb{R}$ bi pripadajuća distribucija izgledala ovako:

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & \dots & x_m \\ p_1 & p_2 & p_3 & p_4 & \dots & p_m \end{pmatrix}$$

pri čemu za svaki $i \in \{1, 2, 3, \dots, n\}$ postoji $j \in \{1, 2, 3, \dots, m\}$ takav da je $X(\omega_i) = x_j$, brojevi x_1, x_2, \dots, x_m su sortirani uzlazno, a brojevi p_1, p_2, \dots, p_m predstavljaju sve vjerojatnosti pa mora vrijediti: $0 \leq p_j \leq 1$, $j = 1, 2, \dots, m$ i $p_1 + p_2 + \dots + p_m = 1$.

Broj p_j je vjerojatnost da slučajna varijabla X poprili vrijednost x_j , tj. $P(X = x_j) = p_j$. Preciznije rečeno, p_j je vjerojatnost događaja kojeg čine svi elementarni događaji koje slučajna varijabla X preslikava u x_j , tj. $p_j = P(\{\omega \in \Omega | X(\omega) = x_j\})$ (što kraće zapisujemo s $P(X = x_j)$).

Primjer 1.

Neka je zadan prostor elementarnih događaja $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$ i neka su p_1, p_2, p_3, p_4, p_5 pripadajuće vjerojatnosti, tj. $P(\omega_j) = p_j$, $j = 1, 2, \dots, 5$. Definirajmo funkciju $X : \Omega \rightarrow \mathbb{R}$ na sljedeći način:

$$X(\omega_1) = X(\omega_3) = x_1, X(\omega_2) = x_3, X(\omega_4) = X(\omega_5) = x_2,$$

gdje su x_1, x_2 i x_3 međusobno različiti realni brojevi. Distribucija slučajne varijable X je sada jednaka:

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 \\ p_1 + p_3 & p_4 + p_5 & p_2 \end{pmatrix}$$

Primjer 2.

Prostorom elementarnih događaja $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$ opisujemo rezultate bacanja simetrične kockice sa šest strana pri čemu svaki broj ima istu vjerojatnost pojavljivanja. Zanima nas hoće li na kockici pasti prost broj, složen broj ili broj koji nije ni složen ni prost. Slučajnu varijablu koja odgovara toj situaciji možemo definirati ovako:

$$X(\omega_1) = 0, \quad X(\omega_2) = X(\omega_3) = X(\omega_5) = -1, \quad X(\omega_4) = X(\omega_6) = 1.$$

Pripadajuća distribucija je: $X \sim \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{3} \end{pmatrix}$.

Neka je $f : \mathbb{R} \rightarrow \mathbb{R}$ proizvoljna funkcija. Tada je $Y = f \circ X = f(X)$ slučajna varijabla s distribucijom

$$Y \sim \begin{pmatrix} f(x_1) & f(x_2) & f(x_3) & f(x_4) & \dots & f(x_m) \\ p_1 & p_2 & p_3 & p_4 & \dots & p_m \end{pmatrix}$$

Može se dogoditi da za dvije različite vrijednosti x_j i x_k vrijedi $f(x_j) = f(x_k)$. Tada moramo udružiti tako dobivena dva ista razreda u jedan čija je pripadna vjerojatnost $p_j + p_k$.

Primjer 3.

Zadana je slučajna varijabla X čija je distribucija $X \sim \begin{pmatrix} -1 & 0 & 1 & 2 & 3 & 4 \\ \frac{1}{12} & \frac{1}{4} & \frac{1}{6} & \frac{1}{12} & \frac{1}{6} & \frac{1}{4} \end{pmatrix}$. Distribucija slučajne varijable $Y = f(X)$, gdje je $f(x) = x + 2$ je

$$Y = f(X) \sim \begin{pmatrix} f(-1) & f(0) & f(1) & f(2) & f(3) & f(4) \\ \frac{1}{12} & \frac{1}{4} & \frac{1}{6} & \frac{1}{12} & \frac{1}{6} & \frac{1}{4} \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{12} & \frac{1}{4} & \frac{1}{6} & \frac{1}{12} & \frac{1}{6} & \frac{1}{4} \end{pmatrix}$$

Distribucija slučajne varijable $Z = g(X)$, gdje je $g(x) = x^2$ je

$$Z \sim \begin{pmatrix} (-1)^2 & 0^2 & 1^2 & 2^2 & 3^2 & 4^2 \\ \frac{1}{12} & \frac{1}{4} & \frac{1}{6} & \frac{1}{12} & \frac{1}{6} & \frac{1}{4} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 4 & 9 & 16 \\ \frac{1}{12} & \frac{1}{4} & \frac{1}{6} & \frac{1}{12} & \frac{1}{6} & \frac{1}{4} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 4 & 9 & 16 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{12} & \frac{1}{6} & \frac{1}{4} \end{pmatrix}$$

Primjetite male izmjene u zapisu distribucije koje su se pojavile zbog toga što su se dva razreda stopila u jedan.

Neka su $X : \Omega \rightarrow \mathbb{R}$ i $Y : \Omega \rightarrow \mathbb{R}$ dvije slučajne varijable s distribucijama

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_m \\ p_1 & p_2 & \dots & p_m \end{pmatrix} \text{ i } Y \sim \begin{pmatrix} y_1 & y_2 & \dots & y_n \\ q_1 & q_2 & \dots & q_n \end{pmatrix}$$

i neka je $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ proizvoljna funkcija dvije varijable. Sada možemo definirati novu slučajnu varijablu $Z : \Omega \rightarrow \mathbb{R}$ sa $Z = f(X, Y)$. Njene vrijednosti su $z_{i,j} = f(x_i, y_j)$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$, a pripadajuće vjerojatnosti su $p_{i,j} = P(X = x_i \text{ i } Y = y_j)$. Ako su slučajne varijable X i Y nezavisne, onda vrijedi $P(X = x_i \text{ i } Y = y_j) = P(X = x_i)P(Y = y_j) = p_i q_j$ i imamo $p_{i,j} = p_i q_j$. Tada je distribucija slučajne varijable Z jednaka

$$Z \sim \begin{pmatrix} z_{1,1} & z_{1,2} & \dots & z_{i,j} & \dots & z_{m,n} \\ p_1 q_1 & p_1 q_2 & \dots & p_i q_j & \dots & p_m q_n \end{pmatrix}.$$

U slučaju da u prvom retku ima istih vrijednosti, potrebno je pripadajuće stupce spojiti u jedan s istom vrijednosti slučajne varijable i sumom pripadajućih vjerojatnosti (kao što je to napravljeno u Primjeru 3 na stranici 67). Ako broevi u prvom retku nisu sortirani, onda se i za to treba pobrinuti.

Primjer 4.

Zadane su distribucije nezavisnih slučajnih varijabli X i Y :

$$X \sim \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{pmatrix} \text{ i } Y \sim \begin{pmatrix} -1 & 0 & 1 & 3 \\ \frac{1}{10} & \frac{1}{5} & \frac{2}{5} & \frac{3}{10} \end{pmatrix}.$$

Odredite distribuciju slučajne varijable $Z = Y - X$.

Vrijednosti $z_{i,j} = y_j - x_i$ i $p_{i,j}$ se najpreglednije zapisuju pomoću dvije tabele:

		Y				q_j				
		-1	0	1	3	p_i	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{10}$
X	0	$\boxed{-1}$	0	1	3	$\frac{1}{2}$	$\boxed{\frac{1}{20}}$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{3}{20}$
	1	-2	$\boxed{-1}$	0	2	$\frac{1}{4}$	$\frac{1}{40}$	$\boxed{\frac{1}{20}}$	$\frac{1}{10}$	$\frac{3}{40}$
	2	-3	-2	$\boxed{-1}$	1	$\frac{1}{4}$	$\frac{1}{40}$	$\frac{1}{20}$	$\boxed{\frac{1}{10}}$	$\frac{3}{40}$

Sada nije teško napisati distribuciju varijable Z :

$$Z \sim \begin{pmatrix} -3 & -2 & \boxed{-1} & 0 & 1 & 2 & 3 \\ \frac{1}{40} & \frac{3}{40} & \boxed{\frac{1}{5}} & \frac{1}{5} & \frac{11}{40} & \frac{3}{40} & \frac{3}{20} \end{pmatrix}.$$

Npr. $P(Z = -1) = P(X = 0 \text{ i } Y = -1) + P(X = 1 \text{ i } Y = 0) + P(X = 2 \text{ i } Y = 1) = \frac{1}{20} + \frac{1}{20} + \frac{1}{10} = \frac{1}{5}$.

Zadaci

Zadatak 1.

Zadana je slučajna varijabla X s distribucijom

$$X \sim \begin{pmatrix} -2 & 1 & 3 & 4 & 5 \\ 7p & p & 6p & p & \frac{2}{17} \end{pmatrix}$$

Nadite pripadajuće vjerojatnosti.

Rješenje.

Suma vjerojatnosti mora biti jednaka 1, dakle mora vrijediti:

$$7p + p + 6p + p + \frac{2}{17} = 1 \Rightarrow 15p = \frac{15}{17} \Rightarrow p = \frac{1}{17}$$

Sve vjerojatnosti su brojevi iz intervala $[0, 1]$ pa je zadana distribucija jednaka

$$X \sim \begin{pmatrix} -2 & 1 & 3 & 4 & 5 \\ \frac{7}{17} & \frac{1}{17} & \frac{6}{17} & \frac{1}{17} & \frac{2}{17} \end{pmatrix}$$

Zadatak 2.

Zadana je slučajna varijabla X s distribucijom

$$X \sim \begin{pmatrix} -2 & 0 & 2 & 3 & 4 \\ 4p & 2p & 2p & 5p & 2p \end{pmatrix}$$

Nadite pripadajuće vjerojatnosti.

Zadatak 3.

Odredite distribuciju slučajne varijable D koja opisuje rezultate bacanja nesimetrične kockice sa šest strana koja "favorizira" veće brojeve tako da za svaki broj $a \in \{4, 5, 6\}$ i svaki broj $b \in \{1, 2, 3\}$ vrijedi da je vjerojatnost da padne broj a dva puta veća od vjerojatnosti da padne broj b .

Zadatak 4.

Zadana je slučajna varijabla X s distribucijom

$$X \sim \begin{pmatrix} -2 & -1 & 0 & 1 & 3 \\ \frac{1}{8} & \frac{1}{16} & \frac{3}{8} & \frac{1}{8} & \frac{5}{16} \end{pmatrix}$$

Nadite distribucije slučajnih varijabli $Y = f(X)$, $Z = g(X)$ i $W = h(X)$ ako je $f(x) = x - 3$, $g(x) = x^3$ i $h(x) = (x - 4)^2$.

Zadatak 5.

Zadane su distribucije nezavisnih slučajnih varijabli X i Y :

$$X \sim \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{pmatrix} \text{ i } Y \sim \begin{pmatrix} -1 & 0 & 1 \\ \frac{2}{7} & \frac{4}{7} & \frac{1}{7} \end{pmatrix}.$$

Odredite distribuciju slučajne varijable $Z = X \cdot Y$.

10 Funkcija vjerojatnosti. Funkcija distribucije vjerojatnosti

Neka je zadana distribucija

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & \dots & x_m \\ p_1 & p_2 & p_3 & p_4 & \dots & p_m \end{pmatrix}.$$

pripadajuću *funkciju vjerojatnosti* $f : \mathbb{R} \rightarrow \mathbb{R}$ definiramo sa

$$f(x) = \begin{cases} p_k, & x = x_k \in \{x_1, x_2, \dots, x_m\}, \\ 0, & x \notin \{x_1, x_2, \dots, x_m\}. \end{cases}$$

Pripadajuću *funkciju distribucije vjerojatnosti* $F : \mathbb{R} \rightarrow \mathbb{R}$ definiramo sa

$$F(x) = \begin{cases} 0, & x < x_1, \\ \sum_{j=1}^k p_j, & x_k \leq x < x_{k+1}, \quad k = 1, 2, \dots, m-1, \\ 1, & x_m \leq x. \end{cases}$$

Napomena.

Pomoću funkcije vjerojatnosti f i funkcije distribucije vjerojatnosti F diskretne slučajne varijable X se vjerojatnost događaja $\{\omega | X(\omega) \leq b\}$, ili kraće $\{X \leq b\}$, može zapisati ovako:

$$P(X \leq b) = \sum_{X(\omega_j) \leq b} P(\omega_j) = \sum_{\substack{k \in \{1, 2, \dots, m\} \\ x_k \leq b}} p_k = F(b).$$

Vrijede i sljedeće formule:

$$\begin{aligned} P(a < X \leq b) &= \sum_{X(\omega_j) \leq b} P(\omega_j) - \sum_{X(\omega_j) \leq a} P(\omega_j) = F(b) - F(a), \\ P(a \leq X \leq b) &= \sum_{\substack{k \in \{1, 2, \dots, m\} \\ a \leq x_k \leq b}} p_k \\ P(X = a) &= f(a) = \begin{cases} p_j, & a = x_j, j \in \{1, 2, \dots, m\} \\ 0, & \text{inače} \end{cases} \end{aligned}$$

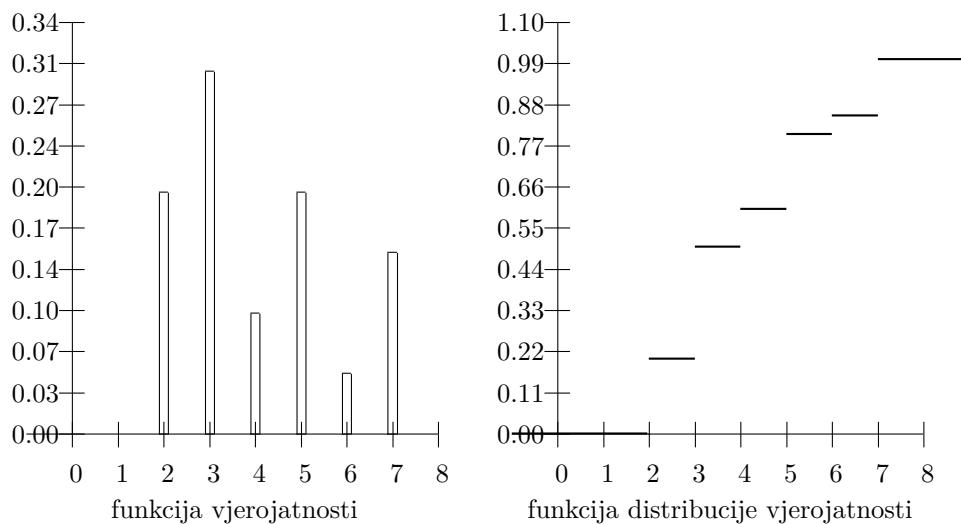
Zadatak 1.

Uz oznake s ove stranice, napišite kako biste izračunali sljedeće vjerojatnosti:

1. $P(X > a)$,
2. $P(X \geq a)$ i
3. $P(a < X < b)$.

Primjer 1.

Zadana je distribucija $\begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 \\ 0.2 & 0.3 & 0.1 & 0.2 & 0.05 & 0.15 \end{pmatrix}$. Skicirajte pripadajuće grafove funkcije vjerojatnosti i funkcije distribucije vjerojatnosti.

Rješenje.**Zadatak 2.**

Skicirajte pripadajuće grafove funkcije vjerojatnosti i funkcije distribucije vjerojatnosti za distribuciju

$$\begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 \\ 0.3 & 0.2 & 0.2 & 0.1 & 0.15 & 0.05 \end{pmatrix}.$$

Zadatak 3.

Neka je X diskretna slučajna varijabla s distribucijom

$$X \sim \begin{pmatrix} -2 & 0 & 1 & 3 & 5 & 6 \\ \frac{1}{6} & \frac{5}{18} & \frac{2}{9} & \frac{1}{18} & \frac{1}{9} & \frac{1}{6} \end{pmatrix}.$$

Izračunajte

1. $P(X = 1)$,
2. $P(X = 4.5)$,
3. $P(-2 < X \leq 3)$ i
4. $P(0 \leq X \leq 5)$.

Rješenje.

1. $P(X = 1) = f(1) = \frac{2}{9},$
2. $P(X = 4.5) = f(4.5) = 0,$
3. $P(-2 < X \leq 3) = \frac{5}{18} + \frac{2}{9} + \frac{1}{18} = \frac{5}{9}$
4. $P(0 \leq X \leq 5) = \frac{5}{18} + \frac{2}{9} + \frac{1}{18} + \frac{1}{9} = \frac{2}{3}.$

11 Očekivanje, varijanca i standardna devijacija diskretne slučajne varijable

Neka je X diskretna slučajna varijabla s distribucijom

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & \dots & x_m \\ p_1 & p_2 & p_3 & p_4 & \dots & p_m \end{pmatrix}$$

Očekivanje diskretne slučajne varijable X , u oznaci $E(X)$, definira se sa

$$E(X) = \sum_{k=1}^m x_k p_k = x_1 p_1 + x_2 p_2 + \dots + x_m p_m.$$

Primjer 1.

Neka je X diskretna slučajna varijabla s distribucijom

$$X \sim \begin{pmatrix} 0 & 1 & 3 & 4 & 6 \\ \frac{7}{30} & \frac{7}{30} & \frac{1}{10} & \frac{7}{30} & \frac{1}{5} \end{pmatrix}$$

Očekivanje slučajne varijable X je jednako:

$$E(X) = \frac{7}{30} \cdot 0 + \frac{7}{30} \cdot 1 + \frac{1}{10} \cdot 3 + \frac{7}{30} \cdot 4 + \frac{1}{5} \cdot 6 = \frac{8}{3}$$

Varianca diskretne slučajne varijable X , u oznaci $Var(X)$, definira se sa

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

tj. varianca $Var(X)$ se može shvatiti kao očekivanje slučajne varijable $Y = f(X)$, gdje je $f(x) = (x - E(X))^2$. Za računanje se koristi jednostavnija formula koja se može izvesti iz definicije:

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

Primjer 1. (nastavak)

U gornjem primjeru bismo imali:

$$\begin{aligned} X^2 &\sim \begin{pmatrix} 0 & 1 & 9 & 16 & 36 \\ \frac{7}{30} & \frac{7}{30} & \frac{1}{10} & \frac{7}{30} & \frac{1}{5} \end{pmatrix} \\ E(X^2) &= \frac{7}{30} \cdot 0 + \frac{7}{30} \cdot 1 + \frac{1}{10} \cdot 9 + \frac{7}{30} \cdot 16 + \frac{1}{5} \cdot 36 = \frac{181}{15} \end{aligned}$$

Sada imamo:

$$Var(X) = E(X^2) - E(X)^2 = \frac{181}{15} - \left(\frac{8}{3}\right)^2 = \frac{223}{45}.$$

Standardna devijacija diskretne slučajne varijable je drugi korijen njezine varijance:

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

Primjer 1. (nastavak)

U našem primjeru imamo: $\sigma(X) = \sqrt{\frac{223}{45}} \approx 2.2261$.

Neka je $\mu = E(X) = \frac{8}{3}$ i $\sigma = \sigma(X) = \sqrt{\frac{223}{45}}$. Izračunajte $P(\mu - \sigma \leq X \leq \mu + \sigma)$.

$$\begin{aligned} P(\mu - \sigma \leq X \leq \mu + \sigma) &= P\left(\frac{8}{3} - \sqrt{\frac{223}{45}} \leq X \leq \frac{8}{3} + \sqrt{\frac{223}{45}}\right) \approx P(0.44 \leq X \leq 4.89) = \\ &= \sum_{\substack{k \in \{1, 2, 3, 4, 5\} \\ 0.44 \leq x_k \leq 4.89}} p_k = p_2 + p_3 + p_4 = \frac{7}{30} + \frac{1}{10} + \frac{7}{30} = \frac{17}{30} \approx 0.5667 \end{aligned}$$

Primjer 2.

Očekivanje slučajne varijable X s distribucijom $X \sim \begin{pmatrix} -1 & 0 & 1 & 2 & a \\ 2p & 2p & \frac{5}{21} & p & \frac{1}{21} \end{pmatrix}$ je $E(X) = \frac{3}{7}$.

Odredite nepoznanice a i p te izračunajte varijancu i standardnu devijaciju slučajne varijable X .

Izračunajte $P(X = 2)$, $P(X = 3)$, $P(-1 \leq X < 1)$, $P(0 \leq X \leq 2)$ i $P(\mu - \sigma \leq X \leq \mu + \sigma)$, gdje je $\mu = E(X)$ i $\sigma = \sigma(X)$.

Rješenje.

Za vjerojatnosti u drugom retku distribucije mora vrijediti da im je suma jednaka 1 pa imamo:

$$2p + 2p + \frac{5}{21} + p + \frac{1}{21} = 1 \Rightarrow p = \frac{1}{7}.$$

Sada imamo sve vjerojatnosti u distribuciji: $X \sim \begin{pmatrix} -1 & 0 & 1 & 2 & a \\ \frac{2}{7} & \frac{2}{7} & \frac{5}{21} & \frac{1}{7} & \frac{1}{21} \end{pmatrix}$.

Nepoznanicu a sada lagano dobijemo pomoću poznatog očekivanja:

$$E(X) = -1 \cdot \frac{2}{7} + 0 \cdot \frac{2}{7} + 1 \cdot \frac{5}{21} + 2 \cdot \frac{1}{7} + a \cdot \frac{1}{21} = \frac{3}{7} \Rightarrow a = 4$$

Sada smo u potpunosti odredili distribuciju slučajne varijable: $X \sim \begin{pmatrix} -1 & 0 & 1 & 2 & 4 \\ \frac{2}{7} & \frac{2}{7} & \frac{5}{21} & \frac{1}{7} & \frac{1}{21} \end{pmatrix}$.

Varijanca i standardna devijacija se sada lagano računaju:

$$E(X^2) = (-1)^2 \cdot \frac{2}{7} + 0^2 \cdot \frac{2}{7} + 1^2 \cdot \frac{5}{21} + 2^2 \cdot \frac{1}{7} + 4^2 \cdot \frac{1}{21} = \frac{13}{7},$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{13}{7} - \left(\frac{3}{7}\right)^2 = \frac{82}{49} \approx 1.67347, \sigma(X) = \sqrt{Var(X)} = \sqrt{\frac{82}{49}} \approx 1.29363.$$

$$P(X = 2) = \frac{1}{7}, P(X = 3) = 0,$$

$$P(-1 \leq X < 1) = \frac{2}{7} + \frac{2}{7} = \frac{4}{7}, P(0 \leq X \leq 2) = \frac{2}{7} + \frac{5}{21} + \frac{1}{7} = \frac{2}{3},$$

$$\mu = \frac{3}{7} \approx 0.43, \sigma \approx 1.29 \Rightarrow P(\mu - \sigma \leq X \leq \mu + \sigma) = P(-0.86 \leq X \leq 1.72) = \frac{2}{7} + \frac{5}{21} = \frac{11}{21}$$

Zadaci

Zadatak 1.

Zadana je slučajna varijabla X s distribucijom

$$X \sim \begin{pmatrix} 0 & 1 & 2 & 3 & 5 \\ \frac{3}{19} & \frac{2}{19} & \frac{5}{19} & \frac{5}{19} & \frac{4}{19} \end{pmatrix}$$

Izračunajte očekivanje, varijancu i standardnu devijaciju slučajne varijable X .

(Rješenje: $E(X) = \frac{47}{19}$, $Var(X) = \frac{964}{361}$)

Zadatak 2.

Zadana je slučajna varijabla X s distribucijom

$$X \sim \begin{pmatrix} -2 & -1 & 1 & 2 & 3 & 5 & 7 & 8 & 9 \\ \frac{3}{28} & \frac{5}{28} & \frac{1}{28} & \frac{1}{14} & \frac{1}{7} & \frac{1}{28} & \frac{3}{28} & \frac{1}{14} & \frac{1}{4} \end{pmatrix}$$

Izračunajte očekivanje, varijancu i standardnu devijaciju slučajne varijable X .

(Rješenje: $E(X) = \frac{111}{28}$, $Var(X) = \frac{13691}{784}$)

Zadatak 3.

Zadana je slučajna varijabla X s distribucijom

$$X \sim \begin{pmatrix} -2 & 0 & 1 & 3 & 4 \\ 3p & 3p & 7p & \frac{1}{11} & 7p \end{pmatrix}$$

Izračunajte pripadne vjerojatnosti i zatim izračunajte očekivanje, varijancu i standardnu devijaciju slučajne varijable X . (Rješenje: $p = \frac{1}{22}$, $E(X) = \frac{35}{22}$, $Var(X) = \frac{2053}{484}$)

Zadatak 4.

Zadana je slučajna varijabla X s distribucijom

$$X \sim \begin{pmatrix} -2 & -1 & 0 & 3 & 4 \\ 3p & \frac{1}{4} & 3p & p & \frac{1}{2}p \end{pmatrix}$$

Izračunajte pripadne vjerojatnosti i zatim izračunajte očekivanje, varijancu i standardnu devijaciju slučajne varijable X . (Rješenje: $p = \frac{1}{10}$, $E(X) = -\frac{7}{20}$, $Var(X) = \frac{1211}{400}$)

Zadatak 5.

Zadana je slučajna varijabla X s očekivanjem $E(X) = \frac{111}{28}$ i distribucijom

$$X \sim \begin{pmatrix} -1 & 2 & 3 & a & 8 \\ 2p & 9p & \frac{1}{7} & \frac{1}{4} & 6p \end{pmatrix}$$

Izračunajte nepoznatu vrijednost a i pripadajuće vjerojatnosti te zatim izračunajte varijancu i standardnu devijaciju slučajne varijable X . (Rješenje: $p = \frac{1}{28}$, $a = 5$, $Var(X) = \frac{5403}{784}$)

Zadatak 6.

Očekivanje slučajne varijable X s distribucijom

$$X \sim \begin{pmatrix} -2 & -1 & 0 & 1 & a \\ p & p & 2p & \frac{1}{10} & \frac{1}{10} \end{pmatrix}$$

je $E(X) = -\frac{3}{10}$. Odredite nepoznate a i p te izračunajte varijancu slučajne varijable X . (Rješenje: $p = \frac{1}{5}$, $a = 2$, $Var(X) = 1.41$)

12 Diskretna uniformna distribucija

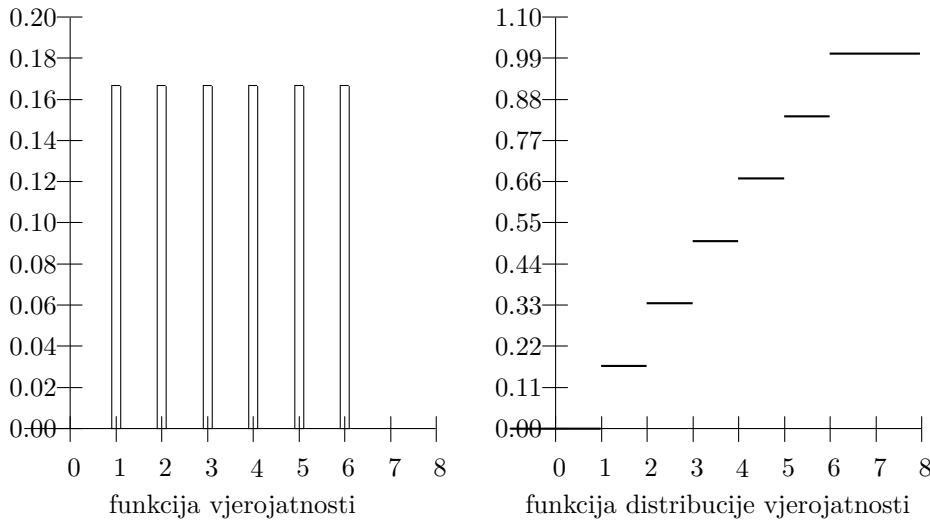
Diskretna uniformna distribucija slučajne varijable X je zadana s

$$X \sim \left(\begin{array}{ccccc} x_1 & x_2 & x_3 & \dots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{array} \right).$$

Jednostavno rečeno, diskretna uniformna distribucija je karakterizirana konačnim skupom vrijednosti koje su sve jednakovjerojatne:

$$P(X = 1) = P(X = 2) = P(X = 3) = \dots = P(X = n) = \frac{1}{n}.$$

Grafovi funkcija vjerojatnosti i funkcije distribucije vjerojatnosti diskretne uniformne distribucije (bacanje simetrične kockice):



Jednostavan način za dobivanje brojeva s uniformnom distribucijom je preko bacanja simetrične igraće kockice⁵ koja bacanjem daje jednu od vrijednosti 1, 2, 3, 4, 5, 6, svaku od njih s istom vjerojatnosti jednakoj $\frac{1}{6}$. Rezultati bacanja pravilnog novčića, brojevi koji se dobivaju na standardnom ruletu kao i karte koje se dobiju iz dobro promiješanog špila karata također imaju diskretnu uniformnu distribuciju.

Očekivanje slučajne varijable s diskretnom uniformnom distribucijom je jednako aritmetičkoj sredini svih njenih vrijednosti:

$$E(X) = \sum_{k=1}^n \frac{1}{n} x_k = \frac{1}{n} x_1 + \frac{1}{n} x_2 + \frac{1}{n} x_3 + \dots + \frac{1}{n} x_n = \frac{1}{n} (x_1 + x_2 + x_3 + \dots + x_n) = \bar{x}.$$

⁵U RPG igrama (naročito u D&D-u) se kao izvor uniformno distribuiranih brojeva koriste sva Platonova tijela, tj. svi pravilni poliedri – tetraedar, heksaedar (kocka), oktaedar, dodekaedar i ikosaedar. Oznake za te "kockice" (redom $d4$, $d6$, $d8$, $d12$, $d20$) možemo shvatiti kao slučajne varijable s pripadajućom diskretnom uniformnom distribucijom. Koriste se i "kockice" $d2$, $d3$, $d10$ i $d100$ koje se ne mogu predstaviti Platonovim tijelima, ali ih je moguće ili simulirati pomoću njih ($d2$ i $d3$ pomoću $d4$ i $d6$, s tim da se $d2$ može simulirati i bacanjem novčića) ili pomoću polupravilnih poliedara ($d10$). "Kockica" $d100$ se može predstaviti pomoću dvije $d10$ "kockice", pri čemu jedna određuje znamenku jedinica, a druga znamenku desetica.

Varijanca slučajne varijable s diskretnom uniformnom distribucijom je jednaka varijanci svih njenih vrijednosti, analogna tvrdnja vrijedi i za standardnu devijaciju:

$$Var(X) = E(X^2) - E(X)^2 = \bar{x^2} - \bar{x}^2, \sigma(X) = \sqrt{Var(X)}.$$

Primjer 1.

Odredimo distribuciju slučajne varijable X koja predstavlja sumu brojeva koji se pojave nakon bacanja dviju simetričnih kockica, jedne s osam strana i brojevima od 1 do 8 na njima ($d8$) i jedne sa četiri strane i brojevima od 1 do 4 na njima ($d4$) (dakle $X = d8 + d4$).



Rješenje.

Distribucije slučajnih varijabli $d4$ i $d8$ su redom:

$$d4 \sim \left(\begin{array}{cccc} 1 & 2 & 3 & 4 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array} \right), \quad d8 \sim \left(\begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \frac{1}{8} & \frac{1}{8} \end{array} \right).$$

Vjerojatnost da npr. na kockicama padnu brojevi 1 i 7 je $\frac{1}{4} \cdot \frac{1}{8} = \frac{1}{32}$, a vjerojatnost da je suma brojeva koji padnu na te dvije kockice jednaka 8 je $4 \cdot \frac{1}{32} = \frac{1}{8}$, jer se suma 8 može dobiti na četiri načina ($8 = 1 + 7 = 2 + 6 = 3 + 5 = 4 + 4$). Traženu distribuciju možemo lagano dobiti iz sljedeće tabele u kojoj su popisane sve sume koje možemo dobiti bacanjem ovih dviju kockica:

$d4 \setminus d8$	1	2	3	4	5	6	7	8
1	2	3	4	5	6	7	8	9
2	3	4	5	6	7	8	9	10
3	4	5	6	7	8	9	10	11
4	5	6	7	8	9	10	11	12

Broj pojavljivanja neke sume u gornjoj tabeli pomnožen s gore dobivenom vjerojatnosti $\frac{1}{32}$ daje pripadajuću vjerojatnost pojavljivanja te sume. Distribuciju slučajne varijable X sada nije problem napisati:

$$X \sim \left(\begin{array}{cccccccccc} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{32} & \frac{2}{32} & \frac{3}{32} & \frac{4}{32} & \frac{4}{32} & \frac{4}{32} & \frac{4}{32} & \frac{4}{32} & \frac{3}{32} & \frac{2}{32} & \frac{1}{32} \end{array} \right).$$

Primjer 2.

U igri D&D igrači vode svoje likove kroz priču u kojima se nerijetko moraju boriti. Tomislav vodi svog lika, hrabrog Sir Robina i u slučaju borbe se mora opredijeliti za jednu od sljedeće dvije mogućnosti:

- borba s dvoručnim mačem kojeg drži objema rukama (i s kojim je iznimno spretan, što mu daje bonus 4 u napadu) ili
- s dva manja mača (s kojima nije toliko vješt pa je bonus 0), ali s kojima može napadati istovremeno, dakle može odjednom napasti dva puta.

Napad počinje bacanjem "kockice" s dvadeset strana, $d20$ (vidi fusnotu ⁵ na stranici 76); broju koji padne se dodaje bonus i ako je rezultat veći od ili jednak 15 onda je napad uspio⁶.

⁶Ovo je pojednostavljena verzija preuzeta iz pravila za DnD, verzija 3.5; naime, ako na kockici $d20$ padne recimo 19 ili 20, napad je izuzetno dobro uspio što znači da bi napadač mogao biti uspešniji u tom napadu nego što je to uobičajeno. Radi jednostavnosti, taj slučaj se neće uzeti u obzir.

Kolika je vjerojatnost da napad Sir Robina uspije ako on koristi dvoručni mač, a kolika ako koristi dva manja mača?

- Ako koristi dvoručni mač, za uspješan napad mora zajedno s bonusom (4) dobiti barem 15 na kockici $d20$, tj. na toj kockici mora dobiti barem 11. Od 20 mogućnosti njih 10 je povoljno pa je vjerojatnost uspjeha jednaka $\frac{10}{20} = \frac{1}{2}$.
- Ako uzme dva manja mača, za uspješan napad mora dobiti barem 15 na kockici $d20$. Od 20 mogućnosti njih 6 je povoljno pa je vjerojatnost uspjeha jednaka $\frac{6}{20} = \frac{3}{10}$.

U slučaju uspješnog napada, napadač je ozlijedio protivnika i onda se određuje koliko jedinica zdravlja je protivnik izgubio tim napadom.

- Ako je napad hrabrog Sir Robina dvoručnim mačem uspio, onda napadnuti gubi $d8+2$ jedinice zdravlja, tj. Tomislav baca $d8$ kockicu i broju koji padne doda 2.
- Ako je napad hrabrog Sir Robina s dva manja mača uspio, onda napadnuti gubi $2d6 + 1$ jedinicu zdravlja, tj. Tomislav baca dvije $d6$ kockice i sumi brojeva koji padnu doda 1.

Očekivani broj jedinica zdravlja koje će protivnik izgubiti u slučaju jednog uspješnog napada je

- za napad dvoručnim mačem jednak $E(d8 + 2) = 6.5$,
- za napad s dva manja mača jednak $E(2d6 + 1) = 8$,

što se lagano dobije iz pripadajućih distribucija za $d8$ i $2d6$ (obje su dane u ovom poglavlju).

Da bismo dobili očekivani broj jedinica zdravlja koje će protivnik izgubiti u jednom napadu, moramo dobivena očekivanja pomnožiti⁷ s vjerojatnostima da napad uspije:

- za napad dvoručnim mačem imamo $\frac{1}{2} \cdot 6.5 = 3.25$,
- za napad s dva manja mača imamo $\frac{3}{10} \cdot 8 = 2.4$.

Zaključujemo da se u duljim borbama hrabrom Sir Robinu daleko više isplati boriti dvoručnim mačem, jer će u prosjeku protivnik izgubiti više jedinica zdravlja, što vodi bržoj i sigurnijoj pobjedi u borbi.

U nekim situacijama se međutim ne isplati uzeti oružje koje garantira veći očekivani broj jedinica zdravlja koje će protivnik izgubiti u jednom napadu. Npr. moguća je situacija gdje je u borbi za pobjedu nužno da u najviše tri napada protivnik izgubi više od 30 jedinica zdravlja, inače hrabri Sir Robin sigurno gubi borbu. Tada sigurno neće uzeti dvoručni mač, jer je maksimalni broj jedinica zdravlja koje protivnik može izgubiti tri napada 30 što nije dovoljno i vodi u siguran poraz. S druge strane, ako uzme dva manja mača, maksimalni broj jedinica zdravlja koje protivnik može izgubiti u jednom napadu je 13 (dvije šestice na dvije kockice $d6$ plus 1) pa postoji mala vjerojatnost (izračunajte je!) koja je ipak veća od nule da hrabri Sir Robin pobijedi u takvoj borbi.

⁷Da ta dva broja trebamo pomnožiti pa da dobijemo ono što nas zanima je posljedica jednog kraćeg vjerojatnosnog računa koji je ovdje izostavljen.

Zadaci

Zadatak 1.

Napišite distribuciju uniformne slučajne varijable koja može poprimiti sve cjelobrojne vrijednosti od -3 do 7 . (Rješenje je na stranici 125.)

Zadatak 2.

Pomoću uniformne distribucije (iako je najjednostavnija) se mogu dobiti druge složenije distribucije. Odredite distribuciju slučajne varijable X kojoj su vrijednosti zbroj brojeva koji padnu na dvije bačene standardne kockice (u igri DnD se takva slučajna varijabla označava s $2d6$). Izračunajte $E(X)$, $Var(X)$ i standardnu devijaciju slučajne varijable X . (Rješenje je na stranici 125.)

Zadatak 3.

Odredite distribuciju slučajne varijable X koja modelira bacanje dviju simetričnih kockica, jedne standardne sa šest strana i brojevima od 1 do 6 na njima i jedne sa četiri strane i brojevima od 1 do 4 na njima te uzimanjem sume brojeva koji na njima padnu (dakle $X = d6 + d4$) .

Zadatak 4.

Odredite distribuciju slučajne varijable X koja modelira bacanje dviju simetričnih kockica, jedne standardne sa šest strana i brojevima od 1 do 6 na njima i jedne sa četiri strane i brojevima od 1 do 4 na njima i uzimanjem razlike brojeva koji na njima padnu (dakle $X = d6 - d4$) .

Zadatak 5.

Diskretna slučajna varijabla X ima uniformnu distribuciju $\left(\begin{array}{cccccc} 1 & 2 & 3 & \dots & n-1 & n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} & \frac{1}{n} \end{array} \right)$. Znamo da vrijedi $P(2 \leq X < 6) = \frac{1}{3}$. Izračunajte n . (Rješenje je na stranici 126.)

Zadatak 6.

Zadana je diskretna slučajna varijabla X s distribucijom $\left(\begin{array}{cccccc} 1 & 2 & 3 & \dots & n-1 & n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} & \frac{1}{n} \end{array} \right)$. Izračunajte $E(X)$ i $Var(X)$. Pomoću dobivenog rezultata izračunajte $E(d6)$ i $Var(d6)$. (Rješenje je na stranici 126.)

Zadatak 7.

Bacamo dvije simetrične kockice sa šest strana. Neka je X suma brojeva koje dobijemo bacanjem tih kockica ($X = 2d6$). Izračunajte

1. $E(X)$,
2. $Var(X)$,
3. za $\mu = E(X)$ i $\sigma = \sigma(X)$ izračunajte $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$.

(Rješenje je na stranici 126.)

13 Bernoullijev pokus. Bernoullijeva shema. Binomna distribucija

Bernoullijev pokus je eksperiment čiji ishod je slučajan i u kojem su moguća samo dva rezultata od kojih jedan nazivamo "uspjeh" (to je obično onaj koji nas zanima ili kojeg promatramo), a drugi "neuspjeh". Obično je riječ o eksperimentu s dva moguća rezultata koji se uvijek mogu izraziti preko "da ili ne" pitanja.

- Ako bacimo novčić, hoće li "pasti pismo"?
- Hoće li se roditi djevojčica?
- Jesu li nečije oči zelene boje?
- Je li potencijalni kupac odlučio kupiti proizvod?
- Hoće li strijelac pogoditi metu u jednom gađanju?

Vidimo da oznake "uspjeh" i "neuspjeh" služe samo kao označavanje rezultata i ne trebaju se shvatiti doslovno.

Primjeri Bernoullijevih pokusa:

- Bacanje novčića. Ovdje padanje glave može značiti "uspjeh", a padanje pisma "neuspjeh". Naravno da se možemo odlučiti i za obrnuto označavanje. Pošteni novčić po definiciji ima vjerojatnost uspjeha jednaku 0.5.
- Bacanje standardne kockice, gdje padanje šestice označava "uspjeh", a sve ostalo "neuspjeh".
- Anketiranje slučajnog glasača da bi se ustanovalo hoće li glasati "da" na predstojećem referendumu.

Matematički se takav pokus opisuje slučajnom varijablom X koja poprima samo dvije vrijednosti, 0 i 1, pri čemu se 1 smatra "uspjehom". Ako je p vjerojatnost uspjeha (obično se s $q = 1 - p$ označava vjerojatnost neuspjeha), onda je očekivanje takve slučajne varijable p , a standardna devijacija $\sqrt{p(1-p)}$:

$$\begin{aligned} E(X) &= p \cdot 1 + q \cdot 0 = p \\ E(X^2) &= p \cdot 1^2 + q \cdot 0^2 = p \\ Var(X) = E(X^2) - E(X)^2 &= p - p^2 = p(1 - p) = pq \Rightarrow \sigma(X) = \sqrt{p(1 - p)} = \sqrt{pq} \end{aligned}$$

Bernoullijeva shema se sastoji od konačnog ili beskonačnog niza nezavisnih slučajnih varijabli $X_1, X_2, X_3 \dots$ za koje vrijede sljedeće tvrdnje:

- Za svaki k je vrijednost od X_k ili 1 ili 0.
- Za svaki k je vjerojatnost da X_k poprima vrijednost 1 jednaka istom broju p , tj. za svaki k vrijedi $P(X_k = 1) = p$.

Drugim riječima, Bernoullijeva shema je niz nezavisnih Bernoullijevih pokusa s istom distribucijom. Dvije moguće vrijednosti svake varijable X_k se često zovu "uspjeh" i "neuspjeh" (u nekim slučajevima se koriste izrazi "pogodak" i "promašaj", u ovisnosti o kontekstu).

Neke od slučajnih varijabli povezanih s Bernoullijevom shemom su:

- Broj uspjeha u prvih n pokusa.
- Broj pokusa potrebnih da se dobije r uspjeha.
- Broj pokusa potrebnih da se dođe do uspjeha.

Mi ćemo obraditi samo prvu slučajnu varijablu čija distribucija se zove *binomna distribucija*.

Promotrimo slučajnu varijablu X koja "broji uspjeha" u Bernoullijevoj shemi u prvih n pokusa. Preciznije, X je slučajna varijabla koja može poprimiti vrijednosti iz skupa $\{0, 1, 2, \dots, n\}$ i za koju vrijedi da je $P(X = k)$ vjerojatnost da je u n uzastopnih Bernoullijevih pokusa nastupilo točno k uspjeha, $0 \leq k \leq n$. Neka je vjerojatnost uspjeha jednaka p i vjerojatnost neuspjeha $q = 1 - p$.

Bernoullijevi pokusi u Bernoullijevoj shemi su međusobno nezavisni pa je vjerojatnost da u nizu od n slučajnih pokusa njih k bude uspješno, a $n - k$ neuspješno, jednaka umnošku vjerojatnosti pojedinih pokusa. Njih k je uspješno s vjerojatnošću p , njih $n - k$ je neuspješno s vjerojatnošću q pa je pripadna vjerojatnost jednaka

$$\underbrace{p \cdot p \cdot \dots \cdot p}_{k \text{ puta}} \cdot \underbrace{q \cdot q \cdot \dots \cdot q}_{(n-k) \text{ puta}} = p^k q^{n-k}.$$

Takvih nizova od n pokusa od kojih je k uspješno ima koliko i načina da se od n objekata (pokusa) izabere njih k (koji su uspješni), a to je $\binom{n}{k}$. Dakle vjerojatnost da u nizu od n Bernoullijevih pokusa njih k bude uspješno je jednaka

$$P_k = P(X = k) = \binom{n}{k} p^k q^{n-k}.$$

Distribucija slučajne varijable X se zove *binomna distribucija*, zadana je parametrima n i p i označava se sa

$$X \sim B(n, p) = \begin{pmatrix} 0 & 1 & 2 & \dots & n \\ P_0 & P_1 & P_2 & \dots & P_n \end{pmatrix}.$$

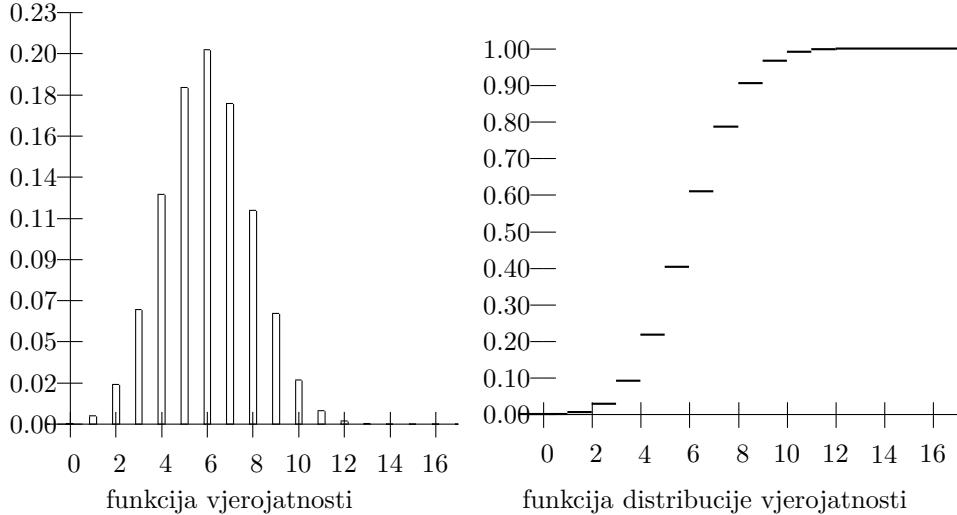
Očekivanje slučajne varijable s binomnom distribucijom se može dobiti direktno iz definicije uz primjenu binomnog teorema:

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

Varijanca jednog Bernoullijevog pokusa je pq . Pokusi u Bernoullijevoj shemi su međusobno nezavisni pa je varijanca binomne slučajne varijable jednaka zbroju varijanci pokusa, tj. imamo:

$$Var(X) = npq \quad \text{i} \quad \sigma(X) = \sqrt{npq}.$$

Grafovi funkcija vjerojatnosti i funkcije distribucije vjerojatnosti binomne distribucije $B(15, 0.4)$:



Zadaci

Zadatak 1. Koliko je očekivanje, a kolika je standardna devijacija binomne slučajne varijable $X \sim B(100, 0.4)$?

$$E(X) = np = 100 \cdot 0.4 = 40$$

$$\sigma(X) = \sqrt{pq} = \sqrt{0.4 \cdot 0.6} = \sqrt{0.24} \approx 0.490$$

Zadatak 2. Koliko je očekivanje, a kolika je standardna devijacija binomne slučajne varijable $X \sim B(75, 0.7)$?

Zadatak 3. Strijelac pogadja metu s vjerojatnošću $p = 0.4$.

- Kolika je vjerojatnost da će u seriji od 11 gađanja točno pet puta pogoditi metu? Riječ je o Bernoullijevoj shemi i binomnoj slučajnoj varijabli $X \sim B(11, 0.4)$. Treba izračunati vjerojatnost $P(X = 5)$:

$$P(X = 5) = \binom{11}{5} 0.4^5 0.6^6 = \frac{11 \cdot 10 \cdot 9 \cdot 8 \cdot 7}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} 0.4^5 0.6^6 = 462 \cdot 0.01024 \cdot 0.046656 \approx 0.2207$$

- Kolika je vjerojatnost da će u seriji od devet gađanja barem dva puta pogoditi metu? $X \sim B(9, 0.4)$, $P(X \geq 2) = ?$

$$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) = 1 - (P(X = 0) + P(X = 1))$$

$$P(X = 0) = \binom{9}{0} 0.4^0 0.6^9 = 1 \cdot 1 \cdot 0.6^9 = 0.010077696$$

$$P(X = 1) = \binom{9}{1} 0.4^1 0.6^8 = 9 \cdot 0.4 \cdot 0.6^8 = 0.060466176$$

$$P(X \geq 2) = 1 - (P(X = 0) + P(X = 1)) = 1 - 0.010077696 - 0.060466176 = 0.929456128$$

- Kolika je vjerojatnost da će u seriji od sedam gađanja najviše tri puta promašiti metu? $X \sim B(7, 0.4)$

Najviše tri puta promašiti u seriji od sedam gađanja je isto što i najmanje $7 - 3 = 4$ puta pogoditi. Dakle traži se $P(X \geq 4) = P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7)$

$$\begin{aligned}
P(X = 4) &= \binom{7}{4} 0.4^4 0.6^3 = \binom{7}{3} 0.4^4 0.6^3 \approx 35 \cdot 0.026 \cdot 0.216 \approx 0.197 \\
P(X = 5) &= \binom{7}{5} 0.4^5 0.6^2 = \binom{7}{2} 0.4^5 0.6^2 \approx 21 \cdot 0.0102 \cdot 0.36 \approx 0.0771 \\
P(X = 6) &= \binom{7}{6} 0.4^6 0.6^1 = \binom{7}{1} 0.4^6 0.6 \approx 7 \cdot 0.00410 \cdot 0.6 \approx 0.0172 \\
P(X = 7) &= \binom{7}{7} 0.4^7 0.6^0 = \binom{7}{0} 0.4^7 \approx 1 \cdot 0.00164 = 0.00164 \\
P(X \geq 4) &= P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7) \approx 0.197 + 0.0771 + 0.0172 + 0.00164 = 0.293
\end{aligned}$$

Zadatak 4. Strijelac pogađa metu s vjerojatnošću 0.7 .

- (a) Izračunajte vjerojatnost da će u seriji od 5 gađanja točno tri puta pogoditi metu.
- (b) Izračunajte vjerojatnost da će u seriji od 6 gađanja barem četiri puta promašiti metu.
- (c) Koliki je očekivani broj pogodaka u seriji od 7000 gađanja?

Zadatak 5. Strijelac pogađa metu s vjerojatnošću 0.6 .

- (a) Izračunajte vjerojatnost da će u seriji od 5 gađanja točno četiri puta promašiti metu.
- (b) Izračunajte vjerojatnost da će u seriji od 6 gađanja barem četiri puta promašiti metu.
- (c) Koliki je očekivani broj pogodaka u seriji od 5000 gađanja?

Zadatak 6. Bacamo simetričan novčić. Izračunajte vjerojatnost da će u seriji od 10 bacanja novčića barem tri puta pasti pismo.

Zadatak 7. Baca se standardna igrača kockica. Uspjehom smatramo kad padne broj šest.

1. Kolika je vjerojatnost uspjeha?

$$\text{Vjerojatnost uspjeha je } p = \frac{1}{6}.$$

2. Kolika je vjerojatnost da će u četiri bacanja broj šest pasti točno jednom?

Neka je X binomna slučajna varijabla s distribucijom $X \sim B(4, \frac{1}{6})$. Tražena vjerojatnost je jednaka

$$P(X = 1) = \binom{4}{1} p^1 q^3 = 4 \cdot \frac{1}{6} \cdot \left(\frac{5}{6}\right)^3 \approx 0.3858.$$

3. Kolika je vjerojatnost da će u četiri bacanja broj šest pasti barem jednom? (Jedan od de Méréovih problema.)

Tražena vjerojatnost je jednaka

$$\begin{aligned}
P(X \geq 1) &= P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 1 - P(X = 0) = \\
&= 1 - \binom{4}{0} p^0 q^4 = 1 - \left(\frac{5}{6}\right)^4 \approx 1 - 0.482 = 0.518
\end{aligned}$$

Zadatak 8. Istovremeno se bacaju dvije standardne igraće kockice. Uspjehom smatramo situaciju u kojoj su na obje kockice pale šestice.

1. Kolika je vjerojatnost uspjeha?
2. Kolika je vjerojatnost da će u 24 bacanja uspjeh nastupiti točno jednom?
3. Kolika je vjerojatnost da će u 24 bacanja uspjeh nastupiti barem jednom? (Jedan od de Méréovih problema.)

Zadatak 9.

Simetričan novčić je bačen četiri puta (sva bacanja su međusobno nezavisna). Kolika je vjerojatnost da broj glava neće biti manji od broja pisama?

Rješenje.

Neka je P broj pisama i G broj glava koji su se pojavili kao ishodi bacanja novčića četiri puta. Iz uvjeta zadatka slijedi da mora vrijediti $G \geq P$, a kako je $G + P = 4$, dobivamo $G \geq 4 - G$, tj. $G \geq 2$. Dakle, za $G \sim B(4, \frac{1}{2})$ trebamo izračunati $P(G \geq 2)$.

Vjerojatnost da u 4 bacanja k puta padne glava je $\binom{4}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{4-k} = \binom{4}{k} \frac{1}{2^4} = \binom{4}{k} \frac{1}{16}$.

Vjerojatnost da broj glava neće biti manji od broja pisama je

$$P(G \geq 2) = P(G = 2 \text{ ili } G = 3 \text{ ili } G = 4) = P(G = 2) + P(G = 3) + P(G = 4) = \frac{1}{16} \left(\binom{4}{2} + \binom{4}{3} + \binom{4}{4} \right) = \frac{1}{16} (6 + 4 + 1) = \frac{11}{16} = 0.6875.$$

Zadatak 10.

Novčić A je bačen tri puta i novčić B je bačen četiri puta (sva bacanja su međusobno nezavisna). Oba novčića su simetrična. Kolika je vjerojatnost da je na oba novčića pao isti broj glava?

Rješenje.

Traženi ishodi će se dogoditi kada na oba novčića padnu 0, 1, 2 ili 3 glave.

Neka je $A \sim B(3, \frac{1}{2})$ i $B \sim B(4, \frac{1}{2})$.

Vjerojatnost da u n bacanja k puta padne glava je $\binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = \binom{n}{k} \frac{1}{2^n}$. Potrebne vjerojatnosti su dane u sljedećoj tabeli:

k=broj glava	0	1	2	3
$P(A = k)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$
$P(B = k)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$

Vjerojatnost da je na oba novčića pao isti broj glava je jednaka:

$$\begin{aligned} P(A = B) &= P((A = 0 \text{ i } B = 0) \text{ ili } (A = 1 \text{ i } B = 1) \text{ ili } (A = 2 \text{ i } B = 2) \text{ ili } (A = 3 \text{ i } B = 3)) = \\ &= P(A = 0 \text{ i } B = 0) + P(A = 1 \text{ i } B = 1) + P(A = 2 \text{ i } B = 2) + P(A = 3 \text{ i } B = 3) = \\ &= P(A = 0)P(B = 0) + P(A = 1)P(B = 1) + P(A = 2)P(B = 2) + P(A = 3)P(B = 3) = \\ &= \frac{1}{8} \cdot \frac{1}{16} + \frac{3}{8} \cdot \frac{4}{16} + \frac{3}{8} \cdot \frac{6}{16} + \frac{1}{8} \cdot \frac{4}{16} = \frac{35}{128} = 0.2734375 \end{aligned}$$

Zadatak 11.

Vjerojantost da strijelac u n nezavisnih gađanja pogodi metu barem jednom je jednaka $r = 0.7626953125$.

Izračunajte n , ako je vjerojatnost da on pogodi metu u jednom gađanju jednaka $p = 0.25$. (Rješenje je na stranici 127.)

14 Poissonova distribucija

Poissonova distribucija se primjenjuje na događaje za koje vrijede sljedeći uvjeti:

1. događaji se mogu brojati nenegativnim cijelim brojevima,
2. događaji su međusobno nezavisni, tako da nastup jednog događaja ne utječe na nastupe niti jednog od sljedećih događaja,
3. prosječan broj nastupa događaja u danom vremenskom periodu (ili na danoj površini ili u danom volumenu) je poznat i konstantan,
4. moguće je odrediti broj nastupa događaja (npr. broj stranaka u baci tokom nekog vremenskog perioda), ali je besmisленo pitati koliko puta događaj nije nastupio (za razliku od binomne distribucije)

Za takve događaje, a oni su dosta česti, Poissonova distribucija daje, preko pripadajućih vjerojatnosti, očekivane frekvencije događaja.

Poissonova distribucija se može koristiti i u "suprotnom smjeru", za testiranje slučajnosti danih podataka. Ako se podaci dovoljno dobro podudaraju s teoretskim vrijednostima dobivenim iz Poissonove distribucije, onda je to jaka podrška tvrdnji da je slučaj, a ne nešto drugo, uzrok događajima na osnovu kojih su dobiveni ti podaci.

Poissonova distribucija je diskretna distribucija kojom se izražava vjerojatnost pojave događaja unutar fiksiranog vremenskog intervala ako je poznato prosječno vrijeme između pojave dva uzastopna događaja i ako su događaji međusobno neovisni. Također se može koristiti za druge intervale kojima se izražavaju udaljenost, površina ili volumen. Tipični primjeri su vremena između dolazaka uzastopnih kupaca u trgovinu i broj bombi koje (neovisno i slučajno) pogode slučajno izabrano područje iz grupe područja jednakih površina.

Ovaj drugi primjer je korišten za vrijeme Drugog svjetskog rata kada su njemačke snage raketnim bombama gadali London. Postavilo se pitanje jesu li neki dijelovi Londona u većoj opasnosti da budu pogodeni zbog strateških ciljeva (npr. tvornice) koji su se u njima nalazili. Drugim riječima, nije bilo jasno jesu li projektili padali na London slučajno, ili su pogadali unaprijed zadane ciljeve. Da dođe do odgovora, britanski statističar R. D. Clarke je podijelio dio grada koji je došao pod udar projektila na $676 (= 26 \cdot 26)$ područja od kojih je svako imalo površinu od jedne četvrtine kvadratnog kilometra. Za svako područje je odredio broj projektila koji su ga pogodili i sastavio je sljedeću tabelu u kojoj se vidjelo koliko je područja pogodilo dani broj projektila.

broj pogodaka, k	0	1	2	3	4	5 ili više
broj područja pogodjenih k puta	229	211	93	35	7	1
teoretski broj pogodaka prema Poissonovoj distribuciji	226.74	211.39	98.54	30.62	7.14	1.57

Ovime je Clarke pokazao da su projektili na London padali slučajno prema predvidivom obrascu određenim Poissonovom distribucijom i time dokazao da strateški važniji dijelovi Londona nisu bili u većoj opasnosti od drugih, odnosno da nema potrebe za evakuacijom iz onih područja grada u kojima se nalaze strateški ciljevi.

Poissonova distribucija je određena parametrom $\lambda \geq 0$, čije značenje je očekivani broj događaja tijekom zadanoj vremenskog intervala. Npr. ako se događaji pojavljuju svake 4 minute, a zanima nas njihovo pojavljivanje tijekom vremenskog intervala od 10 minuta, onda se kao model uzima Poissonova distribucija s parametrom $\lambda = 10/4 = 2.5$.

Da slučajna varijabla X ima Poissonovu distribuciju se označava s $X \sim Poi(\lambda)$.

Neka je

$$Poi(\lambda) = \begin{pmatrix} 0 & 1 & 2 & \cdots & k & \cdots \\ p_0 & p_1 & p_2 & \cdots & p_k & \cdots \end{pmatrix}.$$

Vjerojatnost p_k da slučajna varijabla X s Poissonovom distribucijom poprilično vrijednost k je jednaka

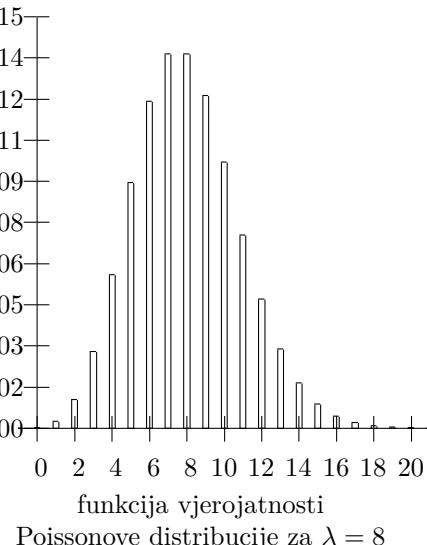
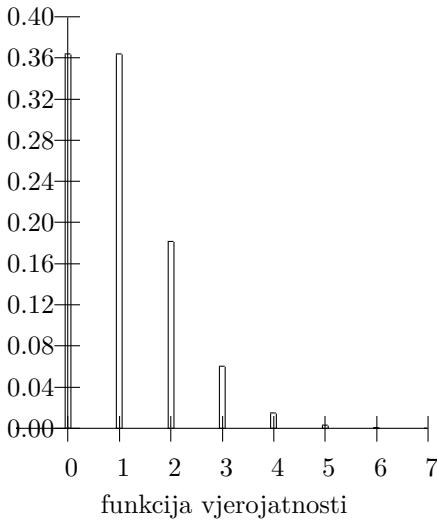
$$p_k = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

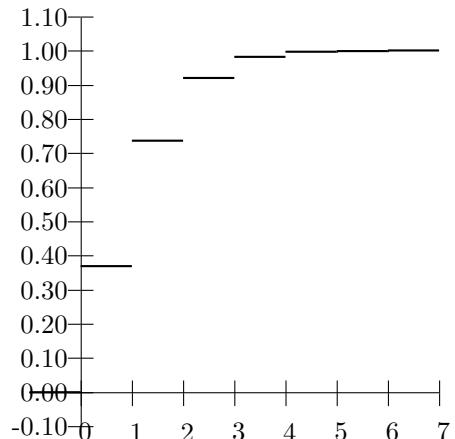
Provjerimo da je suma svih vjerojatnosti distribucije jednaka jedan:

$$\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

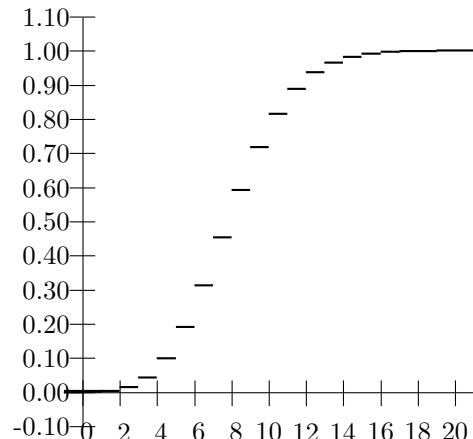
Vrijedi još i: $E(X) = \lambda$, $Var(X) = \lambda$, $\sigma(X) = \sqrt{\lambda}$.

$$E(X) = \sum_{k=0}^{\infty} k p_k = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} k \lambda \frac{\lambda^{k-1}}{k!} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$





funkcija distribucije vjerojatnosti
Poissonove distribucije za $\lambda = 1$



funkcija distribucije vjerojatnosti
Poissonove distribucije za $\lambda = 8$

Poissonova distribucija se koristi za modeliranje broja pojavljivanja različitih događaja u danom prostoru ili vremenu poput:

- broja vozila koji prođu nekom kontrolnom točkom (dovoljno dalekom od semafora) tijekom nekog danog vremena,
- broja tipfelera napravljenih prilikom tipkanja jedne stranice teksta,
- broja telefonskih poziva upućenih nekoj telefonskoj centrali tijekom jedne minute,
- broja pristupa nekom web serveru tijekom jedne minute,
- broja pregaženih životinja nađenih na danoj duljini ceste,
- broja zvijezda u danom volumenu prostora,
- broja žarulja koje pregore tijekom danog vremenskog intervala...

Poissonova distribucija $Poi(\lambda)$ se koristi kao aproksimacija binomne distribucije $B(n, p)$ u slučajevima kada je n velik i p malen, pri čemu je $\lambda = np$.

Primjer 1.

Neka je $X \sim Poi(5)$. Izračunajte $P(X = 2)$, $P(X = 5)$, $P(X < 3)$ i $P(X > 4)$.

1. $P(X = 2) = \frac{5^2 e^{-5}}{2!} \approx 0.084224$
2. $P(X = 5) = \frac{5^5 e^{-5}}{5!} \approx 0.175467$
3. $P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{5^0 e^{-5}}{0!} + \frac{5^1 e^{-5}}{1!} + \frac{5^2 e^{-5}}{2!} \approx 0.006738 + 0.033690 + 0.084224 = 0.124652$
4. $P(X > 4) = 1 - P(X \leq 4) = 1 - (P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)) = 1 - \left(\frac{5^0 e^{-5}}{0!} + \frac{5^1 e^{-5}}{1!} + \frac{5^2 e^{-5}}{2!} + \frac{5^3 e^{-5}}{3!} + \frac{5^4 e^{-5}}{4!} \right) \approx 1 - 0.440493 = 0.559507$

Primjer 2.

Recepција у хотелу prima u prosjeku dva telefonska poziva na sat. Želimo izračunati vjerojatnosti da će broj poziva tokom jednog sata biti 0, 1, 2, 3, 4, 5 ili više.

Rješenje.

Ako u prosjeku imamo dva poziva na sat, onda je $\lambda = 2$. Dakle slučajna varijabla koja modelira ovu situaciju ima distribuciju $X \sim Poi(2)$. U sljedećoj tabeli su dane vjerojatnosti da će tokom jednog sata recepcija primiti 0, 1, 2, 3, 4, 5 ili više poziva.

broj poziva	0	1	2	3	4	5 ili više
vjerojatnost	$P(X = 0)$	$P(X = 1)$	$P(X = 2)$	$P(X = 3)$	$P(X = 4)$	$P(X \geq 5)$
	0.13534	0.27067	0.27067	0.18045	0.09022	0.05265

Primjer 3.

U jednom vatrogasnem domu su izračunali da u prosjeku dobivaju 2.1 lažnu dojavu o požarima na dan. Uz pretpostavku da možemo primjeniti Poissonovu distribuciju, izračunajte vjerojatnost da će u jednom danu biti

1. niti jedna lažna dojava,
2. četiri lažne dojave,
3. najviše tri lažne dojave,
4. barem dvije lažne dojave.

Rješenje.

Ako je riječ o Poissonovoj distribuciji, onda promatramo slučajnu varijablu $X \sim Poi(2.1)$ pa imamo:

1. $P(X = 0) = \frac{2.1^0 e^{-2.1}}{0!} = e^{-2.1} \approx 0.12246$,
2. $P(X = 4) = \frac{2.1^4 e^{-2.1}}{4!} \approx 0.099231$
3. $P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \approx 0.83864$,

4. $P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4) + \dots = 1 - (P(X = 0) + P(X = 1)) \approx 1 - 0.37961 = 0.62039.$

Primjer 4.

Poznato je da je 5% ljudi ljevoruko. Izračunajte vjerojatnost da se u grupi od 100 ljudi nalazi barem dvoje ljevorukih.

Rješenje.

Koristit ćemo Poissonovu distribuciju $X \sim Poi(5)$, jer je prosječan broj ljevorukih u grupi od 100 ljudi jednak $100 \cdot 0.05$. Tražena vjerojatnost je približno jednaka:

$$P(X \geq 2) = 1 - (P(X = 0) + P(X = 1)) \approx 0.96.$$

Zadaci

Zadatak 1.

Neka je $X \sim Poi(2.5)$. Izračunajte $P(X = 2)$, $P(X = 3)$, $P(X \leq 3)$ i $P(X > 5)$.

Zadatak 2.

Poznato je da se u izvjesnoj knjizi od 250 stranica nalazi 50 slučajno raspoređenih štamparskih grešaka. Nadite vjerojatnost da se na stotoj stranici te knjige ne nalazi niti jedna štamparska greška. Kolika je vjerojatnost da se na toj stranici nalaze najviše dvije štamparske greške? (Uputa: prosječan broj štamparskih grešaka po stranici je $50/250 = 0.2$.)

Zadatak 3.

Za jednog konobara je ustanovljeno da mu u prosjeku iz ruku ispadne i razbijje se 0.05 tanjura na sat. Kolika je vjerojatnost da mu tokom jednog sata niti jedan tanjur neće ispasti iz ruku?

VJEROJATNOST I STATISTIKA

PRIMJER 2. KOLOKVIJA

Napomena: sve decimalne brojeve izračunajte na četiri decimale točno, uz zaokruživanje, ako je to potrebno.

1. (a) U posudi se nalazi jedanaest zelenih, devet plavih i sedam crvenih kuglica. Iz nje se na slučajan način bez vraćanja izvlače četiri kuglice (tj. tako da sve kuglice u posudi imaju istu vjerojatnost da će biti izvučene).

Izračunajte vjerojatnost da će biti izvučene kuglice točno dviju boja.

- (b) Bacamo dvije simetrične kockice, crvenu i plavu. Izračunajte vjerojatnost da je suma brojeva palih na kockicama veća od 6 (događaj A). Izračunajte vjerojatnost da je na obje kockice pao paran broj (događaj B). Izračunajte vjerojatnost događaja A ako znate da se dogodio događaj B .

2. Šahist pobijeduje u jednoj partiji s vjerojatnošću 0.23.

(a) Izračunajte vjerojatnost da će šahist u nizu od 13 partija pobijediti točno 10 puta.

(b) Izračunajte vjerojatnost da će šahist u nizu od 8 partija izgubiti ili remizirati barem 6 puta.

U oba slučaja napišite koji događaj smatraste uspjehom i koja je distribucija slučajne varijable koju koristite u rješenju.

3. Za slučajnu varijablu X se zna da ima distribuciju oblika

$$X \sim \left(\begin{array}{cccccc} -3 & 0 & 1 & 3 & 6 & a \\ \frac{1}{4} & 2p & 2p & 3p & 3p & 8p \end{array} \right)$$

i očekivanje $E(X) = \frac{67}{24}$. Odredite nepoznanice a i p , napišite distribuciju slučajne varijable X i izračunajte njenu standardnu devijaciju.

Izračunajte $P(\mu - \sigma \leq X \leq \mu + \sigma)$, (pri čemu je $\mu = E(X)$ i $\sigma = \sigma(X)$).

SKICE RJEŠENJA

1. (a) (11 zelenih, 9 plavih, 7 crvenih; izvučene kuglice su točno dvije boje)

Od broja načina da izvučemo kuglice najviše dviju boja (gdje kuglice treće boje privremeno zanemarujemo) treba oduzeti brojeve načina da od tih kuglica dviju boja izvučemo samo kuglice iste boje. To treba napraviti za sve parove boja i dobivene rezultate zbrojiti.

$$\text{zelene i plave kuglice: } \binom{11+9}{4} - \binom{11}{4} - \binom{9}{4} = 4389.$$

$$\text{zelene i crvene kuglice: } \binom{11+7}{4} - \binom{11}{4} - \binom{7}{4} = 2695.$$

$$\text{plave i crvene kuglice: } \binom{9+7}{4} - \binom{9}{4} - \binom{7}{4} = 1659.$$

Ukupan broj načina je 8743.

Broj načina na koje možemo izvući 4 kuglica od njih 27 je $\binom{27}{4} = 17550$.

Tražena vjerojatnost je $\frac{8743}{17550} \approx 0.498177$

- (b) Događaji A i B se mogu predstaviti ovako:

$c \setminus p$	1	2	3	4	5	6
1						A
2		B		B	A	AB
3				A	A	A
4		B	A	AB	A	AB
5		A	A	A	A	A
6	A	AB	A	AB	A	AB

$$P(A) = \frac{21}{36} = \frac{7}{12}, P(B) = \frac{9}{36} = \frac{1}{4}, P(A \cap B) = \frac{6}{36} = \frac{1}{6}, P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2}{3}$$

2. (šahist i pobjeđivanje u partiji šaha, $p = 0.23$)

- (a) $X \sim B(13, 0.23)$

$$P(X = 10) = \binom{13}{10} \cdot 0.23^{10} \cdot 0.77^3 \approx 286 \cdot 4.14265e - 07 \cdot 0.456533 \approx 5.40899e - 05$$

- (b) $X \sim B(8, 0.23)$, uspjeh je ako šahist pobijedi

Treba izračunati vjerojatnost da će u seriji od 8 pokusa biti najviše 2 uspjeha.

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$P(X \leq 2) \approx 0.727582 (\approx P(Y \geq 6) \text{ za } Y \sim B(8, 0.77))$$

3. Suma svih vjerojatnosti u drugom retku distribucije mora biti jednaka 1 pa dobivamo jednadžbu

$$\frac{1}{4} + 2p + 2p + 3p + 3p + 8p = 1 \text{ kojoj je rješenje } p = \frac{1}{24}. \text{ Iz formule za očekivanje dobivamo jednadžbu:}$$

$$E(X) = \frac{1}{4} \cdot (-3) + \frac{1}{12} \cdot 0 + \frac{1}{12} \cdot 1 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 6 + \frac{1}{3} \cdot a = \frac{67}{24}$$

kojoj je rješenje $a = 7$.

$$E(X^2) = \frac{583}{24} \approx 24.2917, E(X) = \frac{67}{24} \approx 2.79167$$

$$Var(X) = E(X^2) - E(X)^2 \approx 24.2917 - 7.7934 = 16.4983, \sigma(X) = \sqrt{Var(X)} \approx 4.06181$$

$$\mu = 2.79167, \sigma = 4.06181 \Rightarrow$$

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = P(-1.27014 \leq X \leq 6.85347) = \frac{1}{12} + \frac{1}{12} + \frac{1}{8} + \frac{1}{8} = \frac{5}{12} \approx 0.416667$$

15 Kontinuirana slučajna varijabla

U mnogim slučajevima nam za modeliranje slučajnih procesa diskretne slučajne varijable nisu dovoljne i to naročito onda kada ishodi slučajnih procesa nisu izolirani brojevi, nego mogu poprimiti bilo koju vrijednost iz nekog intervala realnih brojeva. Npr. žarulje se testiraju tako da ih se drži uključenima sve dok ne pregore. Slučajna varijabla Y povezana s ovom situacijom mjeri vrijeme do pregaranja u recimo satima. Vrijednosti koje može Y poprimiti su nenegativni realni brojevi, dakle $[0, +\infty)$ pa je Y kontinuirana varijabla.

Primjeri kontinuiranih slučajnih varijabli:

- visina ili težina ljudi na nekom području,
- količina šećera u jednom kilogramu grožđa,
- vrijeme potrebno da se pretrči sto metara,
- vrijeme čekanja na autobus na autobusnoj stanici.

Formalno, *kontinuirana slučajna varijabla* je funkcija sa prostora elementarnih događaja u skup realnih brojeva pri čemu skup vrijednosti te funkcije može biti cijeli interval (ili unija intervala) realnih brojeva ili cijeli skup \mathbb{R} .

Vjerojatnost svakog elementarnog događaja je kod kontinuirane slučajne varijable jednaka nuli. Smisla ima tražiti vjerojatnost da slučajna varijabla poprimi vrijednost iz nekog intervala, što se definira uz pomoć funkcije distribucije vjerojatnosti ili funkcije gustoće vjerojatnosti.

Funkcija gustoće vjerojatnosti kontinuirane slučajne varijable

Funkcija gustoće vjerojatnosti kontinuiranih slučajnih varijabli je funkcija $f : \mathbb{R} \rightarrow \mathbb{R}$ sa sljedećim svojstvima:

1. funkcija f je nenegativna, tj. za svaki $x \in \mathbb{R}$ vrijedi $f(x) \geq 0$

$$2. \int_a^b f(t) dt = P(a \leq X \leq b)$$

$$3. \int_{-\infty}^{+\infty} f(t) dt = 1$$

Kako je vjerojatnost da X poprimi neku konkretnu vrijednost jednaka nuli (što slijedi iz $P(X = a) = \int_a^a f(t) dt = 0$), sve sljedeće vjerojatnosti su međusobno jednake:

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

Funkcija distribucije vjerojatnosti kontinuirane slučajne varijable

Funkciju distribucije vjerojatnosti F kontinuirane slučajne varijable X definiramo formulom

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

gdje je $f(x)$ pripadajuća funkcija gustoće vjerojatnosti. Iz navedenog i Newton–Leibnizove formule slijedi

$$P(a < X \leq b) = F(b) - F(a)$$

i

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

tj. funkcija distribucije vjerojatnosti kontinuirane slučajne varijable je primitivna funkcija pripadne funkcije gustoće vjerojatnosti.

16 Normalna (Gaussova) distribucija. Standardna normalna distribucija

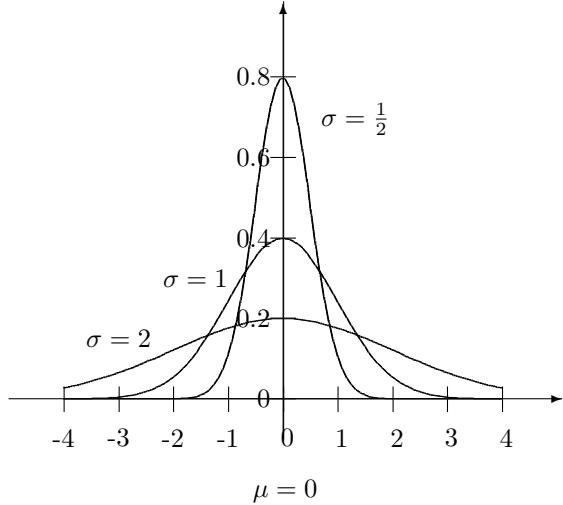
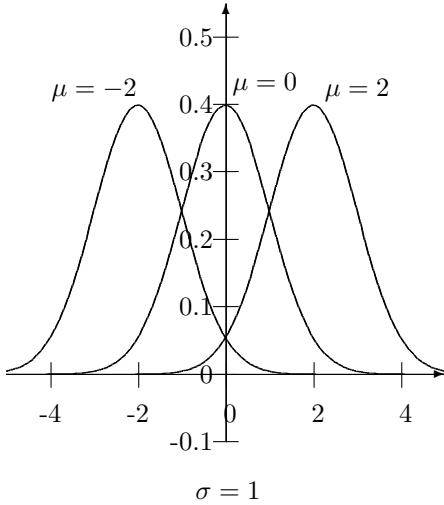
Normalna distribucija, koja se često zove *Gaussova distribucija*, je vrlo važan primjer kontinuirane distribucije primjenjive u raznim područjima. Parametri te distribucije su dva broja, μ (očekivanje) i σ^2 (varijanca). Da neprekidna slučajna varijabla X ima normalnu distribuciju s parametrima μ i σ^2 , označavamo ovako:

$$X \sim N(\mu, \sigma^2).$$

za *standardnu normalnu distribuciju* vrijedu $\mu = 0$ i $\sigma^2 = 1$. Carl Friedrich Gauss je otkrio normalnu distribuciju kada je analizirao neke astronomске podatke i tada je definirao jednadžbu njezine funkcije gustoće:

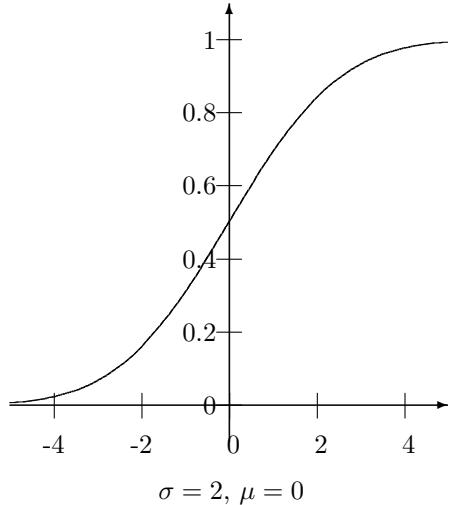
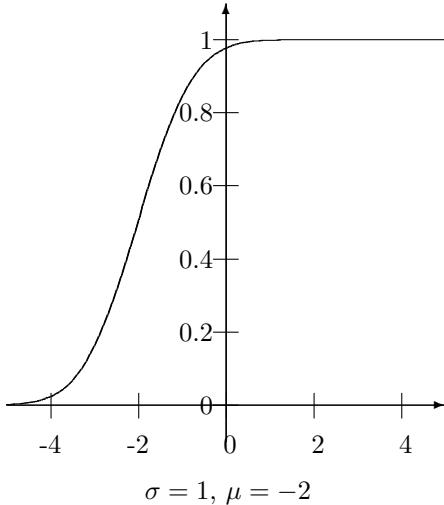
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

Funkcije gustoće vjerojatnosti za različite σ i μ :



Funkcija distribucije vjerojatnosti se definira sa: $F(x) = \int_{-\infty}^x f(t) dt$.

Funkcije distribucije vjerojatnosti:



Vjerojatnost da slučajna varijabla $X \sim N(\mu, \sigma^2)$ poprimi vrijednost manju od ili jednaku b je jednaka (vidi lijevu sliku na stranici 97):

$$P(X \leq b) = \int_{-\infty}^b f(x) dx = F(b)$$

gdje je F pripadajuća funkcija distribucije vjerojatnosti. Da bi izračunali tu vjerojatnost, koristimo se tabelom vrijednosti funkcije distribucije standardne razdiobe $N(0, 1)$, koja je dana na stranici 129. Da bismo mogli koristiti tu tablicu, moramo prvo podatke vezane uz zadatu normalnu distribuciju prevesti u odgovarajuće podatke iz standardne normalne distribucije formulom

$$x^* = \frac{x - \mu}{\sigma}$$

pri čemu vrijedi

$$F(x) = F^*(x^*)$$

gdje je F^* funkcija distribucije vjerojatnosti standardne normalne razdiobe. Sada imamo:

$$P(X \leq b) = F(b) = F^*\left(\frac{b - \mu}{\sigma}\right) = F^*(b^*).$$

Približnu vrijednost za $F^*(b^*)$ očitavamo direktno iz tabele (ili s malo boljih kalkulatora).

Primjer 1.

Da bi našli $F^*(1.23)$, radimo sljedeće:

1. U prvom stupcu tabele tražimo broj 1.2.
2. U prvom retku tabele tražimo broj 0.03.
3. Tražena vrijednost je broj koji se nalazi u presjeku nađenog retka i stupca.

a	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Sada imamo: $F^*(1.23) \approx 0.89065$.

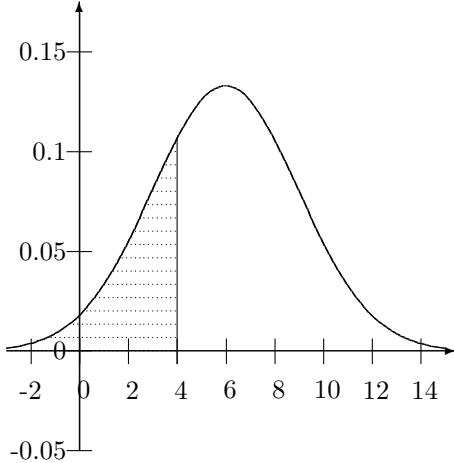
Ako trebamo pronaći vrijednost $F^*(x)$ gdje x ima više od dvije decimalne, onda x zaokružujemo na dva decimalna mesta i u tabeli nalazimo traženu vrijednost za taj zaokruženi broj. Npr. $F^*(2.758) \approx F^*(2.76) \approx 0.99711$.

Ako je $x > 4.09$, onda uzimamo da je $F^*(x) \approx 1$. Npr. $F(5.12) \approx 1$.

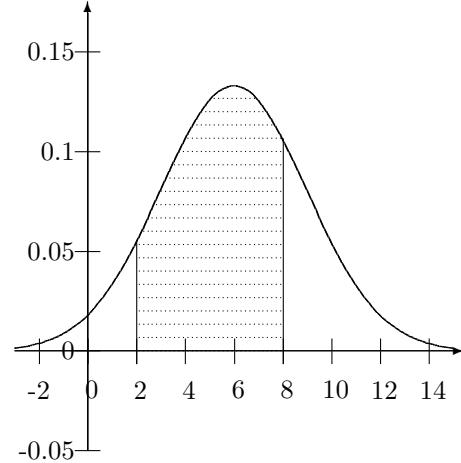
Ako je $x < 0$, onda općenito vrijedi $F(x) = 1 - F(-x)$, što je posljedica identiteta $P\{X \leq x\} + P\{X \geq x\} = 1$ i $P\{X \geq x\} = P\{X \leq -x\}$. Npr. $F^*(-0.54) = 1 - F^*(0.54) \approx 1 - 0.70540 = 0.2946$.

Vjerojatnost da slučajna varijabla $X \sim N(\mu, \sigma^2)$ poprimi vrijednost veću od a i manju od ili jednaku b je jednaka (vidi desnu sliku):

$$P(a < X \leq b) = \int_a^b f(t) dt = F(b) - F(a)$$



Označeno područje ima površinu
 $P(X \leq 4)$.



Označeno područje ima površinu
 $P(-1 \leq X \leq 4)$.

Primjer 2.

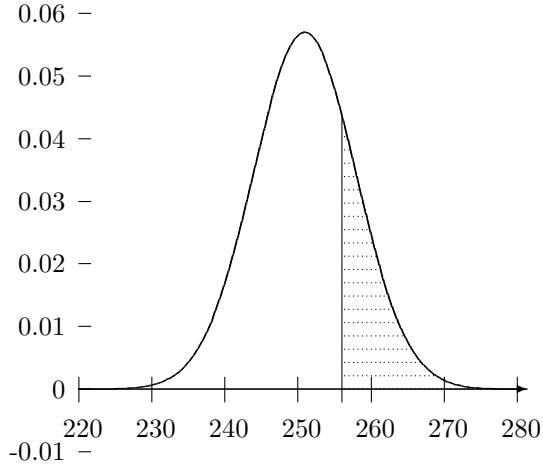
Broj sunčanih dana u jednoj godini na otoku Hvaru normalno je distribuirana slučajna varijabla X s para-

metrima $\mu = 251$ i $\sigma^2 = 49$. Izračunajte vjerojatnost da je broj sunčanih dana u godini

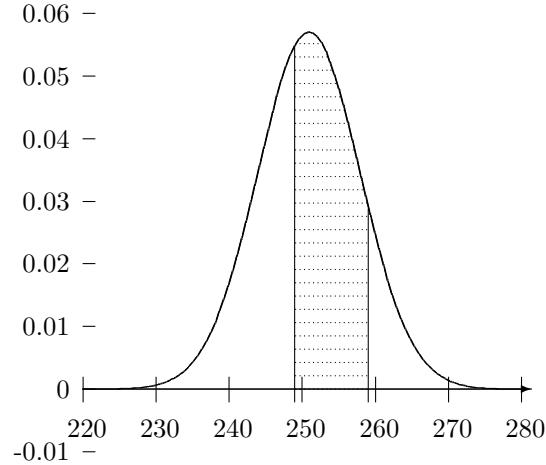
- (a) veći od 256, $X > 256$;
- (b) između 249 i 259, $249 < X \leq 259$?

Rješenje.

$$\begin{aligned}
 \text{(a)} \quad P(X > 256) &= 1 - P(X \leq 256) = 1 - F(256) = 1 - F^*\left(\frac{256 - 251}{7}\right) \approx \\
 &\approx 1 - F^*(0.71) \approx 1 - 0.76115 = 0.23885, \\
 \text{(b)} \quad P(249 < X \leq 259) &= F(259) - F(249) = F^*\left(\frac{259 - 251}{7}\right) - F^*\left(\frac{249 - 251}{7}\right) \approx F^*(1.14) - F^*(-0.29) \\
 &= F^*(1.14) - (1 - F^*(0.29)) \approx 0.87286 - (1 - 0.61409) = 0.48695.
 \end{aligned}$$



$$P(X > 256)$$



$$P(249 \leq X \leq 259)$$

Primjer 3.

Prepostavimo da je temperatura zraka u travnju normalna distribuirana slučajna varijabla T s $\mu = 20^\circ C$ i standardnom devijacijom $\sigma = 3^\circ C$. Nadite vjerojatnost da je temperatura zraka u travnju

- (a) manja od $18^\circ C$;
- (b) između $21^\circ C$ i $23^\circ C$.

Rješenje.

$$\begin{aligned}
 \text{(a)} \quad P(T \leq 18) &= F(18) = F^*\left(\frac{18 - 20}{3}\right) \approx F^*(-0.67) \approx 1 - 0.74857 = 0.25143, \\
 \text{(b)} \quad P(21 < T \leq 23) &= F(23) - F(21) = F^*\left(\frac{23 - 20}{3}\right) - F^*\left(\frac{21 - 20}{3}\right) \approx F^*(1.00) - F^*(0.33) \approx \\
 &\approx 0.84134 - 0.62930 = 0.21204.
 \end{aligned}$$

Zadaci

Zadatak 1.

Očitajte iz tabele vrijednosti funkcije distribucije standardne normalne razdiobe $N(0, 1)$ sljedeće vrijednosti: $F^*(2.71)$, $F^*(1.789)$, $F^*(0.334)$, $F^*(5.23)$, $F^*(0.1212)$, $F^*(-0.34)$, $F^*(-1.3476)$, $F^*(-2.96243)$, $F^*(-7.22)$.

Zadatak 2.

Broj sunčanih dana u jednoj godini na otoku Hvaru normalno je distribuirana slučajna varijabla s parametrima $\mu = 320$ i $\sigma^2 = 81$. Kolika je vjerojatnost da je broj sunčanih dana u godini između 290 i 322?

Zadatak 3.

Broj kišnih dana u jednoj godini na otoku Hvaru normalno je distribuirana slučajna varijabla s parametrima $\mu = 20$ i $\sigma^2 = 4$. Kolika je vjerojatnost da je broj kišnih dana u godini između 6 i 14?

Zadatak 4.

Pretpostavimo da je visina (u cm) studenata veleučilišta normalno distribuirana slučajna varijabla s parametrima $\mu = 171\text{cm}$ i $\sigma^2 = 26.01\text{cm}^2$. Izračunajte vjerojatnost da je visina slučajno odabranog studenta

1. veća od 168 cm,
2. između 170cm i 173.cm.

Zadatak 5.

Pretpostavimo da je visina stanovnika Lastova, u cm, normalno distribuirana slučajna varijabla s parametrima $\mu = 169$ i $\sigma^2 = 10.89$. Izračunajte vjerojatnost da je visina slučajno odabranog stanovnika

- (a) manja ili jednaka 168cm,
- (b) između 165cm i 170cm.

Zadatak 6. Pokažite da za svaku slučajnu varijablu $X \sim N(\mu, \sigma^2)$ vrijedi

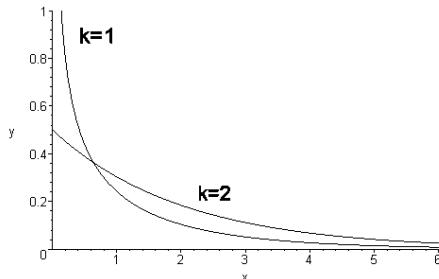
1. $P(\mu - \sigma < X \leq \mu + \sigma) \approx 0.68$,
2. $P(\mu - 2\sigma < X \leq \mu + 2\sigma) \approx 0.95$,
3. $P(\mu - 3\sigma < X \leq \mu + 3\sigma) \approx 0.997$.

17 χ^2 -distribucija

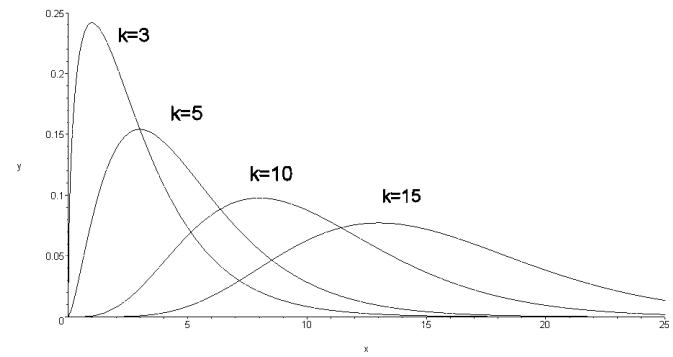
χ^2 -distribucija (hi kvadrat distribucija) ovisi o parametru $k \in \mathbb{N}$ kojeg zovemo stupanj slobode. Funkcija gustoće vjerojatnosti χ^2 -distribucija s k stupnjeva slobode je dana s

$$f_k(x) = c \cdot x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, \quad x > 0$$

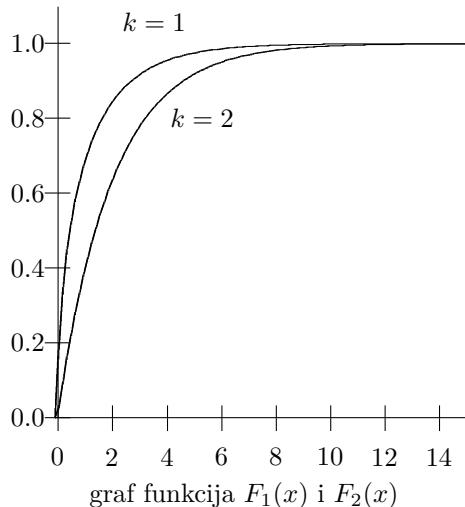
gdje je $c > 0$ konstanta koja ovisi o k takva da vrijedi $\int_0^\infty f_k(t)dt = 1$.



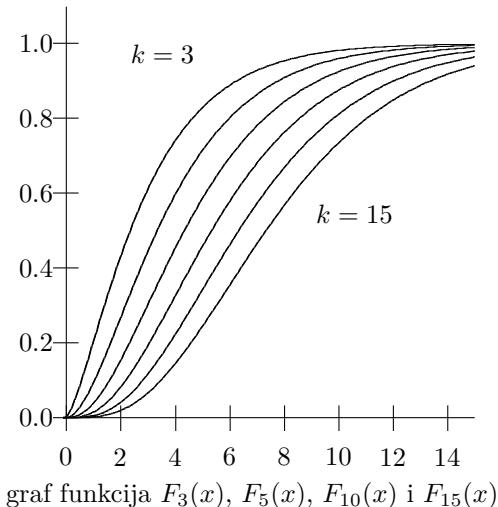
graf funkcija $f_1(x)$ i $f_2(x)$



graf funkcija $f_3(x)$, $f_5(x)$, $f_{10}(x)$ i $f_{15}(x)$



graf funkcija $F_1(x)$ i $F_2(x)$



graf funkcija $F_3(x)$, $F_5(x)$, $F_{10}(x)$ i $F_{15}(x)$

Oznaka za slučajnu varijablu X koja ima χ^2 -distribuciju s k stupnjeva slobode je $X \sim \chi_k^2$. Očekivanje takve varijable je $E(X) = k$, varijanca $D(X) = 2k$, a standardna devijacija $\sigma = \sqrt{2k}$. Za $k \geq 30$ vrijedi da slučajna varijabla $X \sim \chi_k^2$ ima približno normalnu distribuciju $N(k, 2k)$.

Dio III

Statistički testovi

Tipična situacija u kojoj se primjenjuju *statistički testovi* je kad imamo neku pretpostavku, koju nazivamo hipotezom, o nekom statističkom obilježju neke velike grupe ljudi (npr. zanimaju nas visine ili težine svih građana Republike Hrvatske), o seriji nekih proizvoda (npr. neka njihova dimenzija, težina, količina neispravnih proizvoda, ...) itd. Takvu grupu ljudi ili seriju proizvoda zovemo *populacija*.

Gotovo uvijek je nemoguće mjerjenje statističkog obilježja provesti na cijeloj populaciji pa se mjerjenje obavlja na jednoj manjoj grupi ljudi ili manjem dijelu jedne serije proizvoda. Ta manja grupa na kojoj se obavlja mjerjenje se zove *uzorak*. Postoji teorija uzorka koja se bavi time kako odabrati dobar (nepristran) uzorak, no time se nećemo baviti. Statistički testovi nam pomažu da donesemo neki zaključak u vezi promatranog statističkog obilježja na cijeloj populaciji na osnovu mjerena provedenog na uzorku.

Obradit ćemo dva tipa statističkih testova koji se odnose na testiranje hipoteze:

- testiranje hipoteze o očekivanju normalno distribuirane varijable uz poznatu varijancu,
- χ^2 test.

Prilikom testiranja hipoteze uvijek imamo tzv. nul hipotezu (označavamo je s H_0) koju neposredno provjeravamo testom. Rezultat testa može biti njen prihvatanje ili odbacivanje. U slučaju kada nul hipotezu odbacujemo, onda to znači da prihvaćamo alternativnu hipotezu (koju označavamo s H_1) koja može, ali i ne mora, biti negacija nul hipoteze.

Kod testiranja hipoteze možemo napraviti dvije vrste grešaka.

- *Greška prve vrste* se događa kada nul hipotezu odbacimo iako je ona istinita.
Na primjer, u kliničkom testiranju novog lijeka bi nul hipoteza mogla biti da novi lijek u prosjeku nije bolji od lijeka koji se trenutno koristi. Greška prve vrste bi se pojavila ako bismo zaključili da ova dva lijeka daju različite rezultate dok u stvari nema razlike u njihovim primjenama.
- *Greška druge vrste* se događa kada nul hipotezu prihvatom, iako ona nije istinita.
U gornjem primjeru bi ovaj tip greške nastao ako bismo zaključili da oba lijeka u prosjeku polučuju jednakе rezultate u primjeni iako to nije istina.

Sljedeća tabela prikazuje sve moguće rezultate testiranja hipoteze:

	H_0 odbačena	H_0 prihvaćena
H_0 istinita	greška I vrste	ispravna odluka
H_0 neistinita	ispravna odluka	greška II vrste

Greška prve vrste je ozbiljnija i važnije je da obratimo pažnju na to kako da izbjegnemo nju, nego grešku druge vrste. Testovi hipoteze imaju jedan parametar, razinu značajnosti (označava se s α), koji ima značenje vjerojatnosti da ćemo napraviti grešku prve vrste. Taj parametar treba biti što manji, ali ne može biti nula; uobičajeno je da se za njega uzmu vrijednosti $\alpha = 0.05$ ili $\alpha = 0.01$. Ako ne odbacimo nul hipotezu, još uvijek je moguće da je ona neistinita, jer uzorak možda nije bio dovoljno velik da bi se mogla detektirati njena

neistinitost. Općenito, za zadanu razinu značajnosti α nije moguće izračunati vjerojatnost greške druge vrste (označava se s β), no zna se da što je jedna od te dvije vrijednosti manja, to je druga veća. Jakost testa hipoteze je njegova sposobnost da odbaci nul hipotezu kada je ona neistinita, preciznije, vjerojatnost da se neće dogoditi greška druge vrste, a to je $1 - \beta$.

Vrijednost testa je broj kojeg dobijemo na osnovu uzorka. Kritična vrijednost je teoretska granica koja ovisi o razini značajnosti (i ne ovisi o uzorku). Ona dijeli skup vrijednosti koje može poprimiti vrijednost testa na dva dijela. Ako vrijednost testa pripada kritičnom području, onda nul hipotezu odbacujemo, inače je prihvaćamo.

18 Testiranje hipoteze o očekivanju uz poznatu varijancu

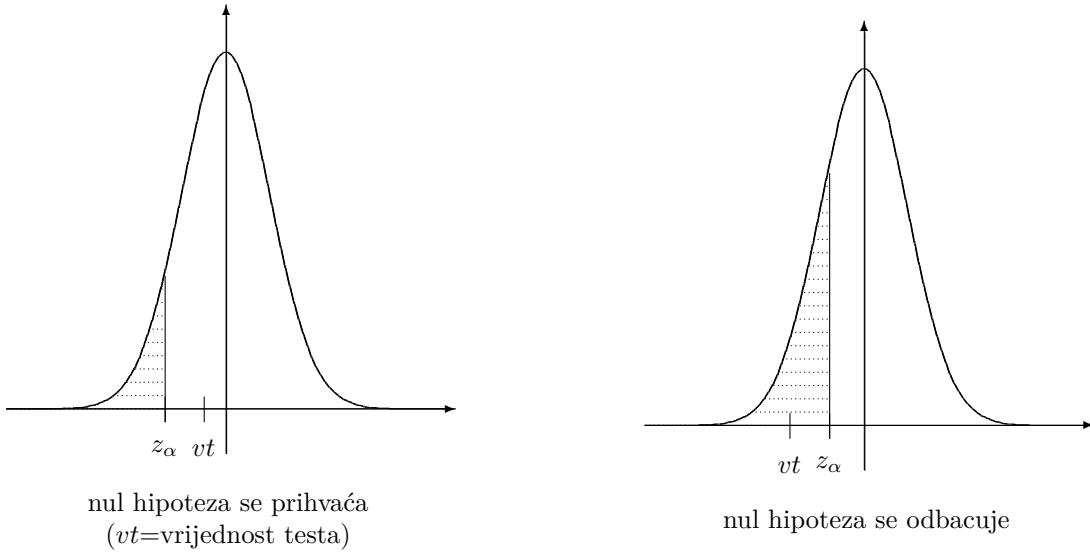
Neka je X slučajna varijabla s nepoznatim očekivanjem μ i poznatom standardnom devijacijom σ , definirana na nekoj populaciji. *Testiranje hipoteze o očekivanju uz poznatu varijancu* se provodi tako da se koristeći aritmetičku sredinu uzorka izračuna vrijednost testa koja se uspoređuje s kritičnom vrijednosti, iz čega se onda izvodi zaključak o prihvaćanju ili odbacivanju nul hipoteze. Test se može provesti ako je X slučajna varijabla s normalnom distribucijom $N(\mu, \sigma^2)$.

Moguće su tri vrste provođenja testa: lijevi, desni i obostrani. U svakom od njih računamo vrijednost testa $\frac{\bar{x}-\mu}{\sigma}\sqrt{n}$. Moraju nam biti poznati razina značajnosti α (obično malen broj, najčešće 0.05 ili 0.01), varijanca σ^2 slučajne varijable X , a ako uzorak nije poznat, onda moramo znati veličinu uzorka n , aritmetičku sredinu \bar{x} uzorka i jesu li svi elementi uzorka manji od μ , veći od μ ili u uzorku ima i brojeva koji su veći od μ i brojeva koji su od njega manji (da bismo mogli odrediti koji test trebamo provesti: lijevi, desni ili obostrani). Nulhipoteza je u sva tri slučaja ista.

$$\text{lijevi } \begin{cases} H_0 : \mu = \mu_0 & \text{nul hipoteza} \\ H_1 : \mu < \mu_0 & \text{alternativna hipoteza} \end{cases}$$

Ako je $\frac{\bar{x}-\mu}{\sigma}\sqrt{n} > z_\alpha = -z_{1-\alpha}$ onda nulhipotezu prihvaćamo (ne odbacujemo) s razinom značajnosti α jer vrijednost testa nije u kritičnom području. U suprotnom nulhipotezu odbacujemo (ne prihvaćamo) s razinom značajnosti α .

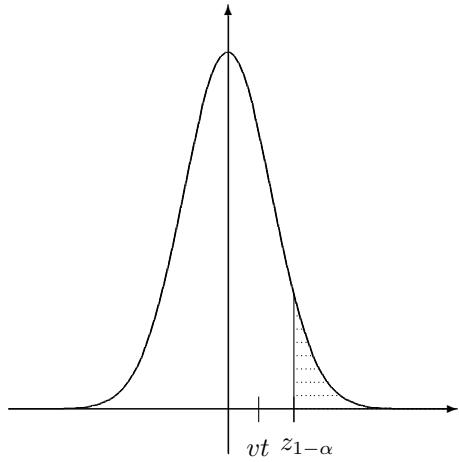
Lijevi test provodimo samo onda kada su sve vrijednost uzorka manje od μ_0 .



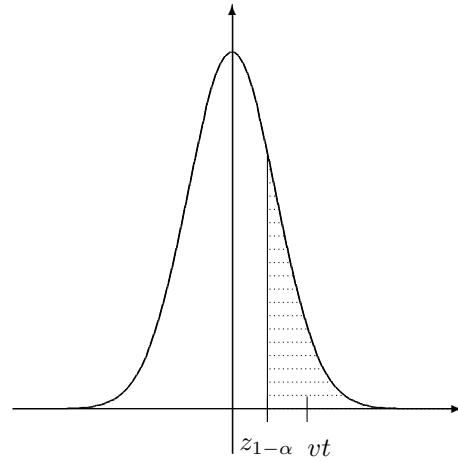
$$\text{desni } \begin{cases} H_0 : \mu = \mu_0 & \text{nul hipoteza} \\ H_1 : \mu > \mu_0 & \text{alternativna hipoteza} \end{cases}$$

Ako je $\frac{\bar{x}-\mu}{\sigma}\sqrt{n} < z_{1-\alpha}$ onda nulhipotezu ne odbacujemo s razinom značajnosti α jer vrijednost testa nije u kritičnom području. U suprotnom nulhipotezu odbacujemo s razinom značajnosti $\alpha = 0.05$.

Desni test provodimo samo onda kada su sve vrijednost uzorka veće od μ_0 .



nul hipoteza se prihvaca

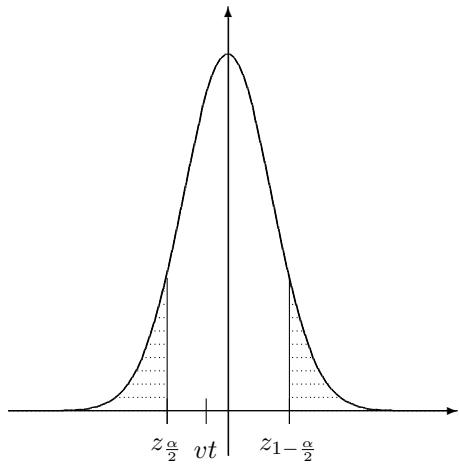


nul hipoteza se odbacuje

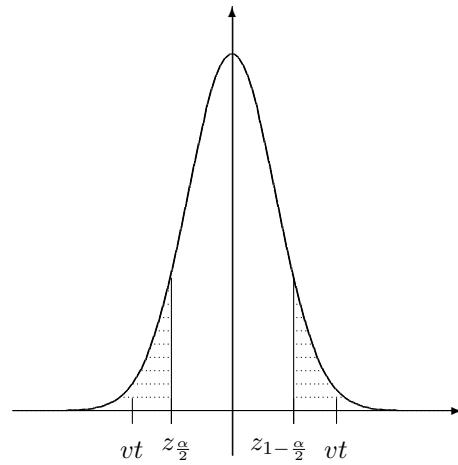
obostrani $\begin{cases} H_0 : \mu = \mu_0 & \text{nul hipoteza} \\ H_1 : \mu \neq \mu_0 & \text{alternativna hipoteza} \end{cases}$

Ako je $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} > z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}}$ i $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < z_{1-\frac{\alpha}{2}}$ onda nulhipotezu ne odbacujemo s razinom značajnosti α jer vrijednost testa nije u kritičnom području. U suprotnom nulhipotezu odbacujemo s razinom značajnosti $\alpha = 0.05$.

Obostrani test provodimo samo onda kada su neke vrijednosti uzorka manje, a neke veće od μ_0 .



nul hipoteza se prihvaca



nul hipoteza se odbacuje

Vrijednosti z_α očitavamo iz tabele na strani 129.

Primjer 1.

U tablici su popisana izmjerena odstupanja promjera matica od propisane vrijednosti (izražena u milimetrima). Uz pretpostavku da je vrijednost odstupanja normalno distribuirana slučajna varijabla s parametrima μ i $\sigma^2 = 0.0036$, odgovarajućim testom testirajte hipotezu $H_0 : \mu = 0.29$ uz razinu značajnosti $\alpha = 0.001$.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
0.27	0.35	0.21	0.28	0.27	0.27	0.31	0.24

Rješenje.

Prvo izračunamo aritmetičku sredinu uzorka:

$$\bar{x} = \frac{0.27 + 0.35 + 0.21 + 0.28 + 0.27 + 0.27 + 0.31 + 0.24}{8} = 0.275.$$

Vrijednost testa je: $\frac{\bar{x} - \mu_0}{\sigma} \sqrt{n} = \frac{0.275 - 0.29}{0.06} \sqrt{8} \approx -0.7071$.

Koristit ćemo obostrani test, jer u uzorku ima vrijednosti i većih i manjih od μ_0 .

$$\begin{cases} H_0 : \mu = 0.29 \text{ mm} & \text{nul hipoteza} \\ H_1 : \mu \neq 0.29 \text{ mm} & \text{alternativna hipoteza} \end{cases}$$

Trebamo izračunati kritične vrijednosti $z_{\frac{\alpha}{2}}$ i $z_{1-\frac{\alpha}{2}}$. Znamo da vrijedi $z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}}$ pa je dovoljno naći samo $z_{1-\frac{\alpha}{2}} = z_{0.9995}$ kojeg tražimo u tabeli tako da nađemo broj unutar tabele koji je najbliži broju 0.9995 te u početku stupca i početku retka u kojima se nalazi taj broj očitavamo traženu kritičnu vrijednost $z_{0.9995}$.

a	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976

Dakle imamo da je $z_{0.9995} = 3.29$ pa prema tome i $z_{0.0005} = -3.29$. Kako je

$$z_{\frac{\alpha}{2}} = -3.29 < \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n} = -0.7071 < z_{1-\frac{\alpha}{2}} = 3.29,$$

(dakle vrijednost testa nije u kritičnom području) nul hipotezu prihvaćamo s razinom značajnosti $\alpha = 0.001$.

Primjer 2.

U tablici su popisana izmjerena odstupanja debljine ploča od propisane vrijednosti (izražena u milimetrima). Uz pretpostavku da je vrijednost odstupanja normalno distribuirana slučajna varijabla s parametrima μ i $\sigma^2 = 0.0016$, odgovarajućim testom testirajte hipotezu $H_0 : \mu = 1.15$ uz razinu značajnosti $\alpha = 0.05$.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1.19	1.17	1.19	1.16	1.18	1.17	1.16	1.19

Rješenje.

Prvo izračunamo aritmetičku sredinu uzorka:

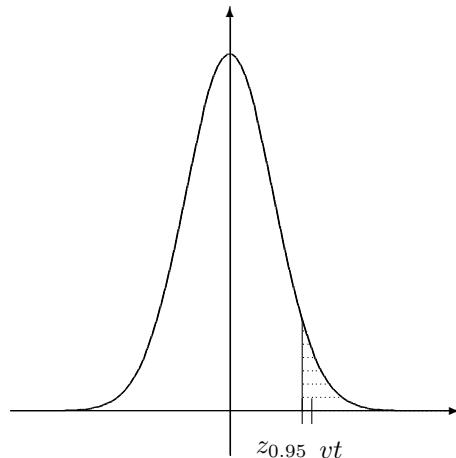
$$\bar{x} = \frac{1.19 + 1.17 + 1.19 + 1.16 + 1.18 + 1.17 + 1.16 + 1.19}{8} = 1.17625.$$

Vrijednost testa je: $\frac{\bar{x} - \mu_0}{\sigma} \sqrt{n} = \frac{1.17625 - 1.15}{0.04} \sqrt{8} \approx 1.8562$.

Koristit ćemo desni test, jer u uzorku su sve vrijednosti veće od μ_0 .

$$\begin{cases} H_0 : \mu = 1.15 \text{ mm} & \text{nul hipoteza} \\ H_1 : \mu > 1.15 \text{ mm} & \text{alternativna hipoteza} \end{cases}$$

Treba izračunati kritičnu vrijednost $z_{1-\alpha} = z_{0.95}$. Iz tabele dobivamo da je $z_{0.95} = 1.64$ (najbliža vrijednost u tabeli je 0.94950).



nul hipoteza se ne prihvata

Vidimo da je vrijednost testa veća od kritične vrijednosti što znači da je u kritičnom području pa nul hipotezu odbacujemo s razinom značajnosti $\alpha = 0.05$.

Primjer 3.

Stroj za proizvodnju matica je provjeren tako da je izmjereno odstupanje promjera matice od propisane

vrijednosti (u mm). U tabeli su navedena odstupanja:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
0.49	0.48	0.49	0.48	0.47	0.47	0.48	0.48

Uz pretpostavku da je vrijednost odstupanja normalno distribuirana slučajna varijabla s parametrima μ i $\sigma^2 = 0.0016$, testirajte hipotezu $H_0 : \mu = 0.5$. Razina značajnosti neka je $\alpha = 0.01$.

Rješenje.

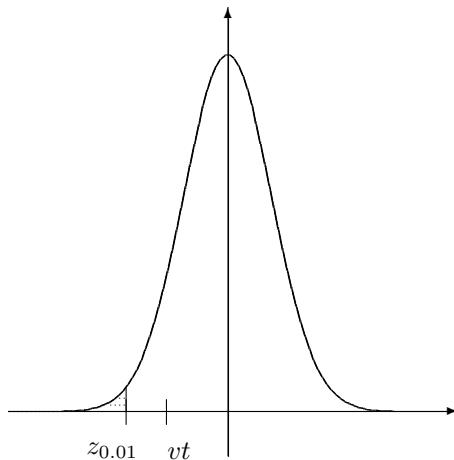
Provodimo lijevo testiranje s razinom značajnosti $\alpha = 0.05$. Iz podataka slijedi $\bar{x} = 0.48$.

$$\begin{cases} H_0 : \mu = 0.5 \text{ mm} & \text{nul hipoteza} \\ H_1 : \mu < 0.5 \text{ mm} & \text{alternativna hipoteza} \end{cases}$$

$$\text{Vrijednost testa je } \frac{\bar{x} - \mu}{\sigma} \sqrt{n} = \frac{0.48 - 0.5}{\sqrt{0.0016}} \sqrt{8} \approx -1.41.$$

$$z_\alpha = z_{0.01} = -z_{0.99} \approx -2.33 < -1.41.$$

Nulhipotezu prihvaćamo s razinom značajnosti $\alpha = 0.01$ jer se vrijednost testa nalazi van kritičnog područja.



nul hipoteza se prihvaca

Zadaci

Zadatak 1.

U tablici su popisana izmjerena odstupanja debljine ploča od propisane vrijednosti (izražena u milimetrima). Uz pretpostavku da je vrijednost odstupanja normalno distribuirana slučajna varijabla s parametrima μ i $\sigma^2 = 0,0025$, odgovarajućim testom testirajte hipotezu $H_0 : \mu = 1.20$ uz razinu značajnosti $\alpha = 0.05$. Navedite koji test koristite i napišite alternativnu hipotezu H_1 .

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1.29	1.28	1.29	1.28	1.27	1.27	1.28	1.28	1.27	1.27

Zadatak 2.

Provjera rada stroja za proizvodnju vijaka. U tablici su popisana izmjerena odstupanja promjera vijaka od propisane vrijednosti (izražena u milimetrima). Uz pretpostavku da je vrijednost odstupanja normalno distribuirana slučajna varijabla s parametrima μ i $\sigma^2 = 0.0064$, testirajte hipotezu $H_0 : \mu = 1$, uz alternativnu hipotezu $H_1 : \mu < 1$. Razina značajnosti testa $\alpha = 0.01$.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
0.91	0.92	0.90	0.91	0.91	0.91	0.94	0.91	0.99	0.91

Zadatak 3.

Provjera rada stroja za proizvodnju matica. U tablici su popisana izmjerena odstupanja promjera matica od propisane vrijednosti (izražena u milimetrima). Uz pretpostavku da je vrijednost odstupanja normalno distribuirana slučajna varijabla s parametrima μ i $\sigma^2 = 0.0047$, odgovarajućim testom (ili lijevi ili desni ili obostrani) testirajte hipotezu $H_0 : \mu = 0.26$, uz razinu značajnosti testa $\alpha = 0.005$.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
0.18	0.18	0.19	0.23	0.21	0.29	0.29	0.27

19 χ^2 test

Nakon što odredimo tabelu frekvencija nekih dobivenih podataka možemo se zapitati odgovara li ona nekim unaprijed poznatim očekivanim frekvencijama, koje se zadaju preko distribucije neke slučajne varijable.

Npr. ako bacimo simetričnu kockicu 60 puta, zbog jednakih vjerojatnosti elementarnih događaja očekujemo da će svi brojevi od 1 do 6 pasti podjednaki broj puta, tj. približno 10 puta. Najčešće to neće biti tako i neki brojevi će se pojaviti rjeđe, a neki češće. Ono što nas zanima je, da li je odstupanje od očekivane frekvencije zanemarivo, tj. nebitno (slučajno), ili su odstupanja takva da bismo trebali posumnjati u simetričnost kockice.

χ^2 test je metoda uspoređivanja frekvencija kojom se provjerava razlikuju li se bitno primijećene frekvencije od očekivanih. Test se koristi za testiranje nul hipoteze koja pretpostavlja da nema bitne razlike između primijećenih i očekivanih rezultata. Nul hipoteza se ili prihvaca ili ne privaća nakon uspoređivanja vrijednosti χ^2 testa s odgovarajućom kritičnom vrijednosti. Što je vrijednost χ^2 testa veća, to je veća vjerojatnost da se nul hipoteza neće prihvati pri čemu se onda smatra da je i neki drugi faktor, a ne samo slučaj, uzrokovao velika odstupanja između očekivanih i primijećenih rezultata.

Neka su eksperimentima dobiveni rezultati nekog pokusa (tj. vrijednosti neke slučajne varijable X) i njihove frekvencije zapisani u ovakvoj tabeli:

X	x_1	x_2	x_3	\cdots	x_{m-1}	x_m
f	f_1	f_2	f_3	\cdots	f_{m-1}	f_m

(13)

Prepostavljamo da slučajna varijabla X čije ishode smo bilježili u eksperimentima ima distribuciju

$$\begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_{m-1} & x_m \\ p_1 & p_2 & p_3 & \cdots & p_{m-1} & p_m \end{pmatrix} \quad (14)$$

Tada, ako je $n = f_1 + f_2 + \cdots + f_m$ dovoljno velik, tj. ako je $np_j \geq 5$ za svaki j , možemo primijeniti χ^2 test. Njime provjeravamo nultu hipotezu (H_0) koja glasi: distribucija mjerene veličine X dane eksperimentalnim podacima u tabeli (13) je jednaka teoretskoj distribuciji danoj tabelom (14). Formalnije rečeno, nulta hipoteza glasi:

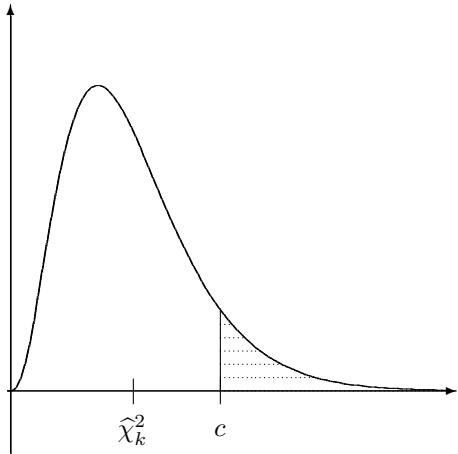
$$H_0 : P(X = x_1) = p_1, P(X = x_2) = p_2, \dots, P(X = x_m) = p_m, \text{ tj.}$$

$$H_0 : X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_{m-1} & x_m \\ p_1 & p_2 & p_3 & \cdots & p_{m-1} & p_m \end{pmatrix}$$

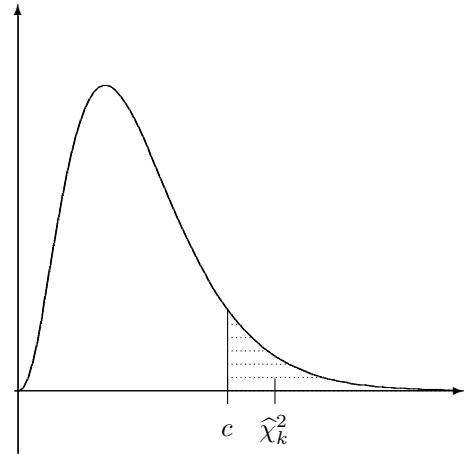
Da bismo to provjerili sa razinom značajnosti α (obično je $\alpha = 0.05$, $\alpha = 0.01$ ili $\alpha = 0.1$), izračunavamo vrijednost χ^2 testa $\hat{\chi}_k^2$, pri čemu je k broj stupnjeva slobode: $k = m - 1$.

$$\hat{\chi}_k^2 = \frac{(f_1 - np_1)^2}{np_1} + \frac{(f_2 - np_2)^2}{np_2} + \cdots + \frac{(f_m - np_m)^2}{np_m}. \quad (15)$$

U tabeli na stranici 128 tražimo kritičnu vrijednost c tako da je $P\{\chi_k^2 \leq c\} = 1 - \alpha =$ u presjeku retka koji počinje stupnjem slobode k i stupca koji počinje s $1 - \alpha$. Ako je $\hat{\chi}_k^2 < c$ onda test potvrđuje nulhipotezu s razinom značajnosti α , a ako je $\hat{\chi}_k^2 \geq c$, onda je $\hat{\chi}_k^2$ u kritičnom području i test odbacuje hipotezu s razinom značajnosti α .



nul hipoteza se prihvaca



nul hipoteza se odbacuje

Primjer 1.

Igrača kockica je bačena 60 puta i dobiveni su sljedeći rezultati:

X	1	2	3	4	5	6
f	5	7	5	14	13	16

Želimo provjeriti je li kockica simetrična (odnosno "poštena") ili imamo razloga posumnjati u tu tvrdnju. To znači da želimo provjeriti odgovaraju li dobiveni brojevi teoretskoj distribuciji za simetričnu kockicu danu tabelom

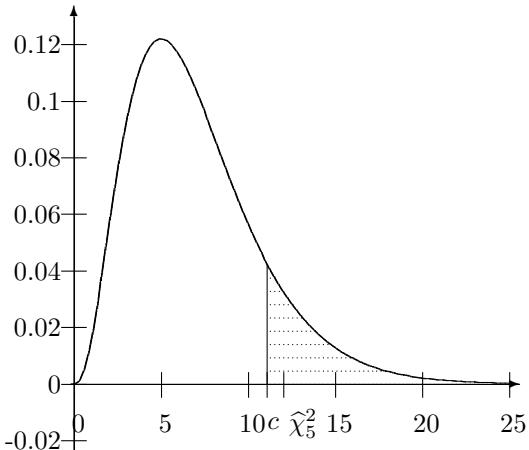
$$\left(\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{array} \right)$$

Nulta hipoteza H_0 : kockica je "poštena", tj. distribucija vrijednosti koje se mogu dobiti njenim bacanjem je jednaka teoretskoj.

Da bismo to provjerili primijenit ćemo χ^2 test s razinom značajnosti $\alpha = 0.05$. Imamo $np_1 = \dots = np_6 = 10$ (što je veće od 5 pa možemo primijeniti χ^2 test), broj stupnjeva slobode je $k = 6 - 1 = 5$ i

$$\hat{\chi}_5^2 = \frac{(5-10)^2}{10} + \frac{(7-10)^2}{10} + \frac{(5-10)^2}{10} + \frac{(14-10)^2}{10} + \frac{(13-10)^2}{10} + \frac{(16-10)^2}{10} = 12$$

U tabeli na stranici 128 tražimo vrijednost c takvu da je $P\{\chi_5^2 \leq c\} = 1 - 0.05 = 0.95$ u presjeku retka koji počinje stupnjem slobode $k = 5$ i stupca koji počinje s 0.95. Očitavamo $c = 11.07$. Jer je $\hat{\chi}_5^2 = 12 \geq 11.07 = c$, tj. jer je $\hat{\chi}_5^2$ u kritičnom području, hipoteza se *odbacuje* s razinom značajnosti $\alpha = 0.05$.



p	0.005	0.001	0.025	0.050	0.100	0.900	0.950	0.975	0.990	0.995
m										
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.002	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.024	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.091	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.210	0.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	0.676	0.381	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	0.599	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278

Primjer 2.

Igrača kockica je bačena 60 puta i dobiveni su sljedeći rezultati:

X	1	2	3	4	5	6
f	11	6	14	13	5	11

Zadatak je analogan prethodnom pa ćemo iskoristiti izračunatu vrijednost za c iz prethodnog zadatka; trebamo provjeriti je li kockica simetrična (odnosno "poštena") ili imamo razloga posumnjati u tu tvrdnju. To znači da želimo provjeriti odgovaraju li dobiveni brojevi teoretskoj distribuciji za simetričnu kockicu. Razina značajnosti je $\alpha = 0.05$.

$$\hat{\chi}_5^2 = \frac{(11-10)^2}{10} + \frac{(6-10)^2}{10} + \frac{(14-10)^2}{10} + \frac{(13-10)^2}{10} + \frac{(5-10)^2}{10} + \frac{(11-10)^2}{10} = 6.8$$

Jer je $\hat{\chi}_5^2 < 11.07 = c$, hipotezu ćemo prihvati s razinom značajnosti od $\alpha = 0.05$.

Zadatak 1.

Igrača kockica je bačena 120 puta i dobiveni su sljedeći rezultati:

X	1	2	3	4	5	6
f	20	14	23	12	26	25

S razinom značajnosti $\alpha = 0.05$ testirajte hipotezu da je kockica simetrična.

Primjer 3.

Novčić je bačen 100 puta i 60 puta je ispala glava. Možemo li s razinom značajnosti $\alpha = 0.05$ zaključiti da je novčić simetričan ili ne?

Tabela koja pripada eksperimentu je

X	pismo	glava
f	40	60

Nulta hipoteza je da se dani podaci slažu s teoretskom distribucijom za bacanje simetričnog novčića:

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

konstatiramo da smijemo provesti χ^2 test, jer je $100 \cdot \frac{1}{2} = 50 > 5$. Broj stupnjeva slobode je $k = 2 - 1 = 1$. Sada računamo $\hat{\chi}_1^2$:

$$\hat{\chi}_1^2 = \frac{(40 - 50)^2}{50} + \frac{(60 - 50)^2}{50} = 4$$

Iz tabele vidimo da je za $P\{\chi_1^2 \leq c\} = 0.95$ $c = 3.841$ i kako je $\hat{\chi}_1^2 > c$, zaključujemo da s razinom značajnosti $\alpha = 0.05$ nultu hipotezu odbacujemo, tj. novčić s razinom značajnosti $\alpha = 0.05$ nije simetričan.

Primjer 4.

U ljevaonici Irongate su se bilježile nesreće koje su se događale tijekom radnog vremena po satima smjene. Izvještaj o nesrećama je dan sljedećom tabelom u kojoj je zabilježeno $n = 168$ nesreća:

sat smjene	1	2	3	4	5	6	7	8
broj nesreća	19	17	15	24	20	26	22	25

Ono što zanima i upravu ljevaonice i sindikat radnika ljevaonice je može li se na osnovu danih podataka ustanoviti da postoji kritični sat smjene tijekom kojeg se nesreće događaju češće nego tijekom drugih sati. Ako takav sat ne postoji, onda bi nesreće trebale biti ravnomjerno raspoređene po satima smjene, tj. trebale bi imati sljedeću distribuciju:

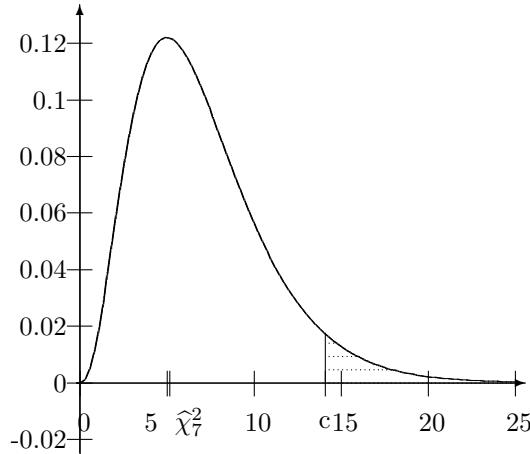
$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \frac{1}{8} & \frac{1}{8} \end{pmatrix}$$

Kako je $168 \cdot \frac{1}{8} = 21 > 5$, možemo primijeniti χ^2 test kojim ćemo uz vrijednost razine značajnosti $\alpha = 0.05$ provjeriti nul hipotezu da su vjerojatnosti događanja nesreća tijekom različitih sati smjene međusobno jednake. Broj stupnjeva slobode je jednak broju sati u smjeni umanjen za jedan, tj. 7. Izračunajmo vrijednost testa:

$$\hat{\chi}_7^2 = \frac{(19 - 21)^2}{21} + \frac{(17 - 21)^2}{21} + \dots + \frac{(22 - 21)^2}{21} + \frac{(25 - 21)^2}{21} \approx 5.143$$

Kritičnu vrijednost c za broj stupnjeva slobode 7 i razinu značajnosti $\alpha = 0.05$ očitavamo iz tabele i dobivamo: $c = 14.067$.

Kako je $\hat{\chi}_7^2 < c$, onda zaključujemo da test potvrđuje nul hipotezu s razinom značajnosti $\alpha = 0.05$.



nul hipoteza se prihvaca

Primjer 5.

Provjerom su na uzorku od 556 sjemenaka, koje mogu biti po boji žute ili zelene, a po obliku glatke ili naborane, dobiveni podaci predstavljeni sljedećom tabelom⁸:

vrsta sjemenki frekvencije	žute glatke 315	zelene glatke 108	žute naborane 101	zelene naborane 32
-------------------------------	--------------------	----------------------	----------------------	-----------------------

χ^2 testom uz razinu značajnosti $\alpha = 0.05$ testirajte nul hipotezu prema kojoj bi se ove sjemenke trebale pojaviti u omjeru $9 : 3 : 3 : 1$.

Navedeni teoretski omjer se može predstaviti distribucijom

$$\left(\begin{array}{cccc} \text{žute glatke} & \text{zelene glatke} & \text{žute naborane} & \text{zelene naborane} \\ \frac{9}{16} & \frac{3}{16} & \frac{3}{16} & \frac{1}{16} \end{array} \right)$$

Budući je $556 \cdot \frac{1}{16} = 34.75 > 5$, možemo primjeniti χ^2 test. Broj stupnjeva slobode je 3. Vrijednost testa je:

$$\hat{\chi}^2_3 = \frac{(315 - 312.75)^2}{312.75} + \frac{(108 - 104.25)^2}{104.25} + \frac{(101 - 104.25)^2}{104.25} + \frac{(32 - 34.75)^2}{34.75} \approx 0.47$$

Kritična vrijednost za broj stupnjeva slobode 3 i razinu značajnosti $\alpha = 0.05$ je $c = 7.815$ i kako je $\hat{\chi}^2_3 < c$, zaključujemo da test potvrđuje nul hipotezu s razinom značajnosti $\alpha = 0.05$.

⁸Uz pomoć eksperimenata koji su uključivali brojanje različitih vrsta sjemenaka poput ovih iz zadatka je ruski biolog Mendel došao do svojih zakona o nasljedivanju koji nose njegovo ime.

Zadaci

Zadatak 1. Tijekom tri uzastopna tjedna jedne godine su dobiveni sljedeći podaci o ukupno 66 ozljeda uzrokovanih ugrizima životinja u nekoj bolnici:

tjedan	25. – 31. kolovoza	1. – 7. rujna	8. – 14. rujna
broj ugriza	27	19	20

χ^2 testom uz vrijednost razine značajnosti $\alpha = 0.1$ provjerite je li jedan od promatranih tjedana bio bitno opasniji što se životinjskih ugriza tiče od ostala dva.

Zadatak 2. Tijekom jedne dulje DnD igre je oktaedarska kockica (ili kraće $d8$; to je oktaedar, pravilno poliedarsko tijelo s 8 strana na kojima su napisani brojevi od 1 do 8) bačena ukupno 168 puta. Sljedećom tabelom su prikazani brojevi koji su dobiveni na taj način:

$d8$	1	2	3	4	5	6	7	8
frekvencije	29	22	18	19	20	23	12	25

χ^2 testom uz vrijednost razine značajnosti $\alpha = 0.01$ provjerite ima li razloga sumnjati da kockica nije simetrična.

Zadatak 3. Sto slučajno odabranih studenata je pitano koji je od brojeva 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 njima najdraži. Rezultati ankete su prikazani u sljedećoj tabeli:

najdraži brojevi	0	1	2	3	4	5	6	7	8	9
frekvencije	5	3	11	10	19	9	11	15	13	4

χ^2 testom uz vrijednost razine značajnosti $\alpha = 0.05$ provjerite postoje li brojevi koje studenti vole više od ostalih.

Zadatak 4. U trgovini odjećom su dobili sljedeće rezultate o kupovini 40 kravata koje se međusobno razlikuju samo po boji:

boja kravate	plava	crvena	zelena	žuta
frekvencije	7	9	14	10

χ^2 testom uz vrijednost razine značajnosti $\alpha = 0.01$ provjerite prodaju li se kravate nekih boja više od ostalih.

Zadatak 5. Iz telefonskog imenika su uzete zadnje znamenke sto uzastopnih telefonskih brojeva i dobivena je sljedeća tabela:

zadnja znamenka tel. broja	0	1	2	3	4	5	6	7	8	9
frekvencije	3	8	15	14	10	7	8	9	11	15

χ^2 testom uz vrijednost razine značajnosti $\alpha = 0.01$ provjerite je li uzimanje zadnjih znamenaka brojeva iz telefonskog imenika dobar način da se dobiju slučajni brojevi.

Zadatak 6. Proizvođač loptica za stolni tenis pakira po pet loptica u jednu kutiju. Loptice mogu biti ili ispravne ili neispravne. Radi kontrole kvalitete proizvoda i načina pakiranja loptica potrebno je odrediti teoretsku distribuciju pakiranja neispravnih loptica po kutijama. Na slučajan je način odabранo 70 kutija za koje su dobiveni sljedeći podaci

broj neispravnih loptica x_i	0	1	2	3	4	5
broj kutija f_i	33	6	6	5	9	11

Na ovom uzorku testirajte nulhipotezu da broj neispravnih loptica po kutiji ima teoretsku distribuciju

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ \frac{4}{18} & \frac{1}{9} & \frac{3}{9} & \frac{3}{18} & \frac{1}{18} & \frac{4}{18} \end{pmatrix}$$

uz razinu značajnosti $\alpha = 0.01$.

Zadatak 7. Proizvođač kockica za iganje je proizveo novu kockicu koja ima sedam strana na kojoj mogu pasti brojevi od 1 do 7 i tvrdi da svaki od brojeva ima jednaku vjerojatnost pojavljivanja. Bacanjem kockice je dobivena sljedeća tabela frekvencija:

broj na kockici	1	2	3	4	5	6	7
frekvencija	8	14	17	11	13	18	19

Na ovim podacima testirajte tvrdnju proizvođača, tj. testirajte nulhipotezu da je navedena kockica sa sedam strana simetrična, tj. ima teoretsku razdiobu

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ p & p & p & p & p & p & p \end{pmatrix}$$

(prvo odredite p) uz razinu značajnosti $\alpha = 0.01$.

Zadatak 8. U kockarnici se može igrati jednostavna igra s jednim pravilom: bacaju se tri kockice i dobitak je proporcionalan broju šestica koje na njima padnu. Praćenjem jednog od igrača dobivena je sljedeća tabela:

broj šestica na tri kockice	0	1	2	3
frekvencije	48	35	15	3

Upravi kockarnice su se navedeni podaci učinili sumnjivim i žele provjeriti igrač li taj igrač s njihovim kockicama koje su sigurno simetrične, ili ih je on neprimjetno zamijenio sa svojim, nesimetričnim kockicama koje favoriziraju padanje šestica. Vjerojatnost p_k padanja k šestica na tri kockice je jednaka $p_k = P(X = k)$, gdje je $X \sim B(3, \frac{1}{6})$.

Na danim podacima testirajte nulhipotezu da igrač igra sa simetričnim kockicama, tj. da dobiveni podaci imaju teoretsku razdiobu

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ p_0 & p_1 & p_2 & p_3 \end{pmatrix}$$

uz razinu značajnosti $\alpha = 0.05$. Na osnovu rezultata zaključite hoće li uprava ljubazno zamoliti tog igrača da napusti kockarnicu ili neće reagirati na njegove uspjehe.

U dosadašnjim primjerima smo eksperimentalne podatke testirali prema teoretskim (očekivanim) vrijednostima. To se, naravno, može raditi jedino ako je teoretska distribucija poznata. Sada ćemo vidjeti primjere u kojima se testiraju dvije grupe eksperimentalnih podataka jedna prema drugoj, tj. provjeravat ćemo imaju li dvije grupe podataka istu (ne nužno poznatu) distribuciju.

Prepostavimo da imamo dvije grupe eksperimentalnih podataka o svim mogućim rezultatima nekog eksperimenta i njihove frekvencije:

X	x_1	x_2	x_3	\dots	x_{r-1}	x_r	
D_1	m_1	m_2	m_3	\dots	m_{r-1}	m_r	$m = m_1 + m_2 + \dots + m_r$
D_2	n_1	n_2	n_3	\dots	n_{r-1}	n_r	$n = n_1 + n_2 + \dots + n_r$
	$s_1 = m_1 + n_1$	s_2	s_3	\dots	s_{r-1}	$s_r = m_r + n_r$	$N = s_1 + s_2 + \dots + s_r = m + n$

Nula hipoteza koju testiramo (s nekom razinom značajnosti α) je da dvije grupe podataka imaju istu distribuciju. Postupak je isti kao i u slučaju testiranja prema teoretskoj distribuciji, jedino se $\hat{\chi}_k^2$ drugačije računa. Prvo računamo vrijednosti p_j , $j = 1, \dots, r$:

$$p_1 = \frac{m_1 + n_1}{N} = \frac{s_1}{N}, p_2 = \frac{s_2}{N}, \dots, p_r = \frac{s_r}{N}.$$

Zatim računamo vrijednosti mp_j i np_j , $j = 1, \dots, r$ i na kraju računamo $\hat{\chi}_k^2$; broj stupnjeva slobode je $k = r - 1$.

$$\hat{\chi}_k^2 = \frac{(m_1 - mp_1)^2}{mp_1} + \dots + \frac{(m_r - mp_r)^2}{mp_r} + \frac{(n_1 - np_1)^2}{np_1} + \dots + \frac{(n_r - np_r)^2}{np_r} \quad (16)$$

Sada gledamo u istu tabelu kao i prije i tražimo takav c da vrijedi $P\{\chi_k^2 \leq c\} = 1 - \alpha$ i ako je $c \leq \hat{\chi}_k^2$ onda je $\hat{\chi}_k^2$ u kritičnom području pa hipotezu odbacujemo, a ako je $c > \hat{\chi}_k^2$, onda je prihvaćamo s razinom značajnosti α .

Primjer 6. Učenici neke škole su postigli sljedeće uspjehe:

ocjena	1	2	3	4	5	
dječaci	40	48	85	42	35	$250 = m$
djevojčice	20	35	77	50	28	$210 = n$
	60	83	162	92	63	460

Željeli bismo ustanoviti jesu li dječaci i djevojčice bili jednakopravni, tj. jesu li distribucije tih dvaju skupova podataka jednake. Idemo redom. Broj stupnjeva slobode je $k = 5 - 1 = 4$.

$$p_1 = \frac{40 + 20}{460} = \frac{60}{460}, p_2 = \frac{83}{460}, p_3 = \frac{162}{460}, p_4 = \frac{92}{460}, p_5 = \frac{63}{460}.$$

$$mp_1 = 250 \cdot \frac{60}{460} \approx 32.61, \quad mp_2 \approx 45.11, \quad mp_3 \approx 88.04, \quad mp_4 = 50, \quad mp_5 \approx 34.24,$$

$$np_1 = 210 \cdot \frac{60}{460} \approx 27.39, \quad np_2 \approx 37.89, \quad np_3 \approx 73.96, \quad np_4 = 42, \quad np_5 \approx 28.76.$$

$$\begin{aligned} \hat{\chi}_4^2 &\approx \frac{(40 - 32.61)^2}{32.61} + \frac{(48 - 45.11)^2}{45.11} + \frac{(85 - 88.04)^2}{88.04} + \frac{(42 - 50)^2}{50} + \frac{(35 - 34.24)^2}{34.24} + \\ &\frac{(20 - 27.39)^2}{27.39} + \frac{(35 - 37.89)^2}{37.89} + \frac{(77 - 73.96)^2}{73.96} + \frac{(50 - 42)^2}{42} + \frac{(28 - 28.76)^2}{28.76} = \\ &\approx 7.14 \end{aligned}$$

U tabeli nalazimo c takav da vrijedi $P\{\chi_4^2 \leq c\} = 0.95$: $c = 9.488$. Kako je $c > \hat{\chi}_4^2$, hipotezu prihvaćamo s razinom značajnosti $\alpha = 0.05$.

Zadaci

Zadatak 9.

Sto ljudi je dobilo cjepivo protiv gripe. Njih 40 je bilo mlađe od 50 godina, a 60 starije. Neki od njih nisu dobili gripu (N), neki su dobili blagi oblik (B), a neki jaki oblik (J) te gripe.

oblik gripe	N	B	J	
< 50	30	6	4	40
≥ 50	36	12	12	60

Koristeći χ^2 test uz razinu značajnosti $\alpha = 0.05$ testirajte nul hipotezu da su reakcije na cjepivo bile jednake u obje grupe.

Rješenje.

$\hat{\chi}_2^2 \approx 2.65$, $P\{\chi_2^2 \leq c\} = 0.95 \Rightarrow c \approx 5.99$ Hipoteza se ne odbacuje, jer je $\hat{\chi}_2^2 < c$.

Zadatak 10.

Ishodi dviju slučajnih varijabli X i Y su A, B, C i D . Mjerenjem se došlo do sljedećih podataka o frekvenciji ishoda tih dviju varijabli:

	A	B	C	D
X	24	46	14	17
Y	43	51	36	11

Koristeći χ^2 test uz razinu značajnosti $\alpha = 0.05$ testirajte nul hipotezu da slučajne varijable X i Y imaju istu distribuciju. Provedite test uz vrijednost razine značajnosti od $\alpha = 0.01$.

Zadatak 11.

U tabeli su popisane frekvencije ishoda dviju slučajnih varijabli, X i Y .

	2	3	4	5	6	7	8	9	10	11	12
X	1	6	7	14	18	18	9	10	10	5	2
Y	1	2	4	4	16	11	17	11	15	8	11

Koristeći χ^2 test uz razinu značajnosti $\alpha = 0.05$ testirajte nul hipotezu da su da slučajne varijable X i Y imaju istu distribuciju.

VJEROJATNOST I STATISTIKA

PRIMJER 3. KOLOKVIJA

Napomena: sve decimalne brojeve izračunajte na četiri decimale točno, uz zaokruživanje, ako je to potrebno.

Za rješavanje ovih zadataka su potrebne odgovarajuće tabele.

1. Pretpostavimo da je masa čipsa kojeg stroj pakira u jednu vrećicu normalno distribuirana slučajna varijabla s parametrima $\mu = 137\text{g}$ i $\sigma^2 = 81\text{g}^2$. Izračunajte vjerojatnost da je masa čipsa u slučajno odabranoj vrećici
 - (a) veća od 146g ($X > 146$);
 - (b) između 134g i 143g ($134 < X \leq 143$).
2. Provjerjen je rad stroja za proizvodnju matica. U tabeli su popisana izmjerena odstupanja promjera matica od propisane vrijednosti (u mm). Uz pretpostavku da je vrijednost odstupanja normalno distribuirana slučajna varijabla s parametrima μ i $\sigma^2 = 0.0036$, odgovarajućim testom testirajte hipotezu $H_0 : \mu = 0.27$ uz razinu značajnosti $\alpha = 0.02$. Navedite alternativnu hipotezu H_1 .

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
0.3	0.24	0.28	0.24	0.28	0.26	0.29	0.26	0.28	0.26

3. Proizvođač loptica za stolni tenis pakira po pet loptica u jednu kutiju. Loptice mogu biti ili ispravne ili neispravne. Radi kontrole kvalitete proizvoda i načina pakiranja loptica potrebno je odrediti teoretsku razdiobu pakiranja neispravnih loptica po kutijama. Na slučajan način su odabrane 102 kutije za koje su dobiveni sljedeći podaci

broj neispravnih loptica x_i	0	1	2	3	4	5
broj kutija f_i	32	14	20	11	10	15

Na ovom uzorku testirajte nulhipotezu da broj neispravnih loptica po kutiji ima **teoretsku razdiobu**

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ \frac{11}{24} & \frac{1}{6} & \frac{5}{24} & \frac{1}{12} & \frac{1}{24} & \frac{1}{24} \end{pmatrix}$$

uz razinu značajnosti $\alpha = 0.01$. (Uputa: χ^2 -test, teoretska razdioba zadana.)

SKICE RJEŠENJA

1. (čips u vrećicama)

$$X \sim N(137, 81), \sigma = \sqrt{81} = 9$$

$$\begin{aligned} \text{(a)} \quad P(X > 146) &\approx P(X^* > 1) = 1 - P(X^* < 1) = \\ &= 1 - F^*(1) \approx 1 - 0.84134 = 0.15866 \end{aligned}$$

$$\begin{aligned} \text{(b)} \quad P(134 < X < 143) &\approx P(-0.33 < X^* < 0.67) = F^*(0.67) - F^*(-0.33) = \\ &= F^*(0.67) - (1 - F^*(0.33)) \approx 0.74857 - (1 - 0.6293) \approx 0.74857 - 0.3707 \approx 0.37787 \end{aligned}$$

2. Provodimo obostrani test.

$$\text{vrijednost testa } vt = \frac{0.269 - 0.27}{0.06} \sqrt{10} \approx -0.0527046$$

alternativna hipoteza je $H_1 : \mu \neq 0.27$.

kritične vrijednosti: $z_{0.01} = -2.33$, $z_{0.99} = 2.33$.

Nul hipoteza se prihvata s razinom značajnosti $\alpha = 0.02$ zato što vrijedi $z_{\frac{\alpha}{2}} < vt < z_{1-\frac{\alpha}{2}}$ (tj. vrijedi $-2.33 < -0.0527046 < 2.33$).

3. (teniske loptice)

Broj stupnjeva slobode je: 5.

Vrijednost testa je jednaka: $\hat{\chi}_5^2 = 40.9626$.

Kritična vrijednost je jednaka: $c = 15.086$.

Budući je vrijednost testa veća od ili jednaka kritičnoj vrijednosti, nul hipotezu ne prihvaćamo s razinom značajnosti $\alpha = 0.01$.

VJEROJATNOST I STATISTIKA

PRIMJER PISMENOG ISPITA

Napomena: sve decimalne brojeve izračunajte na četiri decimale točno, uz zaokruživanje, ako je to potrebno.

Za rješavanje nekih zadataka su potrebne odgovarajuće tabele.

- (a) U posudi se nalazi deset bijelih, devet žutih i jedanaest zelenih kuglica. Iz nje se bez vraćanja na slučajan način izvlači devet kuglica (tj. tako da sve kuglice u posudi imaju istu vjerojatnost da će biti izvučene).

Izračunajte vjerojatnost da će njih osam biti iste boje i jedna neke druge boje.

- (b) Bacamo dvije simetrične kockice, crvenu i plavu. Provjerite jesu li nezavisni sljedeći događaji:

A = suma brojeva palih na kockicama je manja od 11,

B = na plavoj kockici je pao broj strogo veći od broja koji je pao na crvenoj kockici.

- Bacamo standardnu simetričnu kockicu sa šest strana.

(a) Izračunajte vjerojatnost da će u nizu od 9 bacanja kockice šestica pasti točno 6 puta.

(b) Izračunajte vjerojatnost da će u nizu od 8 bacanja kockice šestica pasti najviše 5 puta.

U oba slučaja napišite koji događaj smatrati uspjehom i koja je distribucija slučajne varijable koju koristite u rješenju.

- Za slučajnu varijablu X se zna da ima distribuciju oblika

$$X \sim \left(\begin{array}{cccccc} -4 & -2 & a & 1 & 4 & 7 \\ 2p & 3p & 9p & \frac{2}{17} & 8p & 8p \end{array} \right)$$

i očekivanje $E(X) = \frac{39}{17}$. Odredite nepoznanice a i p , napišite distribuciju slučajne varijable X i izračunajte njenu standardnu devijaciju.

Izračunajte $P(\mu - \sigma \leq X \leq \mu + \sigma)$, (pri čemu je $\mu = E(X)$ i $\sigma = \sigma(X)$).

- Pretpostavimo da je visina (u cm) studenata vеleučilišta normalno distribuirana slučajna varijabla s parametrima $\mu = 170$ cm i $\sigma^2 = 39.69$ cm². Izračunajte vjerojatnost da je visina slučajno odabranog studenta vеleučilišta

(a) veća od 172 cm;

(b) između 169 cm i 173 cm.

- Igrača kockica sa 6 strana služi kao generator uniformno distribuiranih prirodnih brojeva od 1 do 6. Jedna takva kockica je baćena 180 puta i dobiveni brojevi su predstavljeni sljedećom tabelom frekvencija:

X	1	2	3	4	5	6
f	41	28	38	26	30	17

Provjerite je li kockica zaista generator uniformno distribuiranih brojeva, tj. je li ona simetrična ("poštena") koristeći χ^2 test s razinom značajnosti $\alpha = 0.005$. Uputa: provjerite odgovaraju li dobivene frekvencije uniformnoj distribuciji

$$\left(\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{array} \right)$$

SKICE RJEŠENJA

1. (a) (10 bijelih, 9 žutih, 11 zelenih; sve izvučene kuglice osim jedne su iste boje)
 Broj načina da od deset kuglica bijele boje izvučemo njih osam i još jednu od kuglica preostalih boja je $\binom{10}{8} \binom{9+11}{1} = 45 \cdot 20 = 900$.
 Broj načina da od devet kuglica žute boje izvučemo njih osam i još jednu od kuglica preostalih boja je $\binom{9}{8} \binom{11+10}{1} = 9 \cdot 21 = 189$.
 Broj načina da od jedanaest kuglica zelene boje izvučemo njih osam i još jednu od kuglica preostalih boja je $\binom{11}{8} \binom{10+9}{1} = 165 \cdot 19 = 3135$.
 Tražena vjerojatnost je $\frac{900+189+3135}{\binom{30}{9}} = \frac{4224}{14307150} \approx 0.000295237$.

(b) Događaji A i B se mogu predstaviti ovako:

$c \setminus p$	1	2	3	4	5	6	$P(A)$
1	A	AB	AB	AB	AB	AB	$\frac{11}{12}$
2	A	A	AB	AB	AB	AB	$\frac{5}{12}$
3	A	A	A	AB	AB	AB	
4	A	A	A	A	AB	AB	
5	A	A	A	A	A	B	
6	A	A	A	A			

$$P(A \cap B) = \frac{7}{18}$$

Jer vrijedi $P(A) \cdot P(B) = \frac{55}{144} \neq \frac{7}{18} = P(A \cap B)$, događaji A i B nisu nezavisni.

2. (simetrična kockica sa šest strana, $p = 0.166667$)

(a) $X \sim B(9, 0.166667)$

$$P(X = 6) = \binom{9}{6} \cdot 0.166667^6 \cdot 0.833333^3 \approx 84 \cdot 2.14335e - 05 \cdot 0.578704 \approx 0.0010419$$

(b) $X \sim B(8, 0.166667)$, uspjeh je ako padne šestica

Treba izračunati vjerojatnost da će u seriji od 8 pokusa biti najviše 5 uspjeha.

$$P(X \leq 5) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$$

$$P(X \leq 5) = 1 - (P(X = 6) + P(X = 7) + P(X = 8))$$

$$P(X \leq 5) \approx 0.999559$$

3. Suma svih vjerojatnosti u drugom retku distribucije mora biti jednaka 1 pa dobivamo jednadžbu $2p + 3p + 9p + \frac{2}{17} +$

$$8p + 8p = 1 \text{ kojoj je rješenje } p = \frac{1}{34}. \text{ Iz formule za očekivanje dobivamo jednadžbu}$$

$$E(X) = \frac{1}{17} \cdot (-4) + \frac{3}{34} \cdot (-2) + \frac{9}{34} \cdot a + \frac{2}{17} \cdot 1 + \frac{4}{17} \cdot 4 + \frac{4}{17} \cdot 7 = \frac{39}{17} \text{ kojoj je rješenje } a = 0.$$

$$E(X^2) = \frac{284}{17} \approx 16.7059, E(X) = \frac{39}{17} \approx 2.29412$$

$$Var(X) = E(X^2) - E(X)^2 \approx 16.7059 - 5.26298 = 11.4429, \boxed{\sigma(X) = \sqrt{Var(X)} \approx 3.38274}$$

$$\mu = 2.29412, \sigma = 3.38274 \Rightarrow$$

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = P(-1.08862 \leq X \leq 5.67685) = \frac{9}{34} + \frac{2}{17} + \frac{4}{17} = \frac{21}{34} \approx \boxed{0.617647}$$

4. visine studenata, $X \sim N(170, 39.69)$, $\sigma = \sqrt{39.69} = 6.3$

$$(a) P(X > 172) \approx P(X^* > 0.32) = 1 - P(X^* < 0.32) = \\ = 1 - F^*(0.32) \approx 1 - 0.62552 = 0.37448$$

$$(b) P(169 < X < 173) \approx P(-0.16 < X^* < 0.48) = F^*(0.48) - F^*(-0.16) = \\ = F^*(0.48) - (1 - F^*(0.16)) \approx 0.68439 - (1 - 0.56356) \approx 0.68439 - 0.43644 \approx 0.24795$$

5. Broj stupnjeva slobode je: 5.

Vrijednost testa je jednaka: $\hat{\chi}_5^2 = 12.4667$.

Kritična vrijednost je jednaka: $c = 16.75$.

Budući je vrijednost testa manja od kritične vrijednosti, nul hipotezu prihvaćamo s razinom značajnosti $\alpha = 0.005$.

20 Rješenja zadataka

20.1 Događaj. Vjerojatnost

Zadatak 1.

Napišite prostor elementarnih događaja Ω koji sadrži sve ishode bacanja triju novčića. Odredite sljedeće događaje:

$A =$ pala su barem dva pisma

$B =$ pale su najviše dvije glave

$A \cap B = ?$

$A \cup B = ?$

Rješenje.

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

$$A = \{(P, P, G), (P, G, P), (G, P, P), (P, P, P)\}$$

$$B = \{(P, P, P), (P, P, G), (P, G, P), (P, G, G), (G, P, P), (G, P, G), (G, G, P)\}$$

vidi se da je $A \subseteq B$ pa je

$$A \cap B = A$$

$$A \cup B = B$$

Zadatak 3.

U posudi se nalazi pet zelenih i dvije žute kuglice. Slučajno se, bez vraćanja, jedna po jedna izvlače dvije kuglice. Odredite prostor elementarnih događaja za navedeni slučajni pokus. Kako bi izgledao prostor elementarnih događaja ako bismo svaku kuglicu vratili natrag u posudu odmah nakon njenog izvlačenja?

Rješenje. Skup svih kuglica ćemo označiti s $K = \{z_1, z_2, z_3, z_4, z_5, \check{z}_1, \check{z}_2\}$. Prostor elementarnih događaja se sastoji od svih uređenih parova međusobno različitih elemenata skupa K , tj.

$$\Omega = K \times K \setminus \{(z_1, z_1), (z_2, z_2), (z_3, z_3), (z_4, z_4), (z_5, z_5), (\check{z}_1, \check{z}_1), (\check{z}_2, \check{z}_2)\}.$$

Tabelom prostor elementarnih događaja Ω možemo prikazati ovako:

1.k. \ 2.k.	z_1	z_2	z_3	z_4	z_5	\check{z}_1	\check{z}_2
z_1	—	(z_1, z_2)	(z_1, z_3)	(z_1, z_4)	(z_1, z_5)	(z_1, \check{z}_1)	(z_1, \check{z}_2)
z_2	(z_2, z_1)	—	(z_2, z_3)	(z_2, z_4)	(z_2, z_5)	(z_2, \check{z}_1)	(z_2, \check{z}_2)
z_3	(z_3, z_1)	(z_3, z_2)	—	(z_3, z_4)	(z_3, z_5)	(z_3, \check{z}_1)	(z_3, \check{z}_2)
z_4	(z_4, z_1)	(z_4, z_2)	(z_4, z_3)	—	(z_4, z_5)	(z_4, \check{z}_1)	(z_4, \check{z}_2)
z_5	(z_5, z_1)	(z_5, z_2)	(z_5, z_3)	(z_5, z_4)	—	(z_5, \check{z}_1)	(z_5, \check{z}_2)
\check{z}_1	(\check{z}_1, z_1)	(\check{z}_1, z_2)	(\check{z}_1, z_3)	(\check{z}_1, z_4)	(\check{z}_1, z_5)	—	$(\check{z}_1, \check{z}_2)$
\check{z}_2	(\check{z}_2, z_1)	(\check{z}_2, z_2)	(\check{z}_2, z_3)	(\check{z}_2, z_4)	(\check{z}_2, z_5)	$(\check{z}_2, \check{z}_1)$	—

Ako svaku kuglicu vraćamo u posudu odmah nakon što je izvučemo, prostor elementarnih događaja je skup svih uređenih parova skupa K , tj.

$$\Omega = K \times K.$$

Pripadajuća tabela izgleda ovako:

$1.k. \setminus 2.k.$	z_1	z_2	z_3	z_4	z_5	\check{z}_1	\check{z}_2
z_1	(z_1, z_1)	(z_1, z_2)	(z_1, z_3)	(z_1, z_4)	(z_1, z_5)	(z_1, \check{z}_1)	(z_1, \check{z}_2)
z_2	(z_2, z_1)	(z_2, z_2)	(z_2, z_3)	(z_2, z_4)	(z_2, z_5)	(z_2, \check{z}_1)	(z_2, \check{z}_2)
z_3	(z_3, z_1)	(z_3, z_2)	(z_3, z_3)	(z_3, z_4)	(z_3, z_5)	(z_3, \check{z}_1)	(z_3, \check{z}_2)
z_4	(z_4, z_1)	(z_4, z_2)	(z_4, z_3)	(z_4, z_4)	(z_4, z_5)	(z_4, \check{z}_1)	(z_4, \check{z}_2)
z_5	(z_5, z_1)	(z_5, z_2)	(z_5, z_3)	(z_5, z_4)	(z_5, z_5)	(z_5, \check{z}_1)	(z_5, \check{z}_2)
\check{z}_1	(\check{z}_1, z_1)	(\check{z}_1, z_2)	(\check{z}_1, z_3)	(\check{z}_1, z_4)	(\check{z}_1, z_5)	$(\check{z}_1, \check{z}_1)$	$(\check{z}_1, \check{z}_2)$
\check{z}_2	(\check{z}_2, z_1)	(\check{z}_2, z_2)	(\check{z}_2, z_3)	(\check{z}_2, z_4)	(\check{z}_2, z_5)	$(\check{z}_2, \check{z}_1)$	$(\check{z}_2, \check{z}_2)$

20.2 Aritmetička sredina, mod i medijan

Zadatak 4.

Kad dodamo broj 15 jednom popisu brojeva, aritmetička sredina poraste za 2. Kada dodamo broj 1 ovako povećanom popisu brojeva, aritmetička sredina se smanji za 1. Koliko je brojeva bilo u početnom popisu?

Rješenje. Neka je n broj brojeva u originalnom popisu i \bar{x} njihova aritmetička sredina ($n\bar{x}$ je tada njihova suma). Vrijedi: $\frac{n\bar{x} + 15}{n+1} = \bar{x} + 2$ i $\frac{n\bar{x} + 16}{n+2} = \frac{n\bar{x} + 15}{n+1} - 1 = \bar{x} + 1$. Sada dobivamo: $2n + \bar{x} = 13$ i $n + 2\bar{x} = 14$ iz čega slijedi $n = 4$.

Zadatak 5.

Aritmetička sredina pet nenegativnih cijelih brojeva je 15. Njihov medijan je 18. Koja je najveća moguća vrijednost najvećeg od tih brojeva?

Rješenje. Neka su to brojevi $0 \leq x_1 \leq x_2 \leq x_3 \leq x_4 \leq x_5$. Medijan je srednji broj u tom nizu pa je $x_3 = 18$. Iz formule za aritmetičku sredinu dobivamo:

$$\frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = 15 \Rightarrow x_5 = 75 - (x_1 + x_2 + 18 + x_4) = 57 - (x_1 + x_2 + x_4).$$

Da bi najveći broj x_5 bio što veći, vidimo da svi ostali brojevi trebaju biti što manji. Najmanji nenegativni broj je nula pa svi brojevi manji od medijana trebaju biti jednak nula, tj. $x_1 = x_2 = 0$. Preostali nepoznati broj, x_4 , također mora biti što manji, ali ne smije biti manji od $x_3 = 18$ pa vrijedi $x_4 = x_3 = 18$. Sada x_5 lagano izračunamo:

$$x_5 = 57 - (x_1 + x_2 + x_4) = 57 - (0 + 0 + 18) = 57 - 18 = 39.$$

Najveća moguća vrijednost najvećeg od tih brojeva je 39.

Zadatak 6.

Helena je pet dana svaki dan trčala jedan sat i petnaest minuta i tri dana svaki dan jedan sat i trideset minuta. Koliko vremena bi trebala trčati deveti dan da bi joj prosjek trčanja po danu iznosio 85 minuta?

Rješenje. $\frac{5 \cdot 75 + 3 \cdot 90 + t}{9} = 85 \Rightarrow t = 120$

Dakle Helena bi deveti dan trebala trčati dva sata da bi joj prosjek bio 85 minuta trčanja po danu.

20.3 Uvjetna vjerojatnost

Zadatak 5.

Pretpostavite simetričnu varijantu za sve vjerojatnosti koje nisu određene, npr. vjerojatnosti rođenja djevojčice i dječaka su međusobno jednake, tj. 0.5.

1. Obitelj ima dvoje djece, od kojih je stariji dječak. Izračunajte vjerojatnost da oni imaju dva dječaka.
2. Obitelj ima dvoje djece, od kojih je jedno dječak. Izračunajte vjerojatnost da oni imaju dva dječaka.
3. Obitelj ima dvoje djece, od kojih je jedno dječak koji je rođen u utorak. Izračunajte vjerojatnost da oni imaju dva dječaka.

Rješenje. Rješenje dobivamo nalaženjem prostora elementarnih događaja i pripadajućih događaja te primjenom klasične formule za računanje vjerojatnosti "a priori". Elementarni događaji u prva dva zadatka su uređeni parovi (x, y) , gdje su $x, y \in D = \{d, c\}$, d označava dječaka, a c djevojčicu ("curu"), a prostor elementarnih događaja je $\Omega = D \times D = \{(d, d), (d, c), (c, d), (c, c)\}$. Neka prvi član uređenog para označava dijete koje je starije.

$$1. A = \text{obitelj ima dva dječaka} = \{(d, d)\}, P(A) = \frac{1}{4},$$

$$B = \text{starije dijete je dječak} = \{(d, d), (d, c)\}, P(B) = \frac{1}{2}$$

$$A \cap B = \{(d, d)\} = A, P(A \cap B) = \frac{1}{4}, P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{2}.$$

$$2. A = \text{obitelj ima dva dječaka} = \{(d, d)\}, P(A) = \frac{1}{4},$$

$$B = (\text{barem}) \text{ jedno dijete je dječak} = \{(d, d), (d, c), (c, d)\}, P(B) = \frac{3}{4}$$

$$A \cap B = \{(d, d)\} = A, P(A \cap B) = \frac{1}{4}, P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{3}.$$

3. Neka je $W = \{1, 2, 3, 4, 5, 6, 7\}$ skup koji predstavlja dane u tjednu (1 =ponedjeljak, 2 = utorak, ...). Prostor elementarnih događaja možemo definirati ovako: $\Omega = W \times D \times W \times D$ pa su elementarni događaji četvorke oblika (a_1, d_1, a_2, d_2) , gdje je a_j dan u tjednu kada je rođeno dijete d_j ($j = 1, 2$).

$$A = \text{obitelj ima dva dječaka} = W \times \{d\} \times W \times \{d\}, |A| = |W| \cdot |\{d\}| \cdot |W| \cdot |\{d\}| = 49$$

$$B = \text{jedan dječak je rođen u utorak} = (\{2\} \times \{d\} \times W \times D) \cup (W \times D \times \{2\} \times \{d\}) = B_1 \cup B_2 \quad (\text{primijetite da unija nije disjunktna!})$$

$$|B_1| = |B_2| = |W| \cdot |D| \cdot |\{2\}| \cdot |\{d\}| = 14, B_1 \cap B_2 = \{(2, d, 2, d)\}, |B_1 \cap B_2| = 1 \Rightarrow$$

$$|B| = |B_1 \cup B_2| = |B_1| + |B_2| - |B_1 \cap B_2| = 13,$$

$$A \cap B = (\{2\} \times \{d\} \times W \times \{d\}) \cup (W \times \{d\} \times \{2\} \times \{d\}) = C_1 \cup C_2$$

$$|C_1| = |C_2| = |W| \cdot |\{d\}| \cdot |\{2\}| \cdot |\{d\}| = 7, C_1 \cap C_2 = \{(2, d, 2, d)\} \Rightarrow$$

$$|A \cap B| = |C_1 \cup C_2| = |C_1| + |C_2| - |C_1 \cap C_2| = 13$$

$$\text{Dakle, } P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{13}{|\Omega|}}{\frac{27}{|\Omega|}} = \frac{13}{27} \approx 0.4815.$$

20.4 Nezavisni događaji

Zadatak 4.

Neka su A i B nezavisni događaji. Pokažite da su tada nezavisni i događaji A i \bar{B} .

Rješenje. Događaji A i B su nezavisni pa vrijedi $P(A \cap B) = P(A)P(B)$. Treba dokazati da vrijedi $P(A \cap \bar{B}) = P(A)P(\bar{B})$.

S obzirom da je $A \cap \bar{B} = A \setminus B$ i unija $A = (A \setminus B) \cup (A \cap B)$ je disjunktna (pa vrijedi $P(A) = P((A \setminus B) \cup (A \cap B)) = P(A \setminus B) + P(A \cap B)$), vrijedi:

$$P(A \cap \bar{B}) = P(A \setminus B) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(\bar{B})$$

20.5 Diskretna uniformna distribucija

Zadatak 1.

Napišite distribuciju uniformne slučajne varijable koja može poprimiti sve cijelobrojne vrijednosti od -3 do 7 .

Rješenje.

$$\left(\begin{array}{cccccccccc} -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \frac{1}{11} & \frac{1}{11} \end{array} \right)$$

Zadatak 2.

Pomoću uniformne distribucije (iako je najjednostavnija) se mogu dobiti druge složenije distribucije. Odredite distribuciju slučajne varijable X kojoj su vrijednosti zbroj brojeva koji padnu na dvije bačene standardne kockice (u igri DnD se takva slučajna varijabla označava s $2d6$). Izračunajte $E(X)$, $Var(X)$ i standardnu devijaciju slučajne varijable X .

Rješenje.

Vrijednosti koje poprima slučajna varijabla X su redom $2, 3, \dots, 12$. Vrijednost 2 se može dobiti na jedan način, vrijednost $3 = 1 + 2 = 2 + 1$ na dva načina, $4 = 1 + 3 = 2 + 2 = 3 + 1$ na tri načina itd. Broj svih mogućnosti na koje mogu pasti dvije kockice je 36 . Pripadne vjerojatnosti se lako izračunaju primjenom klasične formule vjerojatnosti i sad imamo:

$$X \sim \left(\begin{array}{cccccccccc} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \end{array} \right).$$

$$E(X) = \sum_{k=2}^{12} kp_i = 2\frac{1}{36} + 3\frac{2}{36} + \dots + 7\frac{6}{36} + \dots + 12\frac{1}{36} = 7$$

$$E(X^2) = \sum_{k=2}^{12} k^2 p_i = 2^2 \frac{1}{36} + 3^2 \frac{2}{36} + \dots + 7^2 \frac{6}{36} + \dots + 12^2 \frac{1}{36} = \frac{329}{6} \approx 54.83$$

$$Var(X) = E(X^2) - E(X)^2 \approx 54.83 - 49 = 5.83, \quad \sigma(X) \approx 2.42$$

Zadatak 5.

Diskretna slučajna varijabla X ima uniformnu distribuciju $\left(\begin{array}{ccccccc} 1 & 2 & 3 & \dots & n-1 & n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} & \frac{1}{n} \end{array} \right)$. Znamo

da vrijedi $P(2 \leq X < 6) = \frac{1}{3}$. Izračunajte n .

Rješenje.

$$P(2 \leq X < 6) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) = \frac{1}{n} + \frac{1}{n} + \frac{1}{n} + \frac{1}{n} = \frac{4}{n} = \frac{1}{3} \Rightarrow n = 12.$$

Zadatak 6.

Zadana je diskretna slučajna varijabla X s distribucijom $\begin{pmatrix} 1 & 2 & 3 & \dots & n-1 & n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} & \frac{1}{n} \end{pmatrix}$. Izračunajte $E(X)$ i $Var(X)$. Pomoću dobivenog rezultata izračunajte $E(d6)$ i $Var(d6)$.

Rješenje.

Koristeći identitet $1 + 2 + \dots + n = \frac{n(n+1)}{2}$ dobivamo

$$E(X) = \bar{x} = \frac{1}{n}(1 + 2 + \dots + n) = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}.$$

Kako je $Var(X) = E(X^2) - E(X)^2$, moramo prvo izračunati $E(X^2)$. Uz pomoć identiteta $1^2 + 2^2 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}$ dobivamo

$$E(X^2) = \frac{1}{n}(1^2 + 2^2 + \dots + n^2) = \frac{1}{n} \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}$$

iz čega slijedi

$$Var(X) = E(X^2) - E(X)^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \dots = \frac{1}{12}(n^2 - 1).$$

$$\text{Dakle } E(d6) = \frac{7}{2} \text{ i } Var(d6) = \frac{35}{12}.$$

Zadatak 7.

Bacamo dvije simetrične kockice sa šest strana. Neka je X suma brojeva koje dobijemo bacanjem tih kockica ($X = 2d6$). Izračunajte

1. $E(X)$,
2. $Var(X)$,
3. za $\mu = E(X)$ i $\sigma = \sqrt{Var(X)}$ izračunajte $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$.

Rješenje.

$$X = 2d6 \sim \left(\begin{array}{cccccccccc} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{1}{18} & \frac{1}{12} & \frac{1}{9} & \frac{5}{36} & \frac{1}{6} & \frac{5}{36} & \frac{1}{9} & \frac{1}{12} & \frac{1}{18} & \frac{1}{36} \end{array} \right)$$

$$E(X) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{1}{18} + 4 \cdot \frac{1}{12} + 5 \cdot \frac{1}{9} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{1}{6} + 8 \cdot \frac{5}{36} + 9 \cdot \frac{1}{9} + 10 \cdot \frac{1}{12} + 11 \cdot \frac{1}{18} + 12 \cdot \frac{1}{36} = 7$$

$$X^2 \sim \left(\begin{array}{cccccccccc} 4 & 9 & 16 & 25 & 36 & 49 & 64 & 81 & 100 & 121 & 144 \\ \frac{1}{36} & \frac{1}{18} & \frac{1}{12} & \frac{1}{9} & \frac{5}{36} & \frac{1}{6} & \frac{5}{36} & \frac{1}{9} & \frac{1}{12} & \frac{1}{18} & \frac{1}{36} \end{array} \right), \quad E(X^2) = \frac{329}{6}$$

$$Var(X) = \overline{x^2} - \bar{x}^2 = \frac{329}{6} - 49 = \frac{35}{6}, \quad \sigma(X) = \sqrt{Var(X)} = \sqrt{\frac{35}{6}} \approx 2.4152$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx P(7 - 2 \cdot 2.4152 \leq X \leq 7 + 2 \cdot 2.4152) = P(4.5848 \leq X \leq 9.4152) =$$

$$= P(5) + P(6) + P(7) + P(8) + P(9) = \frac{1}{9} + \frac{5}{36} + \frac{1}{6} + \frac{5}{36} + \frac{1}{9} = \frac{2}{3}$$

20.6 Bernoullijev pokus. Bernoullijeva shema. Binomna distribucija

Zadatak 11.

Vjerojatnost da strijelac u n nezavisnih gađanja pogodi metu barem jednom je jednaka $r = 0.7626953125$.

Izračunajte n , ako je vjerojatnost da on pogodi metu u jednom gađanju jednaka $p = 0.25$.

Rješenje.

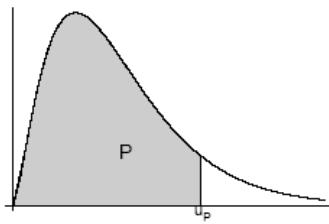
$$X \sim B(n, p)$$

$$r = P(X \geq 1) = 1 - P(X = 0)$$

$$1 - r = P(X = 0) = \binom{n}{0} p^0 (1-p)^n = (1-p)^n$$

$$\ln(1 - r) = \ln((1-p)^n) = n \ln(1-p)$$

$$n = \frac{\ln(1 - r)}{\ln(1 - p)} = \frac{\ln(0.2373046875)}{\ln(0.75)} = \frac{-1.43841036226}{-0.287682072452} = 5$$



Hi kvadrat razdioba $\chi^2(n)$

U tabeli se nalaze vrijednosti c za koje je $P\{\chi_m^2 \leq c\} = p$.

p	0.005	0.001	0.025	0.050	0.100	0.900	0.950	0.975	0.990	0.995
m										
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.002	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.024	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.091	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.210	0.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	0.676	0.381	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	0.599	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	0.857	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	1.152	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	1.479	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	1.834	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	2.214	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	2.617	5.009	5.892	7.041	19.812	22.362	24.736	27.688	29.819
14	4.075	3.041	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	3.483	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	3.942	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	4.416	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	4.905	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	5.407	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	5.921	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	6.447	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	6.983	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	7.529	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	8.085	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	10.520	8.649	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	9.222	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	9.803	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	10.391	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.994
29	13.121	10.986	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.335
30	13.787	11.588	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	17.917	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	24.674	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	31.738	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	39.036	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	46.520	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	54.156	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	61.918	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.170
110	75.550	69.790	82.867	86.792	91.471	129.385	135.480	140.916	147.414	151.948
120	83.852	77.756	91.573	95.705	100.624	140.233	146.567	152.211	158.950	163.648
150	109.142	102.113	117.985	122.692	128.275	172.581	179.581	185.800	193.207	198.360
200	152.241	143.842	162.728	168.279	174.835	226.021	233.994	241.058	249.445	255.264

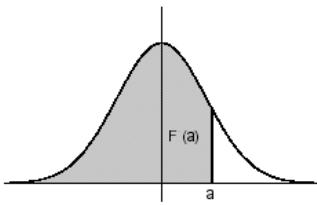


Tabela vrijednosti funkcije distribucije F standardne normalne razdiobe $N(0, 1)$
U tabeli se nalaze vrijednosti p za koje je $F(a) = P\{X \leq a\} = p$, $a \geq 0$.

a	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4.0	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998

Primjer 1.: Približna vrijednost broja $P\{X \leq 1.73\}$ dana u tablici se nalazi u presjeku retka na čijem je početku broj 1.7 i stupca na čijem je vrhu broj 0.03, tj. $P\{X \leq 1.73\} \approx 0.95818$.

Primjer 2.: Da nađemo približnu vrijednost broja $P\{X \leq -0.73\}$ koristimo se identitetima $P\{X \leq a\} + P\{X \geq a\} = 1$ i $P\{X \geq a\} = P\{X \leq -a\}$. Slijedi: $P\{X \leq -0.73\} = P\{X \geq 0.73\} = 1 - P\{X \leq 0.73\} \approx 1 - 0.76730 = 0.2327$.

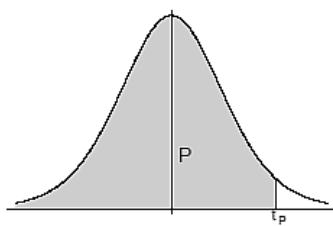


Tabela vrijednosti funkcije distribucije Studentove razdiobe $t(n)$

U tabeli su dane vrijednosti onih t_P u kojima funkcija distribucije Studentove t razdiobe poprima vrijednosti $P \in \{0.750, 0.800, 0.900, 0.950, 0.975, 0.990, 0.995, 0.9975, 0.999, 0.9995\}$ i to za stupnjeve slobode $n \in \{1, 2, \dots, 29, 30, 40, 50, \dots, 120, 130, \infty\}$.

P	0.750	0.800	0.900	0.950	0.975	0.990	0.995	0.9975	0.999	0.9995
n										
1	1.000	1.376	3.078	6.314	12.706	31.821	63.656	127.321	318.289	636.578
2	0.816	1.061	1.886	2.920	4.303	6.965	9.925	14.089	22.328	31.600
3	0.765	0.978	1.638	2.353	3.182	4.541	5.841	7.453	10.214	12.924
4	0.741	0.941	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.476	2.015	2.571	3.365	4.032	4.773	5.894	6.869
6	0.718	0.906	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.689
28	0.683	0.855	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.660
30	0.683	0.854	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
70	0.678	0.847	1.294	1.667	1.994	2.381	2.648	2.899	3.211	3.435
80	0.678	0.846	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
90	0.677	0.846	1.291	1.662	1.987	2.368	2.632	2.878	3.183	3.402
100	0.677	0.845	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
110	0.677	0.845	1.289	1.659	1.982	2.361	2.621	2.865	3.166	3.381
120	0.677	0.845	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
130	0.676	0.844	1.288	1.657	1.978	2.355	2.614	2.856	3.154	3.367
∞	0.674	0.842	1.282	1.645	1.960	2.326	2.576	2.808	3.090	3.297

Literatura

- [1] Neven Elezović, Teorija vjerojatnosti, zbirka zadataka, Element, Zagreb, 1995.
- [2] Nikola Sarapa, Vjerojatnost i Statistika, 1. dio, Kombinatorika, Školska knjiga, Zagreb, 1993.
- [3] Nikola Sarapa, Vjerojatnost i Statistika, 2. dio, Slučajne varijable, Osnove statistike, 1996. Školska knjiga, Zagreb, 1993.
- [4] Seymour Lipschutz, Jack Schiller, Schaum's Outline of Introduction to Probability and Statistics, McGraw-Hill, New York, 1998.
- [5] Nekoliko zadataka je preuzeto s bloga <http://maaminutemath.blogspot.com/>.