STATISTIKA

Teorija vjerovatnoće i statistika predstavljaju granu matematike od izuzetno velike praktične važnosti. U okviru ovog poglavlja biće uvedeni osnovni pojmovi kroz dva primjera. Jedan primjer je analiza rezultata koji su sutdenti ostvarili na nekom ispitu, dok je u drugom primjeru analizirano ispitivanje javnog mnjenja putem ankete.

Pretpostavimo da su nam dati rezultati koje su 100 studenata ostvarili na dva ispita. Svaki ispit je bodovan od 0 do 100 i broj poena je prikazan u tabelama 1 i 2.

100	81	77	72	68	65	63	60	54	48
95	81	77	71	68	65	62	59	53	48
90	81	77	70	67	65	62	58	53	44
88	81	76	69	66	64	62	58	52	43
87	81	76	69	66	64	62	58	52	43
86	79	76	69	66	64	62	57	51	40
86	79	75	68	66	63	62	56	50	39
85	79	74	68	65	63	60	56	50	36
83	77	73	68	65	63	60	55	50	35
82	77	73	68	65	63	60	54	49	30

Tabela 1

83	71	69	67	66	65	64	64	60	58
75	70	68	67	66	65	64	63	60	58
74	70	68	67	66	65	64	63	60	58
74	70	68	67	66	65	64	63	59	57
72	70	68	67	66	65	64	63	59	57
72	69	68	67	66	64	64	63	59	57
72	69	68	67	66	64	64	62	58	56
72	69	68	67	66	64	64	62	58	54
71	69	68	66	66	64	64	61	58	53
71	69	67	66	65	64	64	61	58	53

Tabela 2

Prosječna vrijednost, varijansa i standardna devijacija

Vidimo da se rezultati ova dva ispita dosta razlikuju, na prvom ispitu maksimalno je osvojeno 100 bodova a minimalno 30, dok je za slučaj drugog ispita (rezultati iz tabele 2) najveći broj poena 83 a najmanji 53. Minimalnu vrijednost skupa podataka X obilježavamo sa $\min(X)$ a maksimalnu sa $\max(X)$

Pored ovih informacija ima smisla govoriti o prosječnom broju poena. Ukoliko sa x_i označimo broj poena koje je osvojio student pod brojem i, pri čemu i uzima vrijednosti 1,2,...,100 tada **prosječni broj poena**, označimo ga sa $\overline{P}$, možemo tražiti kao:

$$\overline{P} = \frac{x_1 + x_2 + \ldots + x_{99} + x_{100}}{100}$$

Za prvi ispit dobijamo da je $\overline{P_1} = 65.31$ a za drugi ispit $\overline{P_2} = 64.99$

Gornje razmatranje možemo uopštiti tako da za proizvoljni skup od N brojeva x_i, $i = 1,2,\ldots,N$ definišemo srednju (prosječnu) vrijednost skupa kao:

$$\overline{P} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Da prosječna vrijednost ne karakteriše sve osobine skupa vidi se iz navedenih primjera. Naime oba ispita, iako sa bitno različitim rezultatima imaju bliske prosječne vrijednosti. Postavlja se pitanje kako mjeriti koliko podaci iz nekog skupa odstupaju od prosječne vrijednosti. Jedan od načina je da pronađemo prosječnu vrijednost kvadrata odstupanja, odnosno da sračunamo izraz:

$$V = \frac{(\overline{P} - x_1)^2 + (\overline{P} - x_2)^2 + \ldots + (\overline{P} - x_{100})^2}{100}$$

Veličina sračunata na ovaj način naziva se **varijansom skupa**. Za prvi ispit ona iznosi $V_1 = 180.63$ a za drugi $V_2 = 24.99$. Varijanse se bitno razlikuju što ukazuje na činenicu da kod prvog ispita imamo puno veće rasipanje broja bodova oko srednje vrijednosti nego kod drugog ispita. U opštem obliku za skup od N brojeva x_i, $i = 1,2,\ldots,N$ čija je prosječna vrijednost $\overline{P}$ varijansa je:

$$V = \frac{1}{N}\sum_{i=1}^{N} (\overline{P} - x_i)^2$$

Gornji izraz se može napisati i u sledećem obliku, koji je često prikladiniji za računanje:

$$V = \frac{1}{N}\sum_{i=1}^{N} (\overline{P}^2 - 2\overline{P}x_i + x_i^2) = \frac{1}{N}\sum_{i=1}^{N} \overline{P}^2 - 2\overline{P}\frac{1}{N}\sum_{i=1}^{N} x_i + \frac{1}{N}\sum_{i=1}^{N} x_i^2 = \overline{P}^2 - 2\overline{P}^2 + \frac{1}{N}\sum_{i=1}^{N} x_i^2$$

$$V = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \overline{P}^2$$

Dakle varijansu možemo lako sračunati ukoliko poznajemo prosječnu vrijednost skupa i ako pronađemo prosječnu vrijednost kvadrata elemenata skupa.

S obzirom da varijansa predstavlja srednju vrijednost kvadrata greške ima smisla definisati još jednu veličinu koja je jednaka kvadratnom korijenu varijanse i naziva se **standardna devijacija**:

$$\sigma = \sqrt{V}$$

Za slučaj prvog ispita ona iznosi $\sigma_1 = 13.44$ a u drugom slučaju imamo $\sigma_2 = 5$. Standardna devijacija dakle daje informaciju o tome za koliko se, u prosjeku, razlikuju elementi skupa od prosječne vrijednosti.

Medijan vrijednost skupa

Često se pri analizi podataka očekuje pojavljivanje "ekstremnih" vrijednosti, odnosno može se desiti da neke vrijednosti analiziranog skupa iz nekih razloga enormno odstupaju od ostalih. U takvim slučajevima srednja vrijednost često može biti nepodobna za opisivanje ponašanja elemenata skupa, pa se uvodi pojam medijane kao one vrijednosti u skupu podataka koja ima osobinu da dijeli skup na dva jednako brojna podskupa, prvi sa elementima koji su veći od nje, i drugi sa manjim elementima. Na primjer, voltmetar koji mjeri napon mreže ima osobinu da ponekad pokaže netačnu vrijednost, tako da su rezultati 7 mjerenja: 222V,15V, 251V, 210V, 215V, 231V, 1230V. Srednja vrijednost mjerenja je 340V, a medijana ovog skupa je 222V. Očigledno je u ovom slučaju medijana bliža stvarnoj vrijednosti nego srednja vrijednost, a razlog tome su dva, najvjerovatnije netačna, mjerenja koja su dala nerealne vrijednosti.

Medijana proizvoljnog skupa X obilježava se sa:

$$\mathrm{median}(X)$$

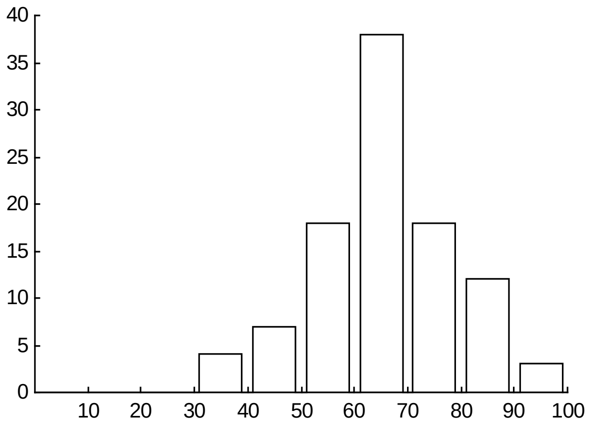
Iako je po definiciji dosta slična sa prosječnom vrijednošću računanje medijan vrijednosti skupa ja puno komplikovanije. U najvećem broju slučajeva podatke treba prvo sortirati i zatim uzeti podatak koji se u sortiranom nizu podataka nalazi u sredini.

Drugi način za izbjegavanje opisanih smetnji (one se nazivaju impulsnim smetnjama jer se javljaju rijetko, a kad se pojave njihov uticaj je veliki) je da se pri statističkoj analizi skupa podataka odbace po nekolike ekstremne vrijednosti Na primjer za skup od 100 mjerenja, eliminiše se na primjer 5 najmanjih vrijednosti i 5 najvećih vrijednosti i analizira se dobijeni skup. U navedenom primjeru, broj mjerenja je mali (7) pa ako odbacimo najmanju i najveću vrijednost tada prosječna vrijednost skupa postaje 225.8V što je prihvatljiv rezultat.

Histogram skupa podataka

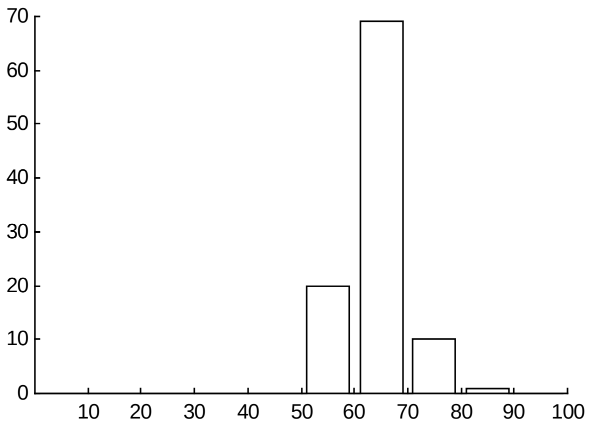
Iz prethodnog razmatranja vidjeli smo da za neki skup podataka možemo sračunati određene veličine koje na neki način opisuju ponašanje elemenata skupa. Problemu opisivanja ponašanja velikog broja elemenata možemo pristupiti i na drugi način. Podijelimo opseg mogućih vrijednosti elemenata (za posmatrani primjer to je opseg od 0 do 100) na neki broj podopsega, recimo 10, i odredimo koliko elemenata skupa leži unitar svakog od podopsega. Na primjer za slučaj prvog ispita u opsegu od 50 do 59 imamo 18 vrijednosti. Broj elemenata skupa koji leže u datom opsegu možemo prikazati tabelarno ili grafički u obliku **histograma**. Histogram crtamo tako što na x-osi nanesemo granice opsega i za svaki opseg crtamo pravougaonik čija je visina proporcionalna broju studenata kod kojih je broj bodova u odabranom opsegu.

Opseg	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-100
Br. stud.	0	0	0	4	7	18	38	18	12	3



Slika 1. Tabelarni prikaz raspodjele broja bodova po opsezima i histogram podataka iz tabele 1.

Opseg	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-100
Br. stud.	0	0	0	0	0	17	68	14	1	0



Slika 2. Tabelarni prikaz raspodjele broja bodova po opsezima i histogram podataka iz tabele 2.

Vjerovatnoća

Na osnovu podataka iz navednih tabela možemo izvući još neke informacije. Neka je svakom opsegu broja bodova pridružena odgovarajuća ocjena. Tada za veliki broj studenata možemo definisati **vjerovatnoću** da je proizvoljno odabrani student dobio neku ocjenu, na primjer 9. Tu vjerovatnoću mogli bismo definisati na osnovu razmotrenih podataka. Naime ocjena 9 se dobija ukoliko je broj bodova između 80 i 89. Od posmatranih 100 studenata za njih 12 je broj bodova u navedenom intervalu. Znači pitamo se kolika je vjerovatnoća da od 100 studenata, birajući jednog studenta odaberemo baš nekoga od tih 12. Ta vjerovatnoća, označimo je sa $p(8)$ iznosi:

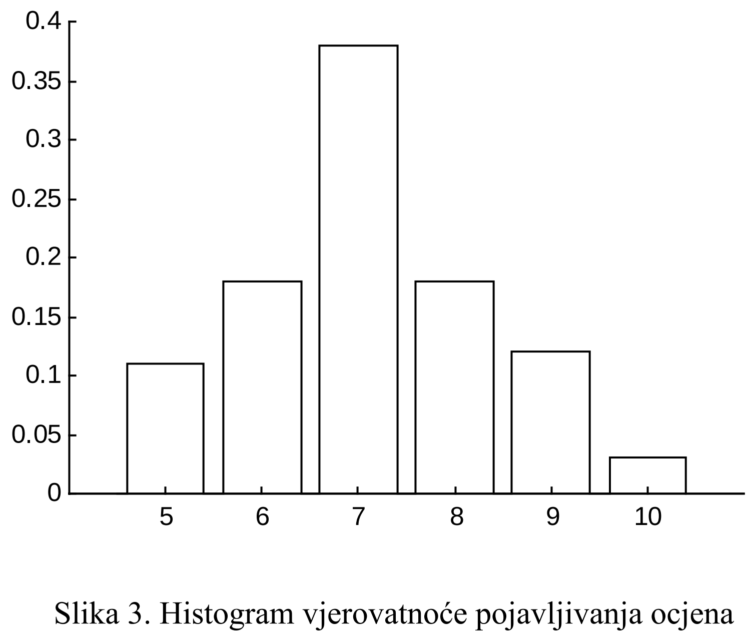
$$p(8) = \frac{12}{100} = 0.12$$

a možemo je izraziti i u procentima pomoživši je sa 100. Tada kažemo da je vjerovatnoća da je proizvoljno odabrani student dobio ocjenu 8 dvanaest posto. Na sličan način možemo odrediti i vjerovatnoće pojavljivanja svih ostalih ocjena, pri čemu za ocjemu 5 treba razmatrati sve one studente koji su osvojili manje od 50 bodova. Prikažimo te podatke tabelarno:

Ocjena	5	6	7	8	9	10
Vjerovatnoća	0.11	0.18	0.38	0.18	0.12	0.03

Tabela 3. Vjerovatnoće pojedinih ocjena na ispitu

Podatke iz gornje tabele takođe možemo prikazati grafički:



Slika 3. Histogram vjerovatnoće pojavljivanja ocjena

Pojam vjerovatnoće je opšti pojam koji se može definisati na više načina. Prvi način je definicija vjerovatnoće "**a priori**". Polazna tačka je prepoznavanje skupa elementarnih događaja i definisanja njihovih vjerovatnoća. Najčešće je to skup jednako vjerovatnih događaja. Primjer za ovako definisanu vjerovatnoću je bacanje kocke i posmatranje rezultata. Svaki se broj od 6 mogućih javlja sa istom vjerovatnoćom i suma svih tih 6 vjerovatnoća je jednaka 1 (sigurno će se pri bacanju kocke pojaviti neki od brojeva između 1 i 6), pa možemo "a priori" reči da je vjerovatnoća pojavljivanja broja 2 pri bacanju kocke $p(2)=1/6$.

Drugi način definicije vjerovatnoće je vjerovatnoća "**a posteriori**". Neka neki eksperiment može uzeti jedan ishod iz skupa ishoda. Ponovimo taj eksperiment n puta i ispitajmo koliko se puta pojavio određeni ishod A. Označimo taj borj sa n_A. Tada možemo vjerovatnoću pojavljivanja ishoda A definisati kao:

$$p_n(A)=\frac{n_A}{n}.$$

Ovo je samo procjena vjerovatnoće i smatra se da kada n teži beskonačnosti, tada $p_n(A)$ teži stvarnoj vjerovatnoći ishoda A koju označavamo sa $p(A)$.

Navedimo neke osobine vjerovatnoće obilježićemo sa A neki događaj a sa $p(A)$ njegovu vjerovatnoću:

- Vjerovatnoća događaja je nenegativna veličina, odnosno $p(A)\geq 0$ za svaki događaj A,
- Vjerovatnoća nekog događaja ne može biti veća od 1, odnosno $p(A)\leq 1$,
- Ukoliko su događaji A i B međusobno isključivi tada je $p(A+B)=p(A)+p(B)$, gdje pod zbirom događaja A i B smatramo događaj kada se desi događaj A ili događaj B.

Za događaj A kod koga je $p(A)=0$ kažemo da je nemoguć, a za događaj A kod koga je $p(A)=1$ kažemo da je siguran događaj. Primjer sigurnog događaja je događaj da je proizvoljno izabrani student dobio ocjenu između 5 i 10, dok je primjer nemogućeg događaja događaj da je student dobio ocjenu 12.

Diskretne slučajne promjenjive

Vratimo se ponovo na analizu rezultata ispita. Ukoliko slučajno biramo nekog studenta tada je i ocjena koju je on dobio na ispitu veličina koja zavisi od tog slučajnog izbora i nazivaćemo je slučajna promjenjiva. U posmatranom primjeru slučajna promjenjiva uzima vrijednosti od 5 do 10, odnosno konačno mnogo diskretnih vrijednosti, pa se zbog toga naziva diskretna slučajna promjenjiva. U tabeli 3 je definisana raspodjela vjerovatnoća posmatrane slučajne promjenjive.

Slučajna promjenjiva x je definisana ako je definisan skup njenih vrijednosti i pripadajuća raspodjela vjerovatnoća. Vjerovatnoća da slučajna promjenjiva uzme bilo koju vrijednost iz skupa vrijednosti je jednaka jedinici, odnosno to je siguran događaj. Raspodjela vjerovatnoća može biti zadata tako da za svaku moguću vrijednost slučajne promjenjive x, obilježićemo je sa x_i odredimo vjerovatnoću njenog pojavljivanja $p(x_i)$, kao što je na primjeru ispita urađeno u tabeli 3. Činjenicu da slučajna promjenjiva mora uzeti jednu od mogućih vrijednosti zapisujemo formulom:

$$\sum_i p(x_i) = 1$$

gdje se sumiranje obavlja po svim mogućim vrijednostima slučajne promjenjive.

Drugi način zadavanja raspodjele vjerovatnoća je da za svaku vrijednost slučajne promjenjive definišemo vjerovatnoću da slučajna promjenjiva ne premaši tu vrijednost. Tu vjerovatnoću označićemo sa $P(x \le x_i)$, a njene vrijednosti za posmatrani primjer analize ocjena na ispitu su date u tabeli 4. Ova vjerovatnoća se lako računa sabirajući vjerovatnoće pojavljivanja svih vrijednosti slučajne promjenjive koje su manje ili jednake vrijednosti x_i:

$$P(x \le x_i) = \sum_{x_n \le x_i} p(x_n)$$

Ocjena x_i	5	6	7	8	9	10
$P(x \le x_i)$	0.11	0.29	0.67	0.85	0.97	1

Tabela 4. Vjerovatnoće da ocjena na ispitu bude manja ili jednaka od zadate ocjene

Matematičko očekivanje, varijansa i standardna devijacija diskretne slučajne promjenjive

Za diskretnu slučajnu promjenjivu i njenu raspodjelu vjerovatnoća možemo definisati nekolika pojma koji su ekvivalentni statističkim pojmovima srednje vrijednosti, varijanse i standardne devijacije. Tako se za promjenjivu x i raspodjelu vjerovatnoća $p(x_i)=$ *vjerovatnoća da slučajna promjenjiva x uzme vrijednost* x_i definiše matematičko očekivanje slučajne promjenjive *x* kao:

$$E(x) = \sum_i x_i p(x_i)$$

varijansa:

$$V(x) = E((x - E(x))^2) = \sum_i (x_i - E(x))^2 p(x_i)$$

$$V(x) = E(x^2 - 2xE(x) + E(x)^2) = E(x^2) - 2E(x)E(x) + E(x)^2 = E(x^2) - E(x)^2$$

i standardna devijacija:

$$\sigma(x) = \sqrt{V(x)}$$

U navedenim formulama sabiranje se vrši po svim mogućim vrijednostima promjenjive *x*.

Za primjer koji razmatramo je: $E(x) = 7.11$, $V(x) = 1.54$ i $\sigma(x) = 1.24$

Kontinualne slučajne promjenjive

Ukoliko slučajna promjenljiva uzima vrijednosti iz konačnog skupa nazivamo je diskretnom slučajnom promjenjivom. Ako je skup vrijednosti slučajne promjenjive kontinualan, nazivamo je kontinualnom slučajnom promjenjivom. Primjer kontinualne slučajne promjenjive je napon elektroenergetske mreže. Naime, napon varira oko svoje nominalne vrijednosti i može uzeti bilo koju vrijednost, s tim što je u najvećem broju slučajeva on blizak vrijednosti 220V. Za kontinualnu slučajnu promjenjivu njenu raspodjelu vjerovatnoća ne možemo prikazati na identičan način kao u opisanom slučaju diskretne slučajne promjenjive jer je vjerovatnoća da napon uzme neku fiksnu vrijednost, recimo tačno 220V veoma mala, teoretski jednaka nuli, pa se stoga za kontinualnu slučajnu promjenjivu x raspodjela vjerovatnoća definiše kao funkcija koja za proizvoljno odabranu vrojednost x_0 daje vjerovatnoću da slučajna promjenjiva ne premaši vrijednost x_0. Tu vjerovatnoću označavamo sa $\mathbf{P}(x < x_0)$. Na ovaj način je definisana funkcija $P_x(x_0)$ i ona se naziva funkcijom raspodjele vjerovatnoća slučajne promjenjive x. Izvod ove funkcije po promjenjivoj x_0, u oznaci $p(x_0)$ naziva se funkcijom gustine vjerovatnoće.

Navešćemo nekolike osobine funkcije gustine raspodjele vjerovatnoća $p(x)$ i funkcije raspodjele vjerovatnoća $P(x)$ kontinualne slučajne promjenjive x:

1. $P(x)$ je neopadajuća funkcija, $p(x)$ je nenegativna funkcija

2. $\lim\limits_{x\to-\infty} P(x) = 0$, $\lim\limits_{x\to+\infty} P(x) = 1$

3. $\int\limits_{-\infty}^{\infty} p(x)dx = 1$

4. $\mathbf{P}(a < x < b) = P(b) - P(a) = \int\limits_{a}^{b} p(x)dx$, gdje je sa $\mathbf{P}(a < x < b)$ označena vjerovatnoća da slučajna promjenjiva uzme neku vrijednost između vrijednosti a i b.

Matematičko očekivanje i varijansa kontinualne slučajne promjenjive x se računaju kao:

$$E(x) = \int\limits_{-\infty}^{\infty} xp(x)dx\,,\ V(x) = \int\limits_{-\infty}^{\infty} (E(x) - x)^2 p(x)dx = \int\limits_{-\infty}^{\infty} x^2 p(x)dx - E(x)^2$$

Gausova raspojela

Različite slučajne promjenjive imaju i različite raspodjele vjerovatnoća. Primijećeno je da je u velikom broju slučajnih procesa prisutna takozvana Gausova raspodjela vjerovatnoća kod koje je funkcija gustine raspodjele vjerovatnoća:

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

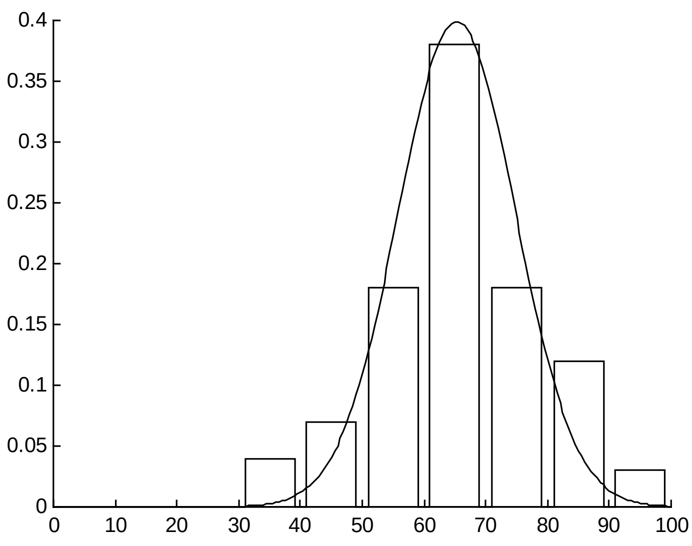
gdje je μ srednja vrijednost, odnosno matematičko očekivanje slučajne promjenjive, a σ njena varijansa. Odnosno ako je x slučajna promjenjiva koja se podvrgava Gausovoj raspodjeli, tada je:

$$E(x) = \mu \text{ i } V(x) = \sigma^2$$

Interesantno će biti prikazati histogram sa slike 1 zajedno sa pripadajućom Gausovom funkcijom gistine raspodjele vjerovatnoće slučajne promjenjive čija je srednja vrijednost jednaka srednjem broju bodova a varijansa jednaka varijansi broja bodova. Ranije smo sračunali da su ovi podaci za razmatrani slučaj 65.31 i 13.44, odnosno pripadna funkcija raspodjele kontinualne slučajne promjenjive je:

$$p(x) = \frac{1}{13.44\sqrt{2\pi}} e^{-\left(\frac{x-65.31}{13.44}\right)^2}$$

Grafik ove funkcije, zajedno sa histogramom vjerovatnoće da se broj bodova nađe u nekom od 10 opsega između 0 i 100 su prikazani na slici 4. Vidmo dosta sličnosti, odnosno možemo zaključiti da je broj bodova koji je neki student dobio na ispitu slučajna promjenjiva čija je raspodjela vjerovatnoća približno Gausova. Gausova raspodjela je od izuzetnog značaja jer se može dokazati da ako je neka slučajna promjenjiva x suma većeg broja promjenjivih čije su raspodjele proizvoljne, tada sa povećanjem broja sabiraka raspodjela promjenjive x teži Gausovoj. Tako se dosta često slučajna promjenjiva sa Gausovom raspodjelom na računaru dobija preko sume od 10-12 slučajnih promjenjivih čija se raspodjela može lakše realizovati na računaru. Najčešće je to uniformna raspodjela koja će biti opisana u nastavku.

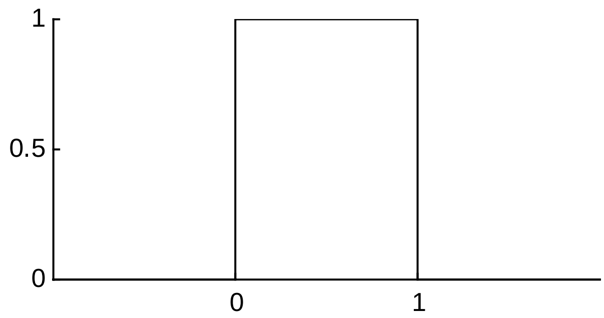


Slika 4. Histogram vjerovatnoća broja bodova i kriva gustine vjerovatnoće odgovarajuće kontinualne slučajne promjenjive

Uniformna raspodjela

Pored Gausove raspodjele, poznate su i druge raspodjele vjerovatnoća. Ranije je pomenuta uniformna raspodjela, kod koje se definišu donja i gornja vrijednost a i b slučajne promjenjive i smatra se da ona sa jednakim vjerovatnoćama uzima bilo koju vrijednost između vrijednosti a i b. Na slici 5 je prikazana uniformna raspodjela slučajne promjenjive koja uzima vrijednosti od 0 do 1. Kada slučajna promjenjiva x ima uniformnu raspodjelu tada je:

$$E(x) = (a+b)/2 \text{ i } V(x) = (b-a)/12$$



Slika 5. Uniformna raspodjela vjerovatnoća

Binomna raspodjela

Raspodjela koja se često koristi je i binomna raspodjela. U njenoj definiciji se polazi od događaja koji mogu imati dva moguća ishoda A i B. Vjerovatnoću ishoda A obilježimo sa p a vjerovatnoću ishoda B sa q. Na primjer može se posmatrati bacanje novčića pri čemu je ishod A “glava” a ishod B “pismo”. Očigledno je $p+q=1$, jer se jedan od dva ishoda mora desiti. Posmatrajmo sada ponavljanje ovakvog događaja, i sa n označimo broj ponavljanja. Od tih n događaja ishod A se pojavio recimo x puta, a ishod B n-x puta. Ukoliko posmatramo x kao slučajnu promjenjivu tada vjerovatnoća da promjenjiva x uzme vrijednost k:

$$p(k) = \binom{n}{k} p^k q^{n-k} \text{ gdje je } \binom{n}{k} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k(k-1)(k-1)\ldots\cdot 2\cdot 1}$$

matematičko očekivanje ove slučajne promjenjive je $E(x) = np$ a standardna devijacija $\sigma(x) = \sqrt{npq} = \sqrt{np(1-p)}$.

Posmatrajmo sada srednji broj pojavljivanja ishoda A, i označimo ga sa y. On se u konkretnom slučaju dobija kada broj k podijelimo sa n. Na ovaj način definisana je slučajna promjenjiva $y = x/n$ kod koje je $E(y) = p$, i $\sigma(y) = \sqrt{p(1-p)/n}$.

Granice pouzdanosti statističkog eksperimenta

Postavimo problem ovako: Sprovodi se anketa među građanima jedne zemlje. Anketno pitanje je tipa da/ne. Ispitivano je 100 ljudi i odgovor "da" se pojavio 46 puta. Nameće se zaključak da možemo procijeniti da je vjerovatnoća pojavljivanja odgovora "da" 0.46. Ispitajmo koliko je ovaj rezultat pouzdan. Rezultat se poklapa sa slučajnom promjenjivom *y*, a za nju imamo da je njena standardna devijacija:

$$\sigma(y) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.46 \cdot 0.54}{100}} = 0.05$$

Veličina standardne devijacije daje informaciju o tome koliko posmatrana slučajna promjenjiva odstupa od svoje očekivane vrijednosti. Na osnovu ove veličine možemo odrediti neki interval $(E(y)-a, E(y)+a)$, i tražiti vjerovatnoću da se stvarna vrijednost (kada bi anketu proširili na čitavu populaciju) nalazi u ovom opsegu. Drugim riječima možemo odrediti pouzdanost rezultata ankete.

Pretpostavimo da anketu izvodimo na sledeći način: odredimo 10 grupa od po 100 ljudi, njih anketiramo, pronađemo ukupnu varijansu i srednju vrijednost tako dobijenih promjenjivih *y*. Formirajmo sada novu promjenjivu *z* koja je jednaka prosječnoj vrijednosti svih promjenjivih *y*. Promjenjiva *z* ima raspodjelu koja je probližno Gausova, što je ranije napomenuto (ona predstavlja sumu 10 slučajnih promjenjivih). Za Gausovu raspodjelu se intervali pouzdanosti mogu tačno odrediti, i dati su u tabeli 5, gdje je $\mu = E(z)$ i $\sigma = \sigma(z)$.

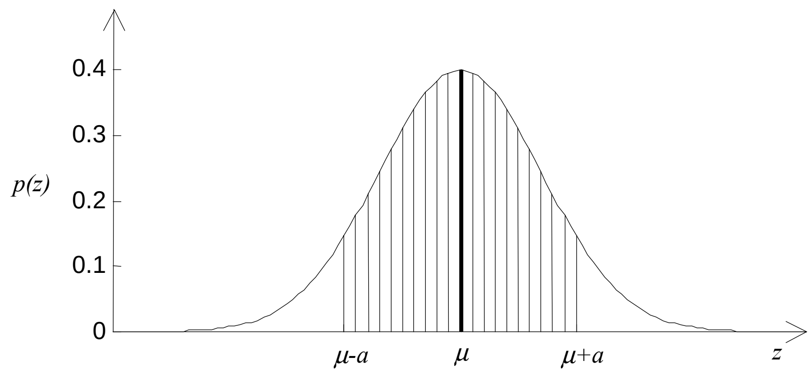
a	$\mathbf{P}(\mu - a < z < \mu + a)$
σ	68%
2σ	95%
3σ	99.7%

Tabela 5. Intervali pouzdanosti za Gausovu raspodjelu

Ovjde je:

$$P(\mu - a < z < \mu + a) = \int_{\mu-a}^{\mu+a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{z-\mu}{\sigma}\right)^2} dz$$

što grafički možemo predstaviti površinom ispod krive gustine raspodjele vjerovatnoća:



S obzirom da je $\sigma(z) = \sqrt{p(1-p)/n} = 0.016$, pri čemu je $n = 1000$ možemo zaključiti da će rezultat aknketiranja čitave populacije biti u granicama od 0.428 do 0.492 sa vjerovatnoćom od 95% (uzeli smo interval $\pm 2\sigma$). Ukoliko zadržimo istu vjerovatnoću pouzdanosti rezultata, za početni primjer sa 100 anketiranih osoba dobijaju se granice pouzdanosti od 0.36 do 0.46.

Ovdje treba napomenuti da bi rezultati bili vjerodostojni moraju slučajne promjenjive y biti međusobno nezavisne. To se postiže ispravnim odabirom populacije za sprovođenje ankete, odnosno vodi se računa o mjestu boravka ispitanika, o godinama, o školskoj spremi, polu, ...

Problem možemo formulisati i na drugi način. Tražimo koliko ispitanika treba uključiti u anketu tako da je rezultat ankete sa vjerovatnoćom 99.7% u intervalu ± 0.01 oko neke očekivane vrijednosti. Ovdje moramo pretpostaviti očekivani odgovor na anketi i neka on iznosi 0.50. Tada imamo da je: $3\sigma = 0.01$, odnosno $\sigma = 0.00333$. Kako je $\sigma = \sqrt{p(1-p)/n}$ to je $n = p(1-p)/\sigma^2$, odnosno $n = 22500$. Znači trebali bi anketirati 22500 ljudi, da bi se ostvarile zadate granice pouzdanosti ankete.