III. Osnovni pojmovi teorije verovatnoće

Neophodni matematički pojmovi za razumevanje teksta u ovoj glavi:[1]

Osnovni pojmovi teorije skupova *(skup, prazan skup, jednaki skupovi, podskup, pravi podskup, univerzalni skup, unija skupova, presek skupova, disjunktni skupovi, razlika skupova, komplement skupa, osnovna pravila algebre skupova)*

Funkcija (preslikavanje)

Operator sabiranja (sumacioni operator) Σ

Operator proizvoda Π

Slučajnost je svuda oko nas. Toga smo često bolno svesni. Ponekad se radujemo zbog toga što su "mnoge stvari u životu nepredvidljive". Jer, kakav bi to život bio kada bi sve bilo predvidljivo…Mnoga pojedinačna dešavanja realnog sveta naše svakodnevice deluju nam kao potpuno nepredvidljiva premda tokom dužeg posmatranja ovih dešavanja uočavamo izvesne pravilnosti. Volimo ponekad (a poneki među nama i često) da se igramo slučajnošću kao u igrama na sreću iako ne možemo unapred da predvidimo ishod svakog pokušaja u ovim igrama. Gledajući svakodnevno vesti o broju novorođene dece u nekoj dovoljno velikoj populaciji uočavamo da su procenti novorođenih dečaka i devojčica retko kad drastično različiti. (Kada bismo se zagledali u procente umrlih u određenim vremenskim periodima u nekoj većoj populaciji ljudi, bili bismo zapanjeni mističnim pravilnostima). Gledajući odrasle ljude oko sebe, mada ne možemo da predvidimo visinu sledeće odrasle osobe koju ćemo sresti, uočavamo da je među onima koje srećemo najviše onih koji su osrednje visoki a znatno manje onih izrazito niskih i izrazito visokih. Dakle, iako u pojedinačnim dešavanjima nepredvidljive, *slučajne* pojave odlikuju izvesne pravilnosti.

U svakodnevnom životu veoma često se srećemo i sa subjektivnim *verovatnoćama* iskazujući stepen svoje uverenosti u dešavanje nekog događaja za koji nije izvesno hoće li se desiti ("Vrlo je verovatno ću se uskoro zaljubiti"). Često se pitamo kolike su *šanse* da se nešto desi ("Kakve su šanse da položim ovaj ispit"). Mnoga pitanja koja istraživači u psihologiji i drugim naučnim oblastima mogu postaviti direktno su povezana sa pitanjem verovatnoće da se nešto desi ("Kolika je verovatnoća da bilo koja osoba koja doživi vrlo stresno iskustvo, tj. iskustvo koje prevazilazi uobičajena ljudska iskustva razvije tzv. posttraumatski stresni poremećaj"). Pri statističkoj analizi podataka dobijenih istraživanjem veoma često (barem implicitno) postavljaju se pitanja u vezi sa verovatnoćom da se nešto desi ("Kolika je verovatnoća da se na slučajnom uzorku u pogledu nekog svojstva dobije razlika određene veličine između određenih grupa ljudi ako u populaciji među tim grupama zapravo nema razlike"). Teorija verovatnoće se na matematički rigorozan način bavi upravo tzv. slučajnim fenomenima pružajući sredstva za dolaženje do

[1] Čitalac koji ne vlada nephodnim pojmovima može konsultovati odrednice **Osnovni pojmovi teorije skupova**, **Funkcija**, **Operator sabiranja (sumacioni operator)** i **Operator proizvoda** u Matematičkom pojmovniku u Dodatku **

odgovora u situacijama neizvesnosti, tj. u situacijama kada ne raspolažemo kompletnim informacijama.

Teorija verovatnoće daje matematičke modele slučajnih pojava. Pri tome, slučaj se u teoriji verovatnoće ne shvata kao haos bez ikakve pravilnosti i bez ikakvih uzročno-posledičnih veza: *slučajne pojave su one pojave u kojima je pojedinačni ishod neizvestan ali postoje određene pravilnosti u raspodeli ishoda u velikom broju ponavljanja.*[2]

Kao što smo već istakli u uvodnoj glavi, statistički postupci, posebno oni koji spadaju u tzv. inferencijalnu statistiku, zasnovani su na teoriji verovatnoće, matematičkoj disciplini koja daje matematičke modele tzv. slučajnih fenomena. Stoga je za razumevanje statističkih postupaka i njihovu primenu neophodno poznavati barem osnovne pojmove teorije verovatnoće. U ovoj glavi izložićemo samo osnovne pojmove ove teorije. Određene pojmove, teoreme i teorijske modele iz teorije verovatnoće koji ne budu izloženi u ovoj glavi izložićemo u okviru teksta za čije razumevanje nam oni budu neophodni. Pri tome, nastojaćemo da izlaganje odabranih delova teorije verovatnoće izvedemo na način za koji verujemo da će biti razumljiv čitaocima kojima je knjiga namenjena trudeći se da, koliko god je to moguće, damo matematički korektan prikaz. Preporučujemo čitaocu da se tokom korišćenja ostalih delova knjige vrati na pojmove koji su obrađeni u ovoj glavi kada god za tim oseti potrebu.

Osnovni pojam u teoriji verovatnoće jeste naravno sam pojam verovatnoće. Videćemo uskoro da definisanje ovog ključnog pojma nije baš jednostavan poduhvat, te da postoje različita određenja i tumačenja samog pojma verovatnoće. Za definisanje pojma verovatnoće neophodno je prethodno definisati nekoliko polaznih pojmova.

III. 1. Polazni pojmovi za definisanje pojma verovatnoće

Slučajni eksperiment (engl. random experiment)

Slučajni, stohastički ili statistički eksperiment je svaki preciziran (planiran) proces posmatranja ili prikupljanja podataka za koji važe sledeći uslovi:

- može se **ponavljati neograničen**, tj. proizvoljan **broj puta** u **istim uslovima**;
- može imati samo **jedan ishod** iz **skupa svih mogućih i međusobno isključivih ishoda** pri čemu su **svi mogući ishodi unapred poznati**;

[2] Teorija verovatnoće je jedna od matematičkih oblasti koje su se relativno kasno razvile. Zasnivanju teorije verovatnoće značajan podsticaj dala je jedna od štetnih ljudskih strasti, kockarska strast. Premda su se i pre toga javljali nagoveštaji probabilističkih pojmova u matematici, prvi jasniji začeci teorije verovatnoće pojavljuju se u 16. veku u knjizi Đirolama Kardana *Liber de ludo aleae (Knjiga o igrama na sreću)*. (Kardano je inače bio lekar, matematičar, fizičar, astronom… i prema određenim tvrdnjama toliko je verovao u astrologiju da je predvidevši tačan dan svoje smrti na taj dan izvršio samoubistvo!). Dakle, u 16. veku formulisano je pravilo prema kojem je verovatnoća da će pravilna kocka pasti na jednu od svojih šest strana jednaka 1/6. U to vreme, kao i vek kasnije, kockanje je bilo vrlo popularno, posebno u aristokratskim krugovima. Godine 1654., poznati francuski kockar sa naučničkom i matematičkom nastrojenošću duha, Antoan Gombo (poznatiji kao Ševalje de Mere/Chevalier de Méré/) obraća se svom prijatelju, poznatom matematičaru i filozofu Blezu Paskalu sa molbom da mu pomogne u rešavanju određenih problema u vezi sa računanjem očekivane frekvencije dobitaka i gubitaka u igrama na sreću, kao i u vezi sa pravednom podelom úloga kada se igra prekine. U prepisci koju je u vezi sa rešavanjem ovih problema Paskal vodio sa svojim savremenikom Fermaom, takođe poznatim matematičarem, postavljene su osnove matematičke teorije verovatnoće (Podroban i zanimljiv prikaz zasnivanja i daljeg istorijskog razvoja teorije verovatnoće može se pročitati u Cowles, 2001).

- pojedinačna pojava određenog ishoda **ne može se predvideti** sa potpunom izvesnošću.

Pri izlaganju teorije verovatnoće u matematičkim knjigama uobičajeno se kao primeri slučajnih eksperimenata navode: bacanje novčića, bacanje kocke, izvlačenje kuglica iz kutije i slično. U psihologiji i srodnim oblastima kao primere onoga što se u teoriji verovatnoće zove slučajnim eksperimentom možemo navesti mnoge postupke sistematskih posmatranja i merenja tokom istraživačkog procesa: anketiranje ispitanika koji pripada slučajnom uzorku u pogledu polne pripadnosti, merenje vremena reakcije na dati stimulus za ispitanika u eksperimentu koji uključuje randomizaciju, zadavanje određenog upitnika ličnosti ili određenog testa znanja ispitaniku iz slučajnog uzorka, računanje prosečne inteligencije jednog slučajnog uzorka iz neke populacije i slično. Naravno, istraživači u ovim oblastima to uobičajeno ne zovu slučajnim eksperimentima već sistematskim posmatranjem, merenjem, testiranjem. Ono što je bitno jeste to da se sa stanovišta teorije verovatnoće ove aktivnosti mogu posmatrati kao slučajni eksperimenti, pod uslovom da se posmatranje, merenje ili testiranje izvodi na tzv. slučajnom uzorku ili po određenim metodološkim pravilima.

Pokušaj (engl. trial)

Pokušaj predstavlja svako pojedinačno izvođenje slučajnog eksperimenta. Na primer, jedno bacanje novčića, merenje vremena reakcije na jedan stimulus kod jednog ispitanika iz slučajnog uzorka...

Ishod (engl. outcome)

Ishod je određeni, realizovani ili mogući rezultat jednog pokušaja. Na primer, mogući ishodi u jednom bacanju novčića su P(ismo) i G(lava) dok su mogući ishodi u dva bacanja novčića PP, GG, GP i PG. Ishodi moraju biti međusobno isključivi, tj. u jednom pokušaju mora biti moguć samo jedan ishod. Na primer, u anketiranju ispitanika u pogledu polne pripadnosti ne sme biti moguće da ispitanik zaokruži oba međusobno isključiva odgovora („muško“, „žensko“), a u psihološkom testiranju jednog ispitanika iz slučajnog uzorka u pogledu depresivnosti (ili merenju znanja ispitanika određenim testom) ispitanik može imati samo jedan složaj odgovora, odnosno jedan rezultat ili skor kojim se iskazuje njegov stepen depresivnosti (ili znanje).

Skup mogućih ishoda (engl. sample space)

Prostor ili skup svih mogućih ishoda nekog slučajnog eksperimenta jeste skup svih mogućih distinktivnih ishoda tog eksperimenta. U nastavku teksta skup mogućih ishoda označavaćemo slovom S (od engleskog Sample space). Na primer, skup ishoda u tri bacanja novčića sadrži sledeće elemente: PPP, PPG, PGP, PGG, GPP, GPG, GGP, GGG. To možemo napisati i ovako: S = {PPP, PPG, PGP, PGG, GPP, GPG, GGP, GGG}. Često za jedan isti slučajni eksperiment možemo definisati različite skupove mogućih ishoda, zavisno od cilja posmatranja. Na primer, ako bacimo dve pravilne kocke i posmatramo parove brojeva sa gornje strane bačenih kocki skup mogućih ishoda možemo opisati sa 36 parova brojeva od 1 do 6, tj. S = {(1,1), (1,2), ... (6,5), (6,6)}. Međutim, ako nas zanima samo zbir brojeva sa gornje strane dveju bačenih kocki a ne pojedinačni brojevi na svakoj od kocki, skup mogućih ishoda bi činilo 11 brojeva od 2

do 12, tj. $S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11,12\}$ (cf. Bartoszynski & Niewiadomska-Bugaj, 2008).

(Slučajan) događaj (engl. random event)

Slučajan događaj je bilo koji skup ishoda, tj. bilo koji podskup skupa svih mogućih ishoda slučajnog eksperimenta. U nastavku teksta u ovoj glavi slučajni događaj zvaćemo prosto događajem. Događaj bismo mogli formalno definisati na sledeći način: $A \subset S$, pri čemu je A oznaka bilo kojeg događaja, a S predstavlja skup svih mogućih ishoda nekog slučajnog eksperimenta. U primeru sa bacanjem tri novčića, događaj *tri „pisma" ili tri „glave"* bio bi sledeći skup: $B = \{PPP, GGG\}$. Događaj može biti prost, tj. može sadržati samo jedan ishod eksperimenta ili složen, tj. može obuhvatati više od jednog ishoda eksperimenta. Na primer, realizacija jedne od šest strana kocke predstavljala bi prost ili elementaran događaj dok bi pojava neparnog broja pri bacanju kocke predstavljala složeni događaj. Prema tome, događaj $A = \{3\}$, koji označava da je kocka pala na trojku bio bi elementaran događaj, a događaj $B = \{1, 3, 5\}$ pri bacanju kocke bio bi složen događaj. Siguran događaj predstavlja skup svih mogućih ishoda nekog slučajnog eksperimenta, a nemoguć događaj predstavlja prazan skup. Siguran događaj ćemo označavati slovom S, a nemoguć događaj oznakom praznog skupa, tj. oznakom $\varnothing$.
Unija dva događaja je događaj koji obuhvata ishode koji su obuhvaćeni jednim ili drugim događajem. Presek dva događaja obuhvata ishode koji su zajednički za oba događaja. Komplement nekog događaja obuhvata sve ishode u skupu svih mogućih ishoda S koji nisu obuhvaćeni tim događajem.
Budući da su događaji zapravo skupovi na njih se može u potpunosti primenjivati algebra skupova. (Čitalac kojem je to potrebno može konsultovati odrednicu **Osnovni pojmovi teorije skupova** u Matematičkom pojmovniku u Dodatku **).

III. 2. Verovatnoća

Verovatnoća (engl. probability) je, uopšteno rečeno, mera mogućnosti ili izvesnosti dešavanja nekog događaja kao rezultata, tj. ishoda slučajnog eksperimenta. Do ove mere dolazi se apriornom logičkom analizom procesa koji vodi određenom ishodu, pretpostavkama ili empirijskom procenom. Iskazuje se razlomcima (1/10), proporcijama (0.28) ili procentima (33%). Postoje različita tumačenja ili određenja pojma verovatnoće: aksiomatsko, statističko (objektivno), klasično i subjektivno.

Aksiomatsko određenje verovatnoće dao je 1933. godine ruski matematičar A. N. Kolmogorov. Prema ovom određenju funkcija verovatnoće svakom događaju A iz skupa svih mogućih ishoda S pridružuje realni broj $P(A)$, verovatnoću događaja A, tako da su ispunjena, tj. tako da važe tri aksioma:[3]

- **Nenegativnost**: $P(A) \geq 0$ za svako A;
- **Normiranost**: $P(S) = 1$ i
- **Prebrojiva aditivnost**: ako prebrojivi podskupovi događaja A_1, A_2,..., A_k postoje i ako su uzajamno isključivi (nemaju zajedničkih elemenata, tj.

[3] Budući da su događaji zapravo skupovi trebalo bi u oznaci verovatnoće događaja A koristiti oznaku $P\{A\}$. Kako se u većini knjiga iz teorije verovatnoće u ovom slučaju koristi obična zagrada odlučili smo da i mi tako postupimo.

svaki sadrži različite ishode od svakog drugog) verovatnoća njihove unije jednaka je zbiru njihovih pojedinačnih verovatnoća:[4]

$$P(A_1 \cup A_2 \cup ... \cup A_k) = P(A_1) + P(A_2) + ... + P(A_k)$$

Dakle, realni brojevi koji se pridružuju svakom događaju, tj. podskupu skupa svih mogućih ishoda, predstavljaju verovatnoće ako ispunjavaju sledeće uslove:

1. nenegativni su;
2. jednaki su jedinici za ceo skup, tj. prostor ishoda, i
3. zbir brojeva dodeljenih uzajamno isključivim događajima jednak je broju dodeljenom njihovoj uniji.

Aksiomatsko određenje verovatnoće važno je za razumevanje matematičkih svojstava verovatnoće i koristi se u dokazima određenih teorema u okviru teorije verovatnoće. Za konkretno računanje verovatnoće u statistici od značaja su klasično (apriorno) i statističko određenje verovatnoće. Ono što je sa praktičnog stanovišta najvažnije uočiti u aksiomatskom određenju verovatnoće jeste da ***verovatnoća može biti jednaka bilo kojem realnom broju od 0 do 1, tj. da ne može biti manja od 0 niti veća od 1.***

Statistički, verovatnoća događaja A, u oznaci $P(A)$, definiše se kao granična vrednost relativne učestalosti (relativne frekvencije) događaja A, tj. kao granična vrednost proporcije ishoda koji pripadaju događaju A u ukupnom broju pokušaja:

$$\boxed{P(A) = \lim_{n \to \infty} \frac{f_A}{n}}$$

Oznakom f_A u gornjoj formuli označena je učestalost događaja A, oznakom n ukupan broj pokušaja, a lim je oznaka za graničnu vrednost ili limes. (Za ovu potrebu dovoljno je limes shvatiti kao broj kojem se približava količnik f_A i n kako se n povećava, tj. približava beskonačnosti). Drugim rečima, količnik f_A i n daje približnu vrednost verovatnoće događaja A, pri čemu je ova približna vrednost utoliko bliža pravoj verovatnoći događaja A ukoliko je n veće.

Verovatnoća događaja A je, dakle, broj kojem se približava relativna frekvencija tog događaja sa približavanjem broja pokušaja ka beskonačnosti. U praksi, budući da se eksperiment može ponoviti samo konačan broj puta, verovatnoću kao graničnu vrednost nemoguće je odrediti (čak i ako ona postoji) pa se verovatnoća ocenjuje, ako je broj pokušaja veliki, jednostavno preko relativne učestalosti, tj. relativne frekvencije događaja:

$$\boxed{P(A) = \frac{f_A}{n}}$$

To bi praktično značilo da, ako, na primer, želimo da statistički odredimo verovatnoću rađanja muškog deteta trebalo bi da beležimo u velikom broju rađanja pol novorođenih beba i da na osnovu velikog broja takvih podataka podelimo broj novorođenih muškog pola sa ukupnim brojem novorođenih. Na ovaj način ocenjeno je da je verovatnoća rađanja muškog deteta 0.515! Dakle, verovatnoća rađanja muškog

[4] Aksiomi koji su prikazani ovde podrazumevaju da je skup svih mogućih ishoda diskretan, tj. prebrojiv. To je učinjeno kako bi i čitalac sa manjim matematičkim predznanjem mogao lakše da razume aksiome. Kada je skup svih mogućih ishoda S neprebrojiv, tj. kontinuiran tada se zbir zamenjuje integralom. Isto tako, iako smo aksiom aditivnosti ovde prikazali za k događaja on važi i za beskonačan broj događaja.

deteta veća je od verovatnoće rađanja ženskog deteta. Ova potonja verovatnoća se ocenjuje da iznosi 0.485. /Nešto veća verovatnoća rađanja muškog deteta mogla bi izgledati i kao velika božanska mudrost (ili mudrost prirode) u obezbeđivanju podjednakog broja muškaraca i žena u ranom odraslom dobu budući da veći broj muške dece – u poređenju sa ženskom decom – umire neposredno po rođenju, u detinjstvu, pa i u adolescentnom periodu. Veću proporciju rađanja muške dece kao argument u prilog Božjeg Proviđenja koristio je u svom veoma zanimljivom članku jedan engleski kraljevski lekar još početkom 18. veka (Arbuthnott, 1710-1712)/.[5]
Ključna ideja koju treba imati na umu pri primeni statistike u odnosu na određivanje verovatnoće na osnovu relativne frekvencije jeste sledeća: za dobro određivanje verovatnoće na statistički način nužno je da raspolažemo *velikim brojem* posmatranja, tj. velikim n.

Zapravo, statističko određenje verovatnoće kao relativne frekvencije utemeljeno je na tzv. *Zakonu velikih brojeva*. Zakon velikih brojeva ima različite formulacije, a formulacija koja je u ovom kontekstu relevantna glasi:[6]

$$\lim_{n\to\infty} P(|p - \pi| > \varepsilon) = 0$$

Pri tome, P(•) je opšta oznaka funkcije verovatnoće, lim je oznaka limesa, n je broj nezavisnih pokušaja ili posmatranja, ε je proizvoljno mali broj veći od nule, p je relativna frekvencija određenog događaja, a π verovatnoća tog događaja. Dakle, prema Zakonu velikih brojeva verovatnoća da će se relativna frekvencija događaja razlikovati od verovatnoće tog događaja za više od proizvoljno malog pozitivnog broja ε teži nuli kako broj nezavisnih pokušaja ili posmatranja (n) teži beskonačnosti. Uočimo da prema ovom obliku Zakona nije granična vrednost same razlike relativne frekvencije i verovatnoće događaja jednaka nuli već granična vrednost verovatnoće da ta razlika bude veća od proizvoljno malog pozitivnog broja ε. Prema tome, ovaj zakon ne tvrdi da se za veliko n relativna frekvencija i verovatnoća garantovano neće nimalo razlikovati, već da je vrlo malo verovatno da će se one razlikovati. Uočimo, isto tako da se Zakon velikih brojeva odnosi na relativnu a ne na apsolutnu frekvenciju. Jedna od čestih zabluda je upravo ona po kojoj frekvencija dešavanja nekog događaja u velikom broju posmatranja mora biti blizu vrednosti n*π, tj. frekvenciji dobijenoj na osnovu π (cf. Falk & Lann, 2013). Postoji i jača forma ovog zakona ali ona zahteva i strože uslove pod kojima važi. Prema *Strogom zakonu velikih brojeva*

$$P(\lim_{n\to\infty} p = \pi) = 1$$

Dakle, kako n teži beskonačnosti, relativna frekvencija se približava („konvergira“) verovatnoći sa verovatnoćom jednakom 1, tj. gotovo sigurno.

Klasično (apriorno) određenje verovatnoće formulisao je francuski matematičar Pjer Simon Laplas u knjizi objavljenoj1812. godine (dakle, baš u vreme kada je Napoleon

[5] Arbuthnott u svom članku prilaže i tabelu broja novorođenih (tačnije krštenih) dečaka i devojčica u Londonu od 1629. do 1740. godine iz koje se vidi da je svake od ovih godina rođeno nešto više muških nego ženskih beba. Zaista, i noviji podaci iz velikog broja istraživanja ukazuju na relativno veći broj začete ili rođene muške dece, dok se u ranom odraslom dobu broj žena i muškaraca ujednačuje (cf. Ellis et al., 2008).

[6] Ovaj tzv. Slabi oblik zakona velikih brojeva prvi je formulisao poznati matematičar Jakob Bernuli u delu objavljenom posle njegove smrti početkom 18. veka.

pokušavao da osvoji Rusiju...) i zasniva se na pretpostavci da su svi ishodi eksperimenta međusobno isključivi i jednako verovatni. Ako neki događaj A obuhvata n_A ishoda od ukupnog broja mogućih ishoda (n), tada je verovatnoća događaja A jednaka odnosu broja ishoda obuhvaćenih događajem A ("povoljnih" ishoda) i ukupnog broja mogućih ishoda:

$$\boxed{P(A) = \frac{n_A}{n}}$$

Verovatnoća, u skladu sa ovom koncepcijom, određuje se matematičkim putem, unapred, pre izvršenog eksperimenta. Za određivanje ukupnog broja mogućih ishoda, posebno kada je taj broj veliki, koriste se pravila koja su razvijena u okviru matematičke discipline kombinatorike (O nekim od ovih pravila čitalac se može, ako se za tim ukaže potreba, informisati u Matematičkom pojmovniku pod odrednicom **Osnovni pojmovi i pravila kombinatorike**). Budući da se u ovom određenju verovatnoće unapred pretpostavlja jednaka verovatnoća svih mogućih ishoda nekog slučajnog eksperimenta, ovo određenje „pati“ od izvesne cirkularnosti.

Primeri određivanja verovatnoće na osnovu klasičnog određenja:

a) Kolika je verovatnoća da će u jednom bacanju novčića pasti 'glava' (misli se, naravno, na stranu novčića na kojoj se nalazi prikaz glave nekog od velikana)?
Broj mogućih ishoda je u ovom slučaju 2 (P, G), a broj „povoljnih“ ishoda jednak je 1. Prema tome:

P(G) = 1/2 = 0.5

b) Kolika je verovatnoća da se iz dobro promešanog špila od 52 karte izvuče karta "pik"?
Budući da su u špilu 52 karte a da karata „pik“ ima 13:

P(pik) = 13/52 = 0.25

c) Kolika je verovatnoća da se u dva nezavisna bacanja pravilnog novčića ne dobije nijedanput „pismo“?
Mogućih ishoda u ovom slučaju je 4 jer je S = {PP, PG, GP, GG}, a samo jedan ishod je „povoljan“ (GG). Prema tome, P(nula „pisama“ u dva bacanja novčića) = 1/4 = 0.25.

d) Kolika je verovatnoća da se u dva nezavisna bacanja pravilne kocke ne dobiju „dve šestice“?
U prvom bacanju postoji 6 mogućih ishoda (kocka može pasti na jednu od 6 strana sa brojevima tačkica od 1 do 6). U drugom bacanju ima takođe 6 mogućih ishoda. Prema osnovnom pravilu kombinatorike broj mogućih ishoda u dva bacanja kocke je 6*6, tj. 36. Budući da postoji samo jedan „nepovoljan“ ishod („dve šestice“, tj. 6 u prvom i 6 u drugom bacanju) preostaje 35 „povoljnih“ ishoda. Stoga je P(ne „dve šestice“ u dva bacanja kocke) = 35/36 = 0.972.

e) Kolika je verovatnoća da se slučajnim razmeštanjem slova A, B, O, R i T, pri čemu sva slova moraju biti iskorišćena u svakom razmeštanju a jedno slovo se može pojaviti samo jedanput u istom razmeštanju, dobije reč TORBA?
Ukupan broj mogućih razmeštaja pri navedenim uslovima dobija se kao broj permutacija bez ponavljanja sa po 5 elemenata: 5*4*3*2*1 = 120. Samo jedan od tih razmeštaja (T, O, R, B, A) daje željenu reč. Prema tome, P(dobijanje reči TORBA slučajnim razmeštanjem slova A, B, O, R i T) = 1/120 = 0.008.

Subjektivno tumačenje verovatnoće određuje ovaj pojam u smislu stepena uverenosti ili verovanja osobe u pogledu dešavanja nekog događaja. Pokazano je da se ovaj stepen uverenosti neke osobe u izvesnost dešavanja nekog događaja može meriti (Bartoszynski & Niewiadomska-Bugaj, 2008). Veoma plastičnu formulaciju subjektivnog tumačenja verovatnoće dali su Edvards, Lindman i Sevidž 1963. godine u članku objavljenom u jednom od najpoznatijih psiholoških časopisa Psychological Review (Psihološka revija): verovatnoća događaja *A* za neku osobu je "cena koju" bi osoba bila "voljna da plati u zamenu za jedan dolar" koji će dobiti ako se događaj *A* desi. Na primer, događaj *Sutra će pasti kiša* ima verovatnoću 1/3 ako bi osoba bila spremna da sada dâ trećinu dolara u zamenu za 1 dolar koji bi dobila ako sutra padne kiša (Edvards, Lindman, & Savage, 1963, str. 197). Proučavanje subjektivne verovatnoće predstavlja zasebnu oblast blisko vezanu za matematičku teoriju izbora i donošenja odluka. Subjektivno tumačenje verovatnoće u osnovi je jednog posebnog pristupa statistici, tzv. bajesovskog ili bejzijanskog pristupa.

Uslovna verovatnoća

Uslovna verovatnoća je mera mogućnosti ili izvesnosti dešavanja događaja B pošto se prethodno desio događaj *A*. Dešavanje događaja *A* je uslov koji treba da bude ispunjen i pod kojim ispitujemo dešavanje događaja *B*. (Budući da je teorija verovatnoće matematička disciplina i da se ne mora nužno primenjivati na dešavanja u realnom svetu ovo „prethodno dešavanje" pri razmatranju uslovne verovatnoće treba shvatiti u najširem smislu jer se ono može odvijati samo u našim mislima, tj. u misaonom eksperimentu).
Uslovna verovatnoća događaja *B*, pod uslovom da se desio događaj *A*, izračunava se po formuli:

$$\mathrm{P}(B|A) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(A)}$$

gde je P($A \cap B$) zajednička verovatnoća događaja *A* i *B*, a P(*A*) verovatnoća događaja *A*.
Na primer, ako je u nekoj populaciji 490 000 muškaraca i 510 000 žena i ako bismo svima pri popisu stanovništva postavili pitanje *Da li se plašite pauka*, pri čemu se na pitanje može odgovoriti samo sa *Da* ili *Ne*, dobijene podatke mogli bismo prikazati u

Tabeli ** (Naravno, u ovom izmišljenom primeru misli se na životinju a ne na „pauka“ koji nosi nepropisno parkirana vozila):[7]

Tabela **: Prikaz moguće raspodele odgovora muškaraca i žena iz jedne zamišljene populacije na pitanje *Da li se plašite pauka* (brojevi u tabeli su u hiljadama ispitanika)

	Da (*Plašim se pauka*)	Ne (*Ne plašim se pauka)*	Ukupno
Muškarci	80	410	490
Žene	350	160	510
	430	570	1 000

Pretpostavimo da je događaj *B* događaj „plašiti se pauka“ (odgovor *Da*), a događaj *A* događaj „biti muškog pola“ i da želimo da odredimo verovatnoću da se slučajno izabrana osoba iz te populacije plaši pauka ako znamo da je reč o muškoj osobi. Pogledajmo, pre svega verovatnoću događaja „plašiti se pauka“ (odgovor *Da*). Verovatnoću da se slučajno odabrana osoba iz date populacije plaši pauka odredili bismo kao količnik broja onih koji su odgovorili *Da* na postavljeno pitanje i ukupnog broja ispitanih: 430 000/1 000 000 = 0.43. Sada se možemo zapitati da li znanje da je reč o muškoj osobi menja ovu verovatnoću? Prema onom što znamo o odnosu muškaraca i žena prema paucima (posebno s obzirom na to da se ne radi o „pauku“ koji nosi automobile...) očekujemo da će informacija da je reč o muškoj osobi unekoliko smanjiti ovu verovatnoću.

Verovatnoću događaja „biti muškog pola“, tj. verovatnoću da će slučajno odabrana osoba iz ove populacije biti muško odredili bismo kao proporciju muškaraca i ona bi bila jednaka 490 000/1 000 000, tj. 0.49, a verovatnoću zajedničkog dešavanja događaja „biti muškog pola“ i događaja „plašiti se pauka“ kao količnik 80 000/1 000 000, što iznosi 0.08. Tada bismo uslovnu verovatnoću događaja „plašiti se pauka pod uslovom da je reč o osobi muškog pola iz date populacije“ izračunali na sledeći način:

$$\frac{\frac{80\,000}{1\,000\,000}}{\frac{490\,000}{1\,000\,000}} = \frac{80\,000}{490\,000} = 0.16$$

Dakle, verovatnoća da se na slučaj izvučena osoba iz date populacije plaši pauka, ako znamo da je reč o osobi muškog pola, iznosi 0.16. Uočimo da uslovna verovatnoća u ovom slučaju zapravo predstavlja proporciju, tj. relativnu učestalost događaja (kao i obična verovatnoća) ali ne u celokupnoj populaciji (muškaraca i žena) već u subpopulaciji koja je ograničena u skladu sa događajem koji predstavlja uslov, tj. u subpopulaciji muškaraca. To je isto kao da smo umesto na celokupnu populaciju pogled ograničili samo na subpopulaciju muškaraca, tj. samo na onaj red ukupne tabele koji se odnosi na muškarce:

	Da (*Plašim se pauka*)	Ne (*Ne plašim se pauka)*	Ukupno
Muškarci	80	410	490

[7] Nadam se da je čitaocu jasno da se ovakva pitanja inače ne postavljaju pri popisu stanovništva. Radi se, dakle, o jednom zamišljenom a ne realnom ispitivanju.

Uočimo, međutim, da se verovatnoća zajedničkog dešavanja dva događaja („biti muško“ i „plašiti se pauka“) računa kao proporcija u odnosu na celokupnu populaciju. Budući da se verovatnoće događaja poput događaja u ovom primeru „biti muško“, „biti žensko“, „plašiti se pauka“ i „ne plašiti se pauka“ određuju na osnovu frekvencija koje se nalaze na marginama tabele, verovatnoće takvih događaja nazivamo uobičajeno marginalnim verovatnoćama.

U vezi sa uslovnom verovatnoćom, veoma je važno imati na umu da, u opštem slučaju, $P(B|A)$ nije jednako $P(A|B)$. Odnos između ove dve uslovne verovatnoće prikazaćemo u nastavku teksta pri prikazivanju Bajesove teoreme.

Statistički nezavisni događaji

Uslovna verovatnoća služi, između ostalog, za definisanje statističke nezavisnosti događaja, pojma koji je veoma važan za razumevanje statističke analize podataka.

Zapamtite: Ako je $P(B|A) = P(B)$ onda su A i B statistički nezavisni događaji. Ako je $P(B|A) = P(B)$ tada je i $P(A|B) = P(A)$.

Dakle, ako dešavanje događaja koji predstavlja uslov ne menja verovatnoću dešavanja drugog događaja, tada za ova dva događaja kažemo da su statistički nezavisni događaji. U našem prethodnom primeru očigledno je da događaji „plašiti se pauka“ i „biti muškog pola“ nisu statistički nezavisni događaji jer je verovatnoća događaja „plašiti se pauka“ jednaka 0.43, a verovatnoća događaja „plašiti se pauka pod uslovom da je reč o muškarcima iz te populacije“ je 0.16.
Uslovna verovatnoća je veoma važna u primeni statistike, pogotovu za razumevanje statističkih postupaka za ispitivanje veza između kategoričkih varijabli (o tome će biti reči u Glavi **).

Osnovne teoreme o verovatnoći

Teorema o komplementu

Komplement događaja A, u oznaci A^c, jeste događaj koji obuhvata sve ishode u skupu mogućih ishoda S koji nisu obuhvaćeni događajem A. Uočimo, dakle, da se komplement uvek definiše u odnosu na određeni skup mogućih ishoda S. Na primer, ako pri jednom bacanju novčića događaj A predstavlja događaj *pala je “glava”* (G), onda bi komplement tog događaja bio događaj *palo je „pismo“* (P). Naime, skup mogućih ishoda u tom slučaju je $S = \{G, P\}$ pa pošto je ishod G obuhvaćen događajem A, događaj A^c obuhvata preostali mogući ishod P.

Ako je verovatnoća događaja A jednaka $P(A)$, tada je verovatnoća komplementa događaja A (u odnosu na S), u oznaci $P(A^c)$:

$$P(A^c) = 1 - P(A).$$

Ova teorema naprosto sledi iz aksioma normiranosti i prebrojive aditivnosti u aksiomatskom određenju verovatnoće koje smo prikazali u prethodnom delu teksta.[8] Naime, prema definiciji komplementa događaja A, očigledno je da su A i A^c uzajamno isključivi događaji i da njihova unija predstavlja zapravo skup svih ishoda S u odnosu na koji se određuje komplement. Prema tome, $P(A) + P(A^c) = P(S) = 1$.
Na primer, ako je verovatnoća da padne "glava" u jednom bacanju novčića jednaka 0.5, tada je verovatnoća da padne "pismo" jednaka 0.5. Isto tako, ako je verovatnoća oboljevanja od depresije u nekoj populaciji jednaka 0.10 onda je verovatnoća neoboljevanja od depresije u toj populaciji jednaka 1 – 0.10, tj. 0.90.

Aditivna teorema

Aditivna teorema tiče se verovatnoće unije dva događaja, tj. verovatnoće da se od dva događaja desi jedan ili drugi, ili oba.
Verovatnoća da se desi događaj A ili događaj B jednaka je zbiru verovatnoća individualnih događaja umanjenom za verovatnoću zajedničkog dešavanja oba događaja:

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

Uočimo da pri razmatranju verovatnoće unije dva događaja logičko "ili" uključuje i mogućnost da se dese oba događaja istovremeno što je nešto drugačije od svakodnevne upotrebe ovog termina gde se uobičajeno podrazumeva da se dešava samo jedan od dva događaja ali ne i oba istovremeno. Budući da se pri računanju zbira verovatnoća $P(A)$ i $P(B)$ verovatnoća istovremenog dešavanja oba događaja uključuje dva puta /i u $P(A)$ i u $P(B)$/, potrebno je od tog zbira oduzeti $P(A \cap B)$.
Ako su događaji **međusobno isključivi**, tj. ako se ne mogu desiti oba istovremeno (tačnije ako je presek dva događaja nemoguć događaj) teorema dobija sledeći oblik:

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

Naime, u tom slučaju je $P(A \cap B)$ jednako nuli, budući da je verovatnoća nemogućeg događaja jednaka nuli.
Uopštenje ovog oblika aditivne teoreme na veći broj međusobno isključivih događaja dat je pri aksiomatskom definisanju verovatnoće u obliku aksioma prebrojive aditivnosti:

$$P(A_1 \cup A_2 \cup \ldots \cup A_k) = \sum_{i=1}^{k} A_i$$

Dakle, ***verovatnoća unije uzajamno isključivih događaja jednaka je zbiru pojedinačnih verovatnoća ovih događaja***.

Multiplikativna teorema

Prema ovoj teoremi verovatnoća zajedničkog dešavanja dva događaja (A i B) jednaka je proizvodu verovatnoće događaja A i uslovne verovatnoće događaja B (ili proizvodu verovatnoće događaja B i uslovne verovatnoće događaja A):

[8] Formalni dokaz ove teoreme, kao i aditivne i multiplikativne teoreme zainteresovani čitalac može pogledati u Hogg & Craig, 1978, str. 13–14.

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

Ako su događaji A i B statistički nezavisni teorema dobija sledeći oblik:

$$P(A \cap B) = P(A)P(B)$$

jer je u tom slučaju uslovna verovatnoća $P(B|A)$ jednaka $P(B)$ a uslovna verovatnoća $P(A|B)$ jednaka $P(A)$. Dakle, verovatnoća zajedničkog dešavanja dva nezavisna događaja jednaka je proizvodu verovatnoća svakog od dva događaja.

Multiplikativna teorema se može generalizovati na veći broj događaja:
$P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1)P(A_2|A_1)(A_3|A_1 \cap A_2) \ldots P(A_k|A_1 \cap A_2 \cap \ldots \cap A_{k-1})$

Teorema o zajedničkoj verovatnoći dva statistički nezavisna događaja može se generalizovati na k ovakvih događaja. Ako su $A_1, A_2,...,A_k$ uzajamno statistički nezavisni događaji tada je njihova zajednička verovatnoća (verovatnoća njihovog zajedničkog dešavanja), u oznaci $P(A_1 \cap A_2 \cap \ldots \cap A_k)$, jednaka proizvodu pojedinačnih verovatnoća tih događaja:

$$P(A_1 \cap A_2 \cap \ldots \cap A_k) = \prod_{i=1}^{k} A_i$$

U statistici se ova generalizacija koristi pri konstrukciji tzv. funkcije verodostojnosti (engl. likelihood function).

Uzajamnu statističku nezavisnost k događaja (pri čemu je $k > 2$) treba razlikovati od statističke nezavisnosti ovih k događaja po parovima. Naime, kada imamo više od dva događaja oni mogu biti nezavisni po parovima ali ne moraju biti uzajamno statistički nezavisni. Ilustrovaćemo to na vrlo jednostavnom primeru (prema Mukhopadhyay, 2000, str. 11). Ako razmatramo bacanje pravilnog novčića dva puta tada skup svih mogućih ishoda predstavlja skup S, $S = \{GG, GP, PG, PP\}$. Definišimo sledeća tri događaja:
A_1 – pala je „glava“ u prvom bacanju, tj. {GG, GP};
A_2 – pala je „glava“ u drugom bacanju, tj. {GG, PG};
A_3 – isti ishod je u oba bacanja, tj. {GG, PP}.
Budući da svaki od ova tri događaja obuhvata po 2 od ukupno 4 moguća ishoda, verovatnoća svakog od njih je 2/4, tj. 1/2. Događaji A_1 i A_2 su statistčki nezavisni jer je $P(A_1 \cap A_2) = P(GG) = 1/4 = P(A_1) * P(A_2)$. Takođe, događaji A_1 i A_3 su statistički nezavisni jer je $P(A_1 \cap A_3) = P(GG) = 1/4 = P(A_1) * P(A_3)$. Na kraju, događaji A_2 i A_3 su statistički nezavisni jer je $P(A_2 \cap A_3) = P(GG) = 1/4 = P(A_2) * P(A_3)$. Ali, $P(A_1 \cap A_2 \cap A_3) = P(GG) = 1/4$ što nije jednako $P(A_1) * P(A_2) * P(A_3)$, budući da je potonja verovatnoća jednaka 1/8. Dakle, iako su statistički nezavisni po parovima, događaji A_1, A_2 i A_3 nisu uzajamno statistički nezavisni.

Treba uočiti i razliku između uzajamno isključivih događaja i statistički nezavisnih događaja: uzajamno isključivi događaji nisu statistički nezavisni jer dešavanje jednog od uzajamno isključivih događaja menja verovatnoću drugog događaja, tj. svodi verovatnoću drugog događaja na nulu. Naime, tada je uslovna verovatnoća ovog drugog događaja ako znamo da se desio prvi događaj jednaka 0.

Teorema o komplementu, aditivna i multiplikativna teorema veoma su korisne kao osnova za računanje verovatnoća u mnogim primenama teorije verovatnoće, pa i u statistici.

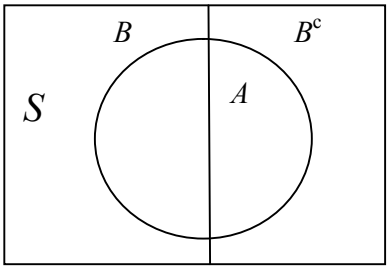
Bajes-Laplasova teorema

Bajes-Laplasova teorema zapravo je samo na drugačiji način iskazana uslovna verovatnoća događaja B pod uslovom dešavanja događaja A. Na osnovu multiplikativne teoreme uslovnu verovatnoću događaja *B* pod uslovom dešavanja događaja *A* možemo iskazati na nešto drugačiji način od onoga koji smo dali u formuli $P(B|A) = P(A \cap B) / P(A)$. Budući da je po multiplikativnom pravilu $P(A \cap B) = P(B)P(A|B)$, zamenom izraza $P(A \cap B)$ u formuli za uslovnu verovatnoću sa $P(B)P(A|B)$ dobijamo:

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

Na taj način dobijamo tzv. Bajes-Laplasovu teoremu ili pravilo koje je navodno formulisao engleski sveštenik Tomas Bajes (Thomas Bayes) u 18. veku.[9] Na osnovu ovog pravila može se jasno uočiti da, u opštem slučaju, $P(B|A) \neq P(A|B)$. Naravno, u nekim posebnim situacijama ove dve verovatnoće mogu biti jednake. Na primer, jednakost ovih verovatnoća dobija se onda kada su verovatnoće događaja *A* i *B* jednake, tj. kada je $P(A) = P(B)$. Bajesovo pravilo je veoma važno u primeni statistike, a predstavljalo je osnovu za razvoj jednog pristupa u statistici, tzv. Bajesovskog pristupa.

Verovatnoću događaja *A* u imeniocu Bajesovog pravila možemo iskazati na drugačiji način. Polazimo od toga da je unija svakog događaja i njegovoog komplementa, prema definiciji komplementa, jednaka događaju koji obuhvata sve moguće ishode (*S*), tj. izvesnom događaju. Dakle, $B \cup B^c = S$. Iskažimo sada događaj *A* preko preseka tog događaja sa skupom svih mogućih ishoda: $A = A \cap S$. Odatle, zamenom *S* sa $B \cup B^c$, dobija se da je $A = A \cap (B \cup B^c)$, što je isto kao $(A \cap B) \cup (A \cap B^c)$.[10]
Ovo se vrlo lako može razumeti ako napravimo vizuelni prikaz nalik tzv. Venovom dijagramu.



[9] Budući da je izvorni engleski izgovor ovog prezimena „Bejz“, u statističkoj literaturi na našem jeziku može se naići i na prideve „Bejzov“, „bejzijanski“. Mi smo u ovom tekstu odlučili da zadržimo oblike „Bajes“, „Bajesov“ i „bajesovski“ jer se u matematičkim knjigama na našem jeziku uobičajeno sreću ovi oblici. Na kraju krajeva, pravilna transkripcija ovog imena možda i nije tako bitna budući da je jedan od najpoznatijih istoričara statistike Stefen Stigler doveo u ozbiljnu sumnju Bajesovo autorstvo nad pomenutom teoremom, pripisujući moguće autorstvo genijalnom slepom matematičaru iz tog vremena N. Saundersonu (cf. Stigler, 1983).

[10] Sve jednakosti u ovom delu slede na osnovu osnovnih pravila algebre skupova koja su data u okviru odrednice **Osnovni pojmovi teorije skupova** u Matematičkom pojmovniku u Dodatku **.

Iz dijagrama vidimo da je celokupni četvorougao izvestan događaj S, dok su događaji B (četvorougao levo od vertikalne linije) i B^c (četvorougao desno od vertikalne linije) uzajamno isključivi (četvorouglovi se nimalo ne preklapaju) i čine zajedno S. Događaj A predstavlja krug čija površina je očigledno sastavljena iz dva nepreklapajuća dela: deo kruga sa leve strane vertikalne linije je presek događaja A i događaja B, a površina kruga **sa** desne strane je presek događaja A i događaja B^c.

Prema tome, $P(A) = P((A \cap B) \cup (A \cap B^c))$. Budući da su događaji $(A \cap B)$ i $(A \cap B^c)$ uzajamno isključivi jer su B i B^c uzajamno isključivi po definiciji komplementa, na osnovu oblika aditivne teoreme za uzajamno isključive događaje sledi da je $P(A) = P(A \cap B) + P(A \cap B^c)$.

Na osnovu multiplikativne teoreme sledi da je $P(A \cap B) = P(B)\, P(A|B)$, a $P(A \cap B^c) = P(B^c)\, P(A|\, B^c)$. Zamenom $P(A \cap B)$ sa $P(B)\, P(A|B)$ a $P(A \cap B^c)$ sa $P(B^c)\, P(A|\, B^c)$ u izrazu $P(A) = P(A \cap B) + P(A \cap B^c)$ dobijamo $P(A) = P(B)\, P(A|B) + P(B^c)\, P(A|\, B^c)$. Na kraju zamenom $P(A)$ u imeniocu Bajesovog pravila sa $P(B)\, P(A|B) + P(B^c)\, P(A|\, B^c)$ dobijamo drugačiji oblik ovog pravila koji se mnogo češće koristi u praktičnim primenama:

$$P(B|A) = \frac{P(B)P(A|B)}{P(B)P(A|B) + P(B^c)P(A|B^c)}$$

Bajesovo pravilo može, na primer, poslužiti za određivanje verovatnoće da osoba ima određenu bolest (B) ako ima određeni simptom (A) u situaciji kada su poznate apriorna verovatnoća oboljevanja od te bolesti, $P(B)$, i uslovna verovatnoća da osoba ima određeni simptom ako ima neku bolest B, tj. $P(A|B)$.

Na primer, ako je prevalencija oboljevanja od određene bolesti (događaj B) u nekoj populaciji 0.2, tada je $P(B) = 0.2$ a $P(B^c) = 0.8$. Ako 90% osoba za koje znamo da su obolele od te bolesti ima određeni simptom i ako se taj simptom javlja kod 5% osoba iako znamo da nisu obolele od te bolesti, tada je $P(A|B) = 0.9$ a $P(A|\, B^c) = 0.05$. U tom slučaju, verovatnoća da osoba koja ima dati simptom jeste obolela od te bolesti može se izračunati na sledeći način:

$$\frac{0.2 * 0.9}{0.2 * 0.9 + 0.8 * 0.05} = 0.82$$

Uslovnu verovatnoću da osoba ima određenu bolest ako ima određeni simptom možemo odrediti jednostavnim deljenjem broja obolelih sa određenim simptomom i ukupnog broja osoba sa datim simptomom. Na primer, ako je u populaciji za koju primenjujemo dati primer 10 000 osoba, tada je:

- broj obolelih 0.2*10 000, tj. 2000;
- broj onih koji nisu oboleli 0.8 * 10 000, tj. 8 000;
- broj obolelih osoba koje imaju određeni simptom jednak je 0.9*2000, tj. 1800;
- broj osoba koje nisu obolele a imaju dati simptom jednak je 0.05*8000, tj. 400.
- ukupan broj osoba sa tim simptomom jednak je 1800 + 400, tj. 2200.

Prema tome, verovatnoća da je osoba obolela od date bolesti ako ima određeni simptom jednaka je 1800/2200, tj. 0.82.

Na taj način, alternativni izraz Bajesove teoreme, iskazan kao količnik frekvencija, imao bi sledeći oblik:

$$\mathrm{P}(B|A) = \frac{\mathrm{f}\,(B \cap A)}{\mathrm{f}(A)}$$

Dakle, verovatnoća događaja *B* pod uslovom da se desio događaj *A*, jednaka je količniku frekvencije zajedničkog dešavanja ova dva događaja i frekvencije događaja *A*, tj. događaja koji predstavlja uslov.

Bajesovo pravilo možemo uopštiti tako što događaj *A* u imeniocu početnog oblika Bajesovog pravila iskažemo preko preseka ovog događaja sa unijom konačnog broja događaja, pri čemu je unija ovih događaja jednaka događaju koji obuhvata sve ishode iz skupa svih mogućih ishoda *S*. Na taj način dobijamo najopštiji oblik Bajesovog pravila:

Ako su A i $B_1, B_2, \ldots, B_n$ događaji u skupu svih mogućih ishoda S, pri čemu je unija svih događaja B_i jednaka S, svi događaji B_i imaju nenultu verovatnoću, i događaji B_i su međusobno isključivi, dakle:

a) $\bigcup_{i=1}^{n} B_i = S$;

b) $(\forall i)$, $\mathrm{P}(B_i) > 0$;

c) $B_i \cap B_j = \varnothing$, za $i \neq j$,

tada je

$$\mathrm{P}(B_j|A) = \frac{\mathrm{P}(B_j)\mathrm{P}(A|B_j)}{\sum_{i=1}^{n} \mathrm{P}(B_i)\mathrm{P}(A|B_i)}.$$

Izraz u imeniocu ovog pravila predstavlja tzv. *formulu potpune verovatnoće*. Formulom potpune verovatnoće možemo verovatnoću događaja *A* iskazati preko zbira verovatnoća preseka događaja *A* i svakog od n uzajamno isključivih događaja B_i, pri čemu su događaji B_i odabrani tako da je njihova unija jednaka skupu svih mogućih ishoda.

Šanse i verovatnoća

U vezi sa izvesnošću, odnosno verovatnoćom dešavanja nekog događaja često se (i u svakodnevnom životu a i u naučnim publikacijama) govori o šansama ili izgledima za dešavanje nekog događaja.

Šanse (engl. Odds) ili izgledi događaja *A*, u oznaci O(*A*), predstavljaju količnik verovatnoće da se događaj *A* desi i verovatnoće da se taj događaj ne desi:[11]

[11] Ponekad se termin šansa događaja (pogotovu termin "chance" u engleskom jeziku) koristi za iskazivanje verovatnoće događaja u obliku procenta. Tako se verovatnoća događaja *A* jednaka 3/4 ponekad iskazuje u obliku šanse tog događaja od 75% (cf. Fulton, Mendez, Bastian, & Musal, 2012). Zbog toga je možda bolje bilo engleski termin "odds" prevesti kao "izgledi". Mi ćemo, ipak, kao

$$\boxed{\mathrm{O}(A)=\frac{\mathrm{P}(A)}{1-\mathrm{P}(A)}}$$

Za razliku od verovatnoće, šanse se kreću u intervalu od 0 do + ∞. Ukoliko je verovatnoća dešavanja nekog događaja jednaka 0.5, šanse tog događaja jednake su jedinici.
Iz obrasca ** jasno je da na osnovu poznatih šansi događaja *A* možemo jednostavno izračunati verovatnoću tog događaja:

$$\boxed{\mathrm{P}(A)=\frac{\mathrm{O}(A)}{1+\mathrm{O}(A)}}$$

Tako, na primer, ako su šanse događaja *A* jednake 4 tada je verovatnoća tog događaja jednaka 0.8.
Ilustrujmo razliku između verovatnoće i šansi na jednom vrlo jednostavnom primeru: verovatnoća da u jednom bacanju pravilne kocke padne „dvojka" jednaka je 1/6 (broj „povoljnih ishoda/ukupan broj mogućih ishoda) ali su šanse da padne „dvojka" jednake 1/5 (broj „povoljnih" ishoda/broj „nepovoljnih" ishoda, pri čemu je zbir broja „povoljnih" ishoda i broja „nepovoljnih" ishoda jednak ukupnom broju mogućih ishoda). Značenje numeričkih vrednosti šansi većih od jedinice najbolje će se razumeti ako se o šansama razmišlja na način koji sledi iz njihove definicije: šanse za neki događaj veće od 1 pokazuju *koliko puta je veća verovatnoća da se događaj desi od verovatnoće da se događaj ne desi.* Na primer, ako su šanse za neki događaj jednake 9 onda je verovatnoća da se događaj desi 0.9, a verovatnoća da se taj događaj ne desi 0.1. Ako su, pak šanse, za neki događaj jednake 99 tada je verovatnoća tog događaja 0.99. Uočimo, isto tako, da su šanse za događaj čija je verovatnoća manja od 0.5 manje od 1, a ukoliko je verovatnoća događaja blizu nuli tada su i šanse za događaj blizu nuli praktično jednake verovatnoći. /Dodatna pojašnjenja i zanimljivi primeri brkanja termina „šanse" (engl. odds) i „verovatnoća" (engl. probability) mogu se naći u Fulton et al., 2012/.

Uslovne šanse (događaja *A* pod uslovom dešavanja *B*) definišemo na isti način kao i šanse, s tom razlikom što sada koristimo uslovne verovatnoće:

$$\boxed{\mathrm{O}(A|B)=\frac{\mathrm{P}(A|B)}{1-\mathrm{P}(A|B)}}$$

U primeru iz Tabele ** šanse događaja „plašiti se pauka" jednake su:

$$\frac{430\,000}{570\,000}=0.75$$

a (uslovne) šanse događaja „plašiti se pauka pod uslovom da je reč o muškarcima iz date populacije" jednake su

---

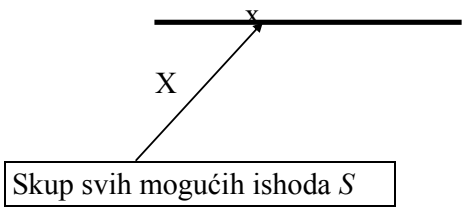
prevod termina "odds" koristiti termin "šanse" jer je takav prevod već ustaljen u metodološkim uđžbenicima na našem jeziku (cf. Todorović, 2008) a i zvuči nam više "statistički" od termina "izgledi". Kako bismo izbegli eventualne nesporazume termin "šanse" u značenju koje je ovde definisano koristićemo isključivo u obliku množine.

$$\frac{\frac{80\,000}{490\,000}}{\frac{410\,000}{490\,000}} = \frac{80\,000}{410\,000} = 0.20\,.$$

Šanse su veoma važan pojam za razumevanje mnogih statističkih postupaka, a u ovoj knjizi šanse ćemo koristiti pri prikazivanju postupaka za ispitivanje veza između kategoričkih varijabli (Glava **).

III. 3. Slučajna varijabla

Kao što smo već pomenuli, izvođenje nekog posmatranja u empirijskom istraživanju može se sa stanovišta teorije verovatnoće posmatrati kao izvođenje slučajnog eksperimenta jer rezultat, ishod opservacije, nije unapred poznat. Ukoliko se, na primer, zadaje neki merni instrument za ispitivanje agresivnosti svaki složaj odgovora na pitanja iz instrumenta se može posmatrati kao određeni ishod u skupu svih mogućih ishoda slučajnog eksperimenta "zadavanje skale agresivnosti slučajno izabranom pojedincu". Svakom složaju odgovora se prema pravilima pridružuje određeni skor (ili mera) koji predstavlja vrednost na nekoj varijabli. Pravilo za dodeljivanje određenog skora određenom složaju odgovora na stavkama skale je slučajna varijabla, a proces zadavanja skale i primene ovog pravila je posmatranje vrednosti slučajne varijable (Stilson, 1966, str.122). Pravilo (funkcija) koje dodeljuje jednu vrednost (realni broj) svakom ishodu iz skupa mogućih ishoda predstavlja slučajnu varijablu koja je definisana na skupu svih mogućih ishoda, tj. na skupu *S*. Dakle, slučajna varijabla (ili slučajna promenljiva, engl. random variable) bi prema ovom određenju bila funkcija koja svakom ishodu iz skupa *S* dodeljuje jednu (i samo jednu) brojčanu vrednost:



Na ovom crtežu X predstavlja **slučajnu varijablu**, a x je **vrednost** slučajne varijable koja odgovara određenom ishodu iz skupa mogućih ishoda..

Slučajna varijabla se može definisati i kao numerička funkcija koja svakom ishodu statističkog eksperimenta pridružuje jedan realan broj. Jedna vrednost slučajne varijable može biti povezana sa više elementarnih ishoda eksperimenta ali svakom elementarnom ishodu odgovara samo jedna vrednost slučajne varijable.

Na primer, *broj „pisama" u dva bacanja novčića* je slučajna varijabla koja može uzeti vrednosti 0, 1 i 2. Naime, mogući ishodi u dva bacanja novčića su: GG (oba puta je pala „glava"), PG (u prvom pokušaju je palo „pismo"a u drugom „glava"), GP (u prvom pokušaju je pala „glava" a u drugom „pismo") i PP (oba puta je palo „pismo"). Vrednost slučajne varijable koja iznosi 0 pridružujemo ishodu GG, vrednost koja iznosi 1 pridružujemo ishodima PG i GP, a vrednost 2 pridružujemo ishodu PP. Uočimo da je vrednost 1 date slučajne varijable povezana sa dva elementarna ishoda

ali da svakom od tih ishoda odgovara samo jedna vrednost (vrednost 1) od mogućih vrednosti slučajne varijable.

Naš poznati statističar B. Ivanović slučajnu varijablu definiše kao varijablu X koja na slučaj može uzimati jednu od svojih mogućih vrednosti x_1, x_2,..,x_n sa odgovarajućim verovatnoćama p_1,p_2,..,p_n tako da je zbir svih verovatnoća od p_1 do p_n jednak 1. Dakle, kada se saberu verovatnoće za sve moguće vrednosti slučajne varijable taj zbir mora biti jednak jedinici (Ivanović, 1966). Za potrebe ovog teksta Ivanovićevo određenje nam deluje sasvim prihvatljivo.

Na osnovu istog slučajnog eksperimenta moguće je definisati više različitih slučajnih varijabli. Na primer, u eksperimentu sa bacanjem novčića dva puta slučajne varijable mogu biti *broj pisama u dva bacanja novčića* ali i *broj „glava" u dva bacanja novčića*. Budući da su u bacanju novčića dva puta mogući ishodi PP, PG, GP i GG slučajna varijabla *broj pisama u dva bacanja novčića* će za ishod GG uzeti vrednost 0, za ishode PG i GP vrednost 1, a za ishod PP vrednost 2. S druge strane, slučajna varijabla *broj „glava"u dva bacanja novčića* će za ishod PP uzeti vrednost 0, za ishode PG i GP vrednost 1, a za ishod GG vrednost 2. Dakle, istim ishodima slučajnog eksperimenta zavisno od toga kako je određena slučajna varijabla mogu se pridružiti različite brojčane vrednosti. Međutim, za jednu istu slučajnu varijablu jednom ishodu slučajnog eksperimenta može se pridružiti samo jedna brojčana vrednost.

Važno je u ovom trenutku uočiti da je pojam slučajne varijable matematički pojam, tj. da je reč o apstraktnom matematičkom entitetu sa određenim svojstvima koja su matematički definisana. Dakle, pojam slučajne varijable spada u pojmove tzv. teorijskog sveta. Razumevanje pojma slučajne varijable veoma je važno jer primena tog pojma iz „teorijskog sveta" na stvarni, realni svet stoji u osnovi statističkog zaključivanja. Tokom daljeg izlaganja (a posebno u Glavi 7) videćemo u kakvoj je vezi taj pojam sa primenom statistike u analizi podataka dobijenih posmatranjem ili merenjem varijabli na uzorcima entiteta (jedinica posmatranja). I kakva je konkretna korist od tog pojma u razumevanju primene statistike.

Diskretne i kontinuirane slučajne varijable

Slučajne varijable mogu biti diskretne (diskontinuirane) i kontinuirane.

Diskretne slučajne varijable mogu uzeti bilo konačan bilo prebrojivo beskonačan broj vrednosti.[12] Na primer, diskretna slučajna varijabla može biti *broj „glava" u 5 bacanja novčića* ili *broj dece u porodici* ili *broj grešaka u eksperimentu*. Očigledno sve ove varijable iz primera mogu uzeti prebrojiv broj vrednosti. Mada je u navedenim primerima broj mogućih vrednosti ne samo prebrojiv, nego i konačan, teorijski je moguće da diskretna slučajna varijabla ima i beskonačan broj vrednosti.

Kontinuirane slučajne varijable mogu teorijski uzeti neprebrojivo beskonačan broj vrednosti. Naime, između bilo koje dve vrednosti takve varijable je neodređeno veliki broj vrednosti.[13] Primeri takvih varijabli su *inteligencija*; *visina*; *autoritarnost*; *brzina*

[12] Prebrojivo beskonačan znači da se skup mogućih vrednosti varijable može postaviti u odnos 1:1 sa pozitivnim celim brojevima tako da je svaki član skupa mogućih vrednosti povezan sa jednim i samo sa jednim pozitivnim celim brojem.

[13] Neprebrojivo beskonačan skup mogućih vrednosti varijable je skup čiji su elementi tako "gusti" da primena pozitivnih brojeva ne može da ih iscrpe.

reagovanja. Bitno je uočiti da mi ove varijable tretiramo kao teorijski kontinuirane bez obzira na to što su konkretne mere dobijene u empirijskim istraživanjima na takvim varijablama diskretne i realno (iz tehničkih razloga) ne mogu predstavljati kontinuum.

Distinkcija između diskretnih i kontinuiranih slučajnih varijabli bitna je pre svega matematički: matematički tretman varijabli i teorijski modeli koje primenjujemo opisujući „ponašanje“ ovih varijabli zavise od toga da li je reč o diskretnoj ili kontinuiranoj varijabli. „Ponašanje“ diskretnih slučajnih varijabli opisujemo distribucijama verovatnoća, a „ponašanje“ kontinuiranih slučajnih varijabli funkcijama gustine.

Distribucije verovatnoće, funkcije gustine i funkcije distribucije slučajnih varijabli

Distribucija verovatnoće diskretne slučajne varijable

„Ponašanje“ diskretne slučajne varijable se u teoriji verovatnoće opisuje skupom parova pri čemu svaki par čine određena moguća vrednost slučajne varijable i toj vrednosti pridružena verovatnoća da varijabla uzme datu vrednost:

X (moguće vrednosti)	P (verovatnoće)
$X = x_1$	$P(X = x_1)$
$X = x_2$	$P(X = x_2)$
.....	.....
$X = x_n$	$P(X = x_n)$
	$\sum_{i=1}^{n} P(X = x_i) = 1$

Uočimo iz Tabele ** da svakoj mogućoj vrednosti slučajne varijable odgovara određena verovatnoća pri čemu je zbir verovatnoća za sve vrednosti slučajne varijable jednak 1, što je nužno po definiciji.

Na primer, tabelarni prikaz distribucije verovatnoće za slučajnu varijablu *broj "pisama" u dva bacanja novčića* izgledao bi ovako:

X	P
X = 0	1/4 (0.25)
X = 1	1/2 (0.50)
X = 2	1/4 (0.25)

Umesto termina distribucija verovatnoće često se koristi termin funkcija mase verovatnoće (engl. probability mass function) ili funkcija verovatnoće (engl. probability function).

Funkcija distribucije za diskretnu slučajnu varijablu

Funkcija distribucije ili kumulativna funkcija distribucije (u daljem tekstu CDF, prema engleskom **C**umulative **d**istribution **f**unction) za diskretnu slučajnu varijablu, u oznaci $F(x)$, daje verovatnoću da slučajna varijabla X uzme vrednost manju od neke određene vrednosti x ili jednaku toj vrednosti:

$$F(x) = P(X \leq x) = \sum_{i:\, x_i < x} p_i$$

Pri tome, p_i je verovatnoća vrednosti x_i, a indeks *i* ispod sumacionog operatora odnosi se na sve vrednosti x_i koje su manje od x. Dakle, vrednost funkcije distribucije za dato x kod diskretnih slučajnih varijabli predstavlja zbir verovatnoća svih vrednosti manjih od x i verovatnoće vrednosti x.

Tabelarni prikaz CDF za *broj "pisama" u dva bacanja novčića* izgledao bi ovako:

$X \leq 0$	0.25
$X \leq 1$	0.75
$X \leq 2$	1

Uočimo da se verovatnoća u CDF za vrednost varijable jednaku 1 dobija sabiranjem verovatnoća za tu vrednost i verovatnoća za sve manje vrednosti:
0.25 + 0.50 = 0.75
Po istom principu, verovatnoća u CDF za vrednost varijable jednaku 2 (što je najveća moguća vrednost varijable iz ovog primera) jednaka je 1, tj. 0.25 + 0.50 + 0.25.

Za grafički prikaz distribucije verovatnoća diskretne slučajne varijable iz ovog primera može se upotrebiti tzv. štapićasti dijagram verovatnoća. Taj dijagram bi izgledao ovako:

**štapići

Uočimo da se na „apscisi“ dijagrama (što i nije prava apscisa jer su vrednosti varijable diskontinuirane) nalaze moguće vrednosti slučajne varijable a na ordinati verovatnoće za svaku od tih vrednosti. Dakle, visina štapića dignutog iz date tačke na „apscisi“ štapićastog dijagrama odgovara verovatnoći da slučajna varijabla uzme vrednost koja odgovara datoj tački.

Grafički prikaz funkcije distribucije (CDF) za varijablu iz primera izgledao bi ovako:
**grafik
Dakle, na ordinati su i u ovom prikazu verovatnoće ali sada tzv. kumulativne verovatnoće. Kada iz određene tačke na „apscisi“ povučemo liniju paralelno sa ordinatom do „susreta“ sa grafikom funkcije distribucije, pa potom paralelu sa „apscisom“, vrednost koju tako nalazimo na ordinati odgovara verovatnoći da varijabla uzme vrednost manju od date vrednosti ili vrednost jednaku datoj vrednosti. Uočimo da je funkcija distribucije za diskretnu slučajnu varijablu „stepeničasta“, tj. diskontinuirana funkcija.

<u>Funkcija gustine verovatnoće kontinuirane slučajne varijable</u>

Za kontinuiranu slučajnu varijablu umesto o verovatnoći da varijabla uzme određenu vrednost (ova je verovatnoća matematički u stvari jednaka nuli!)[14] govorimo o verovatnoći da slučajna varijabla uzme neku vrednost u datom intervalu od donje do gornje granice datog intervala. Umesto o verovatnoći da varijabla uzme neku određenu vrednost x govorimo o gustini verovatnoće varijable X na x /u oznaci $f(x)$/.[15]

Matematički gustina verovatnoće se definiše na sledeći način: ako je X kontinuirana slučajna varijabla i ako je $f(x)$ funkcija od X tako da za bilo koje dve vrednosti X, a i b (tako da je a < b)

$$P(a < X < b) = \int_a^b f(x)\, dx = \lim_{k \longrightarrow \infty} \left[\sum_{i=1}^{k} f(x_i)\, h \right]$$

f(x) je funkcija gustine verovatnoće za X (prema Winkler & Hays, 1975).[16]

Grafički prikaz funkcije gustine za kontinuirane slučajne varijable sastoji se zapravo iz ucrtavanja odgovarajuće neprekidne linije koja je definisana funkcijom gustine za datu varijablu. Apscisa tog grafika je realna osa u matematičkom smislu, a na ordinati su gustine, a ne verovatnoće!.
Na primer, grafik funkcije gustine za jednu kontinuiranu slučajnu varijablu može izgledati ovako:

Grafik **

Verovatnoća da kontinuirana slučajna varijabla uzme neku vrednost u intervalu od a do b dobija se računanjem površine koja nastaje kada se iz tačaka koje predstavljaju granice intervala (tačke a i b na grafiku) povuku paralele sa ordinatom do funkcije gustine.[17] U matematici se ove površine računaju određenim integralom.

Veoma je važno zapamtiti da ***celokupna površina pod grafikom funkcije gustine odgovara verovatnoći da slučajna varijabla uzme bilo koju vrednost u intervalu od najmanje moguće do najveće moguće vrednosti, što je po definiciji slučajne varijable jednako 1***.

---

[14] Čitaocima koji poznaju integralni račun ovo će biti jasno ako uzmu u obzir da se verovatnoća kod kontinuiranih varijabli računa preko integrala: budući da su u tom slučaju gornja i donja granica integrala jednake određeni integral je jednak nuli.

[15] Gustina verovatnoće je verovatnoća da slučajna varijabla uzme vrednost u intervalu od granica a do b: $P(a \le X \le b) = P(X \le b) - P(X < a)$. Ako veličinu intervala b - a označimo sa Δx tada je $b = a + \Delta x$. Verovatnoća intervala u odnosu na Δx je: $\frac{P(a \le X \le b)}{\Delta x} = \frac{P(X \le a + \Delta x) - P(X < a)}{\Delta x}$. Ako a fiksiramo dok Δx varira i recimo približava se 0, tada se ceo količnik menja i približava određenoj graničnoj vrednosti kako Δx ide bliže 0. Ova granica daje gustinu verovatnoće varijable X na vrednosti a. Grubo rečeno, gustina verovatnoće na a za najmanju promenu veličine intervala daje brzinu promene verovatnoće intervala sa gornjom granicom a (prema Winkler & Hays, 1975, str.129).

[16] Izraz $f(x)\, dx$ je površina pravougaonika čija je visina $f(x)$ a osnovica razmak h = (x - dx/2, x + dx/2).

[17] Ovu verovatnoću možemo napisati na dva načina: $P(a \le X \le b)$ ili $P(a < X < b)$ jer su ta dva izraza jednaka pošto je verovatnoća da varijabla uzme određenu vrednost kod kontinuiranih varijabli jednaka nuli.

Funkcija distribucije ili kumulativna funkcija gustine za kontinuiranu slučajnu varijablu

Funkcija distribucije ili kumulativna funkcija gustine (u daljem tekstu CDF, prema engleskom **C**umuative **d**ensity **f**unction)[18] za dato x daje verovatnoću da slučajna varijabla X uzme neku vrednost u intervalu od najmanje moguće vrednosti (x_{min}) do date vrednosti x.

Matematički CDF se definiše na sledeći način:

Ako je $f(x)$ funkcija gustine verovatnoće kontinuirane slučajne varijable, kumulativna funkcija gustine, u oznaci $F(x_a)$,

$$F(x_a) = P(X < x_a) = \int_{-\infty}^{x_a} f(x)\,dx \quad \text{za svako } x_a$$

daje verovatnoću da na slučaj uzeta vrednost varijable padne između $-\infty$ (ili x_{min}) i x_a (prema Winkler & Hays, 1975).

Grafički prikaz funkcije distribucije slučajne varijable čiju funkciju gustine smo prikazali na Grafiku ** dat je na Grafiku **

Kao što se iz Grafika ** može uočiti, na ordinati grafičkog prikaza funkcije distribucije su verovatnoće, baš kao i kod diskretnih varijabli. Prema tome, verovatnoća da kontinuirana slučajna varijabla uzme neku vrednost u intervalu od najniže moguće vrednosti do date vrednosti x očitava se na isti način kao kod diskretnih varijabli: iz tačke na apscisi koja odgovara datoj vrednosti x povučemo liniju paralelno sa ordinatom do preseka sa grafikom funkcije distribucije pa potom paralelu sa apscisom do ordinate. Vrednost koju tako nalazimo na ordinati odgovara traženoj verovatnoći.
Uočimo da je funkcija distribucije za kontinuiranu slučajnu varijablu kontinuirana funkcija. Razlika između funkcija distribucija za diskretnu i kontinuiranu slučajnu varijablu u pogledu kontinuiranosti često se služi za definisanje ova dva tipa slučajnih varijabli:slučajna varijabla je diskretna ako je njena funkcija distribucije stepeničasta a kontinuirana ako je njena funkcija distribucije kontinuirana funkcija.

Kvantil

Kvantil p (ili „kvantil reda p“) za kontinuiranu slučajnu varijablu X, u oznaci q_p, predstavlja najmanju vrednost slučajne varijable za koju važi

$$F(q_p) = P(X \le q_p) = p$$

pri čemu je p verovatnoća, tj. neki broj od 0 do 1, a $F(\cdot)$ funkcija distribucije slučajne varijable X.
Dakle, kvantil p je najmanja vrednost slučajne varijable za koju važi sledeće: verovatnoća da slučajna varijabla uzme neku vrednost manju ili jednaku kvantilu p jednaka je p. Na primer, kvantil 0.5 (čita se kao „kvantil nula pet“) je ona vrednost

[18] Istu skraćenicu CDF za funkciju distribucije namerno koristimo u ovom tekstu i za diskretne i za kontinuirane slučajne varijable bez obzira na to što su engleski nazivi ovih funkcija različiti. To činimo zato što se se u mnogim statističkim paketima funkcije za računanje vrednosti funkcije distribucije označavaju ovom skraćenicom za oba tipa varijabli.

slučajne varijable za koju važi da je verovatnoća da varijabla uzme neku vrednost manju od te vrednosti ili jednaku toj vrednosti jednaka 0.5 ili 50%. (Kvantil 0.5 inače predstavlja vrednost koju u statistici zovemo medijana). Kvantili slučajnih varijabli su veličine koje se veoma često koriste u statističkoj analizi podataka.

Očekivana vrednost i varijansa slučajne varijable

Očekivana vrednost slučajne varijable (engl. Expected value of the random variable)

Očekivanje, očekivana vrednost ili aritmetička sredina diskretne slučajne varijable X, u oznaci $E(X)$ definiše se na sledeći način:

$$E(X) = \sum_i x_i p_i$$

Pri tome, x_i su moguće vrednosti varijable a p_i su verovatnoće tih vrednosti.[19] Indeks i ispod operatora zbira označava da se sabiraju proizvodi vrednosti varijable i odgovarajućih verovatnoća za sve moguće vrednosti varijable.
Na primer, očekivanu vrednost za slučajnu varijablu *broj „pisama" u dva bacanja novčića* izračunali bismo na sledeći način:

$$E(X) = 0*0.25 + 1*0.50 + 2*0.25 = 1$$

Dakle, očekivana vrednost slučajne varijable je zbir svih proizvoda mogućih vrednosti varijable i odgovarajućih verovatnoća.

Očekivana vrednost za kontinuirane slučajne varijable definiše se na sledeći način:

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

pri čemu je $f(x)$ funkcija gustine slučajne varijable.

Smisao očekivane vrednosti najlakše je ilustrovati primerima iz igara na sreću. Na primer, u nekoj lutriji prodato je svih 10 000 srećki a postoje tri vrste dobitka (1 000 evra, 500 evra i 100 evra)(modifikovano prema Freund, 1966, str.90). *Iznos dobitka neke osobe* predstavlja slučajnu varijablu, a očekivani dobitak po srećki predstavlja matematičko očekivanje ili očekivanu vrednost. Ako, dakle, raspodelimo ukupni dobitak od 1600 evra na vlasnike 10 000 srećki to iznosi 0.16 evra po srećki.
Saobrazno definiciji datoj u **, očekivanu vrednost bismo u ovom slučaju izračunali tako što pomnožimo moguće vrednosti slučajne varijable *iznos dobitka neke osobe* (a to su vrednosti 0, 100, 500 i 1000) sa odgovarajućim verovatnoćama i potom sve dobijene proizvode saberemo:
0*0.9997 + 100*0.0001 + 500*0.0001 + 1000*0.0001 = 0.16
Naravno, očekivana vrednost u ovakvim slučajevima ne predstavlja subjektivno očekivanje osobe koja kupuje srećku. Ta osoba subjektivno očekuje da dobije neku veliku nagradu, inače ne bi kupovala srećku za očekivani dobitak od 0.16 evra! Vrednost 0.16 je matematičko očekivanje koje govori o tome koliki dobitak se u

[19] Mada smo verovatnoću događaja do sada obeležavali velikim slovom P u obrascima se radi jednostavnosti verovatnoća za vrednosti slučajne varijable uobičajeno označava malim slovom p.

proseku (po osobi) može očekivati. Uočimo istovremeno i probleme koji se mogu javiti u korišćenju očekivane vrednosti: praktično nijedan vlasnik srećke neće dobiti ništa a očekivana vrednost je ipak veća od nule jer će samo tri vlasnika srećki dobiti znatno više od nule! Znatno bi zapravo bilo smisaonije i prirodnije da u ovom slučaju očekivana vrednost bude 0 ali ona to matematički nije. (U tome se možda krije tajna privlačnosti igara na sreću...).

Očekivanom vrednošću slučajne varijable u statistici se definiše aritmetička sredina populacije i u tom se slučaju uobičajeno umesto oznake $E(X)$ koristi oznaka μ (grčko slovo „mi“).

Varijansa slučajne varijable (engl. Variance of the random variable)

Varijansa slučajne varijable X, u oznaci V(X), predstavlja očekivanu vrednost kvadriranih odstupanja vrednosti varijable od očekivane vrednosti:

$$V(X) = E[X - E(X)]^2$$

Uočimo, dakle, da varijansa predstavlja neku vrstu očekivane vrednosti. Ona, dakle, pokazuje koliko velika kvadrirana odstupanja vrednosti na nekoj slučajno varijabli od njihove očekivane vrednosti možemo u proseku očekivati.
Ukoliko je slučajna varijabla diskretna, tada se varijansa može definisati na osnovu opšte definicije i na sledeći način:

$$V(X) = \sum_i (x_i - \mu)^2 p_i$$

Kao i u izrazu za očekivanu vrednost, x_i su moguće vrednosti varijable a p_i su verovatnoće tih vrednosti. Oznakom μ označena je očekivana vrednost.
Dakle, varijansa diskretne slučajne varijable predstavlja zbir svih proizvoda parova, pri čemu prvi član para predstavlja kvadrirano odstupanje date vrednosti na varijabli od očekivane vrednosti varijable a drugi član para čini verovatnoća za datu vrednost na varijabli. Uočimo, isto tako, da ***varijansa ne može imati negativnu vrednost*** budući da su verovatnoće p_i po definiciji nenegativne a kvadrat kojim se množe verovatnoće ne može biti negativan broj.

Varijansa za kontinuirane slučajne varijable definiše se na sledeći način:

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

pri čemu je $f(x)$ funkcija gustine slučajne varijable.

Varijansa varijable *broj „pisama“ u dva bacanja novčića* bila bi:
$V(X) = (0-1)^2*0.25 + (1-1)^2*0.50 + (2-1)^2*0.25 = 0.25 + 0 + 0.25 = 0.5.$

Varijansu za slučajnu varijablu *iznos dobitka neke osobe* izračunali bismo na sledeći način:
$V(X) = (0 - 0.16)^2*0.9997 + (100 - 0.16)^2*0.0001 + (500 - 0.16)^2*0.0001 + (1000 - 0.16)^2*0.0001 = 0.026 + 0.997 + 24.984 + 99.968 = 125.98.$
Dakle, očekivano kvadrirano odstupanje vrednosti na varijabli od očekivane vrednosti varijable iznosi 125.98. Uočimo da ovako velikoj vrednosti varijanse najviše

doprinose samo tri vrednosti (100, 500 i 1000) koje su znatno udaljene od ogromne većine preostalih vrednosti na varijabli.
U statistici se varijansa slučajne varijable koristi kao varijansa populacije i uobičajeno označava oznakom σ^2 („sigma kvadrat“), a pozitivni kvadratni koren varijanse populacije predstavlja *standardnu devijaciju* populacije, uobičajeno u oznaci σ.

Parametri oblika distribucije slučajne varijable

Kao što sam naziv ovih parametara sugeriše, parametri oblika distribucije (engl. shape parameters) slučajne varijable upućuju na oblik koji ova distribucija ima. Dva ključna parametra oblika distribucije su skjunis i kurtozis.[20] Skjunis ukazuje na simetričnost distribucije, a kurtozis na stepen u kojem su vrednosti slučajne varijable nagomilane na krajevima raspodele.

Skjunis (engl. Skewness), u oznaci α_3, ukoliko je slučajna varijabla diskretna, definisan je na sledeći način:

$$\alpha_3 = \sum_i \frac{(x_i - \mu)^3}{\sigma^3} p_i$$

Oznaka x_i označava moguće vrednosti slučajne varijable, p_i su verovatnoće tih vrednosti, a σ je standardna devijacija, tj. pozitivni kvadratni koren varijanse. Cifra 3 u supskriptu oznake za skjunis potiče od naziva sume u brojiocu ($\sum_i (x_i - \mu)^3 p_i$): ova suma se u teoriji verovatnoće, zbog srodnosti sa momentima iz fizike, zove *treći centralni moment.*[21]

Skjunis pokazuje da li je distribucija slučajne varijable simetrična ili asimetrična: ukoliko je ovaj parametar jednak nuli distribucija varijable je simetrična (iako postoje određeni izuzeci od ovog pravila, cf. Wheeler, 2011a), ako je skjunis manji od nule distribucija je negativno asimetrična a ako je skjunis veći od nule distribucija varijable je pozitivno asimetrična.

Skjunis za distribuciju kontinuirane slučajne varijable definiše se na sledeći način:

$$\alpha_3 = \int_{-\infty}^{+\infty} \frac{(x - \mu)^3}{\sigma^3} f(x)\, dx$$

pri čemu je $f(x)$ funkcija gustine slučajne varijable.

Skjunis za distribuciju slučajne varijable *broj „pisama“ u dva bacanja novčića* izračunali bismo na sledeći način:

$\alpha_3 = [(0 - 1)^3 * 0.25]/0.7071^3 + [(1 - 1)^3 * 0.5]/ 0.7071^3 + [(2 - 1)^3 * 0.25]/ 0.7071^3 = -0.7071 + 0 + 0.7071 = 0.$

[20] Termin skjunis potiče od engleske reči skew što znači kos ili iskošen, dok termin kutrozis potiče od grčkog kurtos (κυρτοσ**) što znači kriv, zasvođen ili izbočen.

[21] Centralni momenti, za razliku od običnih momenata koji uključuju odstupanja vrednosti slučajne varijable u odnosu na nulu, uključuju odstupanja vrednosti na varijabli u odnosu na očekivanu vrednost, tj. aritmetičku sredinu slučajne varijable. Na primer, drugi moment je $\sum_i (x_i - 0)^2 p_i = \sum_i x_i^2 p_i$ a drugi *centralni* moment je $\sum_i (x_i - \mu)^2 p_i$. Očigledno, varijansa je zapravo drugi centralni moment.

Dakle, distribucija ove varijable je simetrična. Na ovom primeru može se lako uočiti i šta znači simetričnost distribucije: uočimo da je aritmetička sredina distribucije jednaka 1 a da je vrednost na varijabli manja od aritmetičke sredine (vrednost 0) podjednako udaljena od aritmetičke sredine kao i vrednost na varijabli koja je veća od aritmetičke sredine (vrednost 2) te da ove jednako udaljene vrednosti imaju jednake verovatnoće. Uopštavanjem sa ovog jednostavnog primera simetrična distribucija za diskretnu slučajnu varijablu mogla bi se opisati kao distribucija u kojoj je broj vrednosti manjih i većih od aritmetičke sredine podjednak, a verovatnoće vrednosti koje su podjednako udaljene od aritmetičke sredine na jednu i drugu stranu jednake.

Kurtozis (engl. Kurtosis), u oznaci α_4, ukoliko je slučajna varijabla diskretna, definisan je na sledeći način:

$$\alpha_4 = \sum_i \frac{(x_i - \mu)^4}{\sigma^4} p_i$$

Oznake imaju isto značenje kao u obrascu za skjunis. Cifra 4 u supskriptu oznake za kurtozis potiče od naziva sume u brojiocu ($\sum_i (x_i - \mu)^4 p_i$): ova suma se inače zove

četvrti centralni moment.

Kurtozis pokazuje u kojoj meri distribucija slučajne varijable ima izražene verovatnoće za vrednosti koje su na krajevima distribucije u odnosu na verovatnoće vrednosti u sredini distribucije: kurtozis definisan obrascem** ima vrednost 3 za tzv. normalnu raspodelu (o kojoj će biti reči u nastavku teksta u ovoj glavi). Često se od vrednosti ovako definisanog kurtozisa oduzima vrednost 3 kako bi se postiglo da kurtozis za tzv. normalnu raspodelu bude jednak nuli. U tom slučaju vrednost kurtozisa manja od nule ukazuje na to da raspodela ima manje razvučene i manje izdignute krajeve od normalne raspodele, dok kurtozis veći od nule ukazuje na raspodelu sa razvučenijim i(li) izdignutijim krajevima od normalne raspodele. Mi ćemo postupak oduzimanja vrednosti 3 od vrednosti kurtozisa (tzv. „kalibrisanje“ kurtozisa) koristiti u ovoj knjizi kada god se budemo bavili realnim podacima budući da je to u skladu sa načinom na koji program SPSS (koji se najčešće koristi u analizama podataka u psihologiji i srodnim oblastima) računa kurtozis.

Kurtozis za distribuciju kontinuirane slučajne varijable definiše se na sledeći način:

$$\alpha_4 = \int_{-\infty}^{+\infty} \frac{(x - \mu)^4}{\sigma^4} f(x)\, dx$$

pri čemu je *f(x)* funkcija gustine slučajne varijable.

Kurtozis za distribuciju slučajne varijable *broj „pisama“ u dva bacanja novčića* izračunali bismo na sledeći način:

$\alpha_3 = [(0 - 1)^4 * 0.25] / 0.7071^4 + [(1 - 1)^4 * 0.5] / 0.7071^4 + [(2 - 1)^4 * 0.25] / 0.7071^4 = 1 + 0 + 1 = 2$.

(Uočimo kako vrednost jednaka aritmetičkoj sredini zapravo nimalo ne doprinosi vrednosti kurtozisa. Dakle, što su vrednosti varijable dalje od aritmetičke sredine, to one više doprinose vrednosti kurtozisa. To je važno uočiti kako bi se razumelo značenje ovog parametra).

Ako oduzmemo 3 od vrednosti kurtozisa dobijamo vrednost -1. Negativna vrednost koja se dobije kada se od kurtozisa koji je računat po obrascu ** oduzme 3 govori o tome da su verovatnoće krajnjih vrednosti ove slučajne varijable relativno niske u odnosu na verovatnoće središnjih vrednosti.
(Čitaocima sa solidnim matematičkim predznanjem koji žele da potpunije razumeju skjunis i kurtozis kao parametre oblika distribucije slučajne varijable preporučujem tekst koji se u referencama nalazi pod Wheeler, 2011a).

Uloga teorije o slučajnim varijablama i distribucijama slučajnih varijabli u statističkoj analizi podataka

Slučajna varijabla, distribucija verovatnoća, funkcija gustine i funkcija distribucije su apstraktni matematički entiteti koji su definisani u okviru teorije verovatnoće na način sličan onome koji smo prikazali (prikaz koji smo mi dali je, naravno, unekoliko pojednostavljen). U okviru teorije verovatnoće definisano je mnoštvo familija funkcija (distribucija verovatnoća za diskretne slučajne varijable i funkcija gustine za kontinuirane slučajne varijable). Svaka familija ovih distribucija i funkcija gustine definisana je parametrima, tj. „opštim" veličinama koje kada uzmu određenu brojnu vrednost iz skupa mogućih vrednosti definišu određenu distribuciju u okviru date familije.
Na primer, binomna familija distribucija verovatnoća definisana je sledećom formulom:

$$\mathrm{Bin}(x; n, \pi) = \binom{n}{x} \pi^x (1-\pi)^{n-x}, \quad \text{za } x = 0, 1, \ldots, n$$

$$\text{pri čemu je } n = 1, 2, \ldots \text{ i } 0 \le \pi \le 1$$

U ovoj formuli $\binom{n}{x}$ je binomni koeficijent,[22] a n i π su parametri distribucije.
Parametar n, kao što se vidi, može uzeti celobrojne vrednosti od 1 pa nadalje, a parametar π vrednosti u segmentu od 0 do 1.
Očekivana vrednost, tj. aritmetička sredina binomne raspodele, u oznaci μ_{Bin}, i varijansa, u oznaci σ^2_{Bin} definisane su na sledeći način:

$$\mu_{Bin} = n*\pi$$

$$\sigma^2_{Bin} = n*\pi*(1-\pi)$$

Uočimo da kada opšte veličine n i π uzmu određene vrednosti tada imamo konkretnu binomnu distribuciju. Na primer, ako je n = 10, a π = 0.515 tada opšti obrazac za binomnu familiju distribucija dobija sledeći oblik:

$$\mathrm{Bin}(x; 10, 0.515) = \binom{10}{x} 0.515^x (1-0.515)^{10-x}, \quad \text{za } x = 0, 1, \ldots, n$$

Sada imamo posla sa funkcijom iz koje za određeno x lako dobijamo vrednost funkcije Bin(x). Na primer, ako je x = 6, tada je Bin(6; 10, 0.515) = $210*0.515^6*0.485^4 \approx 0.22$. Slučajna varijabla bi u ovom slučaju mogla biti *broj*

[22] Binomni koeficijent je objašnjen u Matematičkom pojmovniku pod odrednicom **Osnovni pojmovi i pravila kombinatorike** (tačka 6. Kombinacije).

dečaka u 10 rađanja, a u primeru je binomna raspodela upotrebljena da se izračuna verovatnoća da se u 10 nezavisnih rađanja rodi 6 dečaka. Na osnovu binomne raspodele mogli bismo izračunati i verovatnoću da se u dva bacanja novčića ne dobije nijedno „pismo" na drugačiji način od onog koji smo primenili u trećem primeru kod apriornog određenja verovatnoće. Budući da je n = 2, $\pi = 0.5$ a x = 0, ovu verovatnoću bismo izračunali na sledeći način:

$$P(\text{nula"pisama"}) = \binom{2}{0} 0.5^{0} * (1-0.5)^{2-0} = \frac{2!}{(2-0)!\,0!} 1 * 0.5^{2} = 1*1*0.25 = 0.25$$

Binomna familija distribucija verovatnoća najčešće se koristi kao matematički model za slučajne varijable *broj „uspeha" u n pokušaja* iz slučajnih eksperimenata u kojima su u svakom pokušaju moguća samo dva (uzajamno isključiva) ishoda, pojedinačni pokušaji nezavisni, a verovatnoća „uspeha" u svakom pokušaju konstantna iz pokušaja u pokušaj. „Uspeh" je u ovom slučaju prosto jedan od dva uzajamno isključiva ishoda koji nas zanima. Na primer, ako ovaj model primenimo na eksperiment sa bacanjem novčića, a zanima nas padanje novčića na stranu „glava" onda bi „uspeh" bio da je pala „glava". Mogli bismo, recimo, primenjujući binomnu raspodelu kao matematički model takvog slučajnog eksperimenta odrediti verovatnoću da u 10 pokušaja (bacanja novčića) padne 6 puta „glava", ako pretpostavimo da je verovatnoća padanja„glave" u svakom pokušaju konstantna i iznosi 0.5. Dakle, u tom slučaju parametri binomne raspodele bili bi n = 10 i $\pi = 0.5$. Slučajna varijabla u tom slučaju bila bi *broj „glava" u 10 bacanja novčića* a vrednost te slučajne varijable za koju računamo verovatnoću jeste vrednost 6. Verovatnoću da u 10 bacanja ispravnog novčića padne 6 puta „glava" (a koja je približno jednaka 0.21) izračunali bismo na isti način na koji smo prethodno izračunali verovatnoću da se u 10 nezavisnih rađanja rodi 6 dečaka (ta verovatnoća bila bi približno jednaka 0.21). Uočimo iz primera koje smo prikazali da se jedna ista familija distribucija verovatnoća može primeniti za opis ponašanja različitih slučajnih varijabli.
Dakle, matematičari koji se bave teorijom verovatnoće definisali su veliki broj različitih funkcija (distrubucija verovatnoća ili funkcija gustine) kojima možemo opisivati „ponašanje" različitih slučajnih varijabli. Kakve veze teorija slučajnih varijabli i matematički modeli „ponašanja" slučajnih varijabli imaju sa statističkom analizom podataka? Razumevanje ove veze od ključnog je značaja za razumevanje same statističke analize podataka. Naime, sve varijable kojima se bave istraživači u psihologiji i srodnim oblastima, teorijski se mogu posmatrati kao „otelotvorenja" slučajnih varijabli iz teorije verovatnoće. Prema tome, različite distribucije verovatnoća i funkcije gustine kojima se u teoriji verovatnoće opisuje „ponašanje" slučajnih varijabli koriste se kao modeli „ponašanja" varijabli kojima se bave istraživači u psihologiji i srodnim oblastima, tj. kao modeli raspodele ovih varijabli u populaciji. Na primer, tzv. normalna (ili Laplas-DeMoavr-Gausova) funkcija gustine (koju ćemo prikazati u nastavku teksta) često se smatra adekvatnim matematičkim modelom za opis distribucije vrednosti mnogih psiholoških varijabli (inteligencije, ekstraverzije, neuroticizma, verbalne sposobnosti) u populaciji. Drugim rečima, pretpostavlja se da kada bismo raspolagali merama na tim varijablama za sve članove populacije onda bi se distribucija tih mera (mere i verovatnoće tih mera) mogla matematički opisati modelom normalne funkcije gustine.[23] Mnoge psihološke

[23] Na ovoj pretpostavci zasniva se primena mnogih statističkih postupaka u psihologiji premda nije sasvim izvesno da je ova pretpostavka opravdana. U Glavi ** razmotrićemo probleme koji se u

karakteristike ljudi posmatraju se kao da predstavljaju rezultat sabiranja uticaja mnoštva nezavisnih faktora, te otuda i pretpostavka o sveprisutnoj normalnosti distribucija velikog broja psihololoških varijabli. Za one varijable za koje znamo da su rezultat nekih drugih mehanizama, a ne sabiranja uticaja mnogih nezavisnih činilaca, neopravdano je pretpostaviti da se normalno distribuiraju. Na primer, u takve varijable bi spadale varijable koje nastaju kao rezultat multiplikativnog dejstva malog broja činilaca.

S druge strane, svaki potencijalni rezultat slučajnog uzorka na nekoj varijabli pre izvođenja istraživanja teorijski se posmatra kao slučajna varijabla sa istom distribucijom koju ima data varijabla u populaciji (o tome ćemo podrobnije govoriti u Glavi 7). Dakle, pre nego što izmerimo inteligenciju datog ispitanika iz slučajnog uzorka, njegov potencijalni rezultat se može posmatrati kao slučajna varijabla koja ima normalnu raspodelu, tj. istu raspodelu koju ima varijabla inteligencija u populaciji. Ovakav pristup je moguć samo ako je uzorak slučajan (slučajni uzorak ćemo objasniti u Glavi 7).

Na kraju, statističke mere (statistici) koje dobijamo na osnovu rezultata ispitanika iz slučajnog uzorka (npr. aritmetička sredina uzorka) teorijski se mogu posmatrati kao slučajne varijable čije „ponašanje“ se može opisati određenim distribucijama verovatnoće ili funkcijama gustine slučajnih varijabli. Takav tretman statističkih mera dobijenih na slučajnim uzorcima sledi na osnovu jedne od ključnih teorema iz teorije o slučajnim varijablama:

Ako je X slučajna varijabla svaka funkcija od X je takođe slučajna varijabla.

Dakle, budući da se, pre izvođenja istraživanja, pojedinačni rezultati (opservacije ili elementi) slučajnog uzorka mogu posmatrati kao slučajne varijable onda su i statistici koji su funkcija tih slučajnih varijabli takođe slučajne varijable. Tvrdnja da su statistici funkcija slučajnih varijabli sledi naprosto iz načina, tj. formula kojima su statistici definisani. Na primer, aritmetička sredina kao statistik, u oznaci M, definisana je na sledeći način:

$$M = \frac{\sum_{i=1}^{n} X_i}{n}$$

pri čemu su X_i rezultati pojedinih jedinica posmatranja iz slučajnog uzorka, a n broj rezultata, tj. veličina uzorka.

Očigledno, ako se pre izvođenja istraživanja svako X_i posmatra kao slučajna varijabla onda je i funkcija od tih slučajnih varijabli, u ovom slučaju statistik M, slučajna varijabla. Na tretiranju statistika kao slučajnih varijabli zasnovani su postupci statističkog zaključivanja koje ćemo razmatrati u ovoj knjizi počevši od Glave 7.

Normalna ili DeMoavr-Laplas-Gausova funkcija gustine

Familija normalnih funkcija gustine definisana je sledećim izrazom:

---

statističkoj analizi podataka javljaju kada imamo razloga da sumnjamo u opravdanost ove pretpostavke. Moglo bi se svakako postaviti i opštije pitanje: da li su istraživači u psihologiji i srodnim oblastima odabrali prave teorijske modele iz teorije verovatnoće za opis „ponašanja“ varijabli kojima se bave i da li je možda potrebno definisati nove teorijske modele koji su primereniji „ponašanju“ ovih varijabli. Na ovo veoma teško pitanje koje zadire u same temelje primene statistike u psihologiji i srodnim oblastima nemamo jasan odgovor ali bi istraživači u ovim oblastima svakako trebalo da ga budu svesni.

$$\mathcal{N}(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad -\infty < x < \infty$$

U ovom izrazu x su moguće vrednosti slučajne varijable X, π je konstanta koja iznosi 3,1415 (Ludolfov broj) a e je konstanta (2.7183) koja predstavlja osnovu prirodnog, Neperovog logaritma. Opšte veličine μ i σ su tzv. parametri kojima je definisana familija normalnih funkcija i predstavljaju aritmetičku sredinu i standardnu devijaciju (pozitivni kvadratni koren varijanse) slučajne varijable, tim redom. Kada parametri μ i σ uzmu određene vrednosti tada je u okviru ove familije definisana određena normalna funkcija gustine ili normalna raspodela.[24] Normalna raspodela se u statistici uobičajeno skraćeno označava nekom varijantom velikog slova N sa oznakama parametara u zagradi. Mi ćemo je u ovom tekstu označavati na sledeći način:

$\mathcal{N}(\mu,\sigma)$. **font Matura Script Capitals

Na primer, normalnu raspodelu sa parametrima μ = 100 i σ = 15 označili bismo ovako:

$\mathcal{N}(100,15)$.

Grafički prikaz normalne raspodele sa parametrima μ = 100 i σ = 15 dat je na Grafiku **.

Grafik **

Kao što iz Grafika ** možemo da uočimo normalna funkcija gustine je zvonasta i simetrična u odnosu na najvišu ordinatu, tj. ordinatu dignutu iz tačke koja je jednaka aritmetičkoj sredini varijable. Kriva se asimptotski približava X osi. Budući da funkcija ima samo jedan vrh, normalna raspodela spada u tzv. unimodalne raspodele (raspodele koje imaju jednu vrednost koja je najčešća). Pošto je ukupna površina ispod krive jednaka 1 ili 100% (po definiciji funkcije gustine slučajne varijable), površine levo ili desno od najviše ordinate jednake su 0.5 (ili 50%). Kao i kod svake funkcije gustine, verovatnoća da slučajna varijabla sa normalnom funkcijom gustine uzme neku vrednost u određenom intervalu predstavljena je površinom koja se dobija kada se iz granica intervala povuku paralele sa ordinatom do krive. Na primer:

- površina između ordinata koje su dignute u tačkama μ - σ i μ + σ iznosi 0.6826;
- površina između ordinata koje su dignute u tačkama μ - 2σ i μ + 2σ iznosi 0.9544;
- površina između ordinata koje su dignute u tačkama μ - 3σ i μ + 3σ iznosi 0.9974;

Uočimo, dakle, da je verovatnoća da slučajna varijabla koja se „ponaša“ po normalnoj raspodeli, tj. koja ima normalnu funkciju gustine uzme neku vrednost u intervalu ograničenom vrednošću koja je za jednu standardnu devijaciju manja od aritmetičke sredine i vrednošću koja je za jednu standardnu devijaciju veća od aritmetičke sredine iznosi 0.6826 (ili 68.26%). Isto tako, verovatnoća da slučajna varijabla koja ima normalnu funkciju gustine uzme neku vrednost u intervalu od vrednosti koja je za tri standardne devijacije manja od aritmetičke sredine do vrednosti koja je za tri standardne devijacije veća od aritmetičke sredine iznosi 0.9974 (gotovo 100%).

[24] Bez obzira na to što je naziv „normalna funkcija gustine“ matematički korektan naziv za ovu funkciju, budući da je reč o kontinuiranim slučajnim varijablama, mi ćemo u ovom tekstu, jer je to u psihologiji uobičajeno, koristiti i naziv „normalna raspodela“ .

Uočimo i da je kvantil 0.5, tj. medijana za normalnu raspodelu baš vrednost iznad koje je najviša ordinata, a ta vrednost je istovremeno i aritmetička sredina.

Na Grafiku ** prikazane su površine pod normalnom krivom koje su veoma važne za statističku analizu podataka.

Grafik **

Za statističku analizu podataka posebno su važni sledeći kvantili normalne raspodele:
$q_{0.025} = \mu - 1.96\sigma$
$q_{0.975} = \mu + 1.96\sigma$
$q_{0.005} = \mu - 2.58\sigma$
$q_{0.995} = \mu + 2.58\sigma$
Dakle,
Uočimo i da je kvantil 0.5 za varijablu koja ima normalnu raspodelu jednak aritmetičkoj sredini varijable:

$$q_{0.5} = \mu$$

/Istovremeno, kvantil 0.5 za varijablu koja ima normalnu raspodelu jednak je medijani i modalnoj vrednosti (modu) varijable/.

Činjenica da slučajna varijabla X ima određenu raspodelu piše se uobičajeno tako što se posle oznake varijable stavlja znak ~ („tilda“) i potom se daje oznaka raspodele. Na primer, skraćenica $X \sim Bin(n, \pi)$ označava da varijabla X ima binomnu raspodelu sa parametrima n i π, a $X \sim \mathcal{N}(\mu,\sigma)$ označava da varijabla X ima normalnu raspodelu sa parametrima μ i σ.

Ako $X \sim \mathcal{N}(\mu, \sigma)$ onda je funkcija distribucije za varijablu X, u oznaci $F(x_a)$, data sledećim izrazom:

$$F(x_a) = \int_{-\infty}^{x_a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx, \qquad -\infty < x_a < \infty$$

Standardizovana normalna funkcija gustine

U statističkoj analizi podataka veoma je važna normalna raspodela čija je aritmetička sredina jednaka 0, a standardna devijacija jednaka 1, tj. normalna raspodela sa parametrima $\mu = 0$ i $\sigma = 1$. Ako μ i σ iz opšte formule za familiju normalnih raspodela zamenimo sa 0 i 1, dobićemo obrazac kojim je definisana tzv. standardizovana normalna funkcija gustine:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Slučajna varijabla koja ima standardizovanu normalnu raspodelu uobičajeno se označava slovom Z te je u ovom obrascu x iz opšte formule zamenjeno sa z. Isto tako, standardizovana funkcija gustine se u teoriji verovatnoće uobičajeno označava sa $\phi(z)$, a ne $\mathcal{N}(z)$ pa smo i mi koristili ovu uobičajenu oznaku. Uočimo, dakle, da je $\phi(z)$ specijalni slučaj opšteg obrasca normalne raspodele ** u kojem je u imeniocu količnika koji se nalazi ispred izraza e izostavljena σ jer je jednaka 1, dok je u

eksponentu umesto $\left(\frac{x-\mu}{\sigma}\right)$ stavljeno z, jer je $z=\left(\frac{x-\mu}{\sigma}\right)$. Dakle, varijabla Z je naprosto linearna transformacija varijable X. (Linearne transformacije će biti podrobno objašnjene u Glavi 6).

Funkcija distribucije za varijablu koja ima standardizovanu normalnu raspodelu, uobičajeno u oznaci $\Phi(z_a)$, definiše se na sledeći način:

$$\Phi(z_a)=\int_{-\infty}^{z_a}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz, \qquad -\infty < z_a < \infty$$

Grafički prikaz standardizovane normalne funkcije gustine dat je na Grafiku **, a standardizovana normalna funkcija distribucije na Grafiku **.
Za statističku analizu podataka veoma su važni sledeći kvantili standardizovane normalne raspodele (umesto opšte znake q_p za kvantil, u slučaju standardizovane normalne raspodele uobičajeno se koristi oznaka z_p):

$z_{0.025} = -1.96$
$z_{0.975} = +1.96$
$z_{0.005} = -2.58$
$z_{0.995} = +2.58$

<u>Zapamtite</u>:

Verovatnoća da slučajna varijabla koja ima standardizovanu normalnu raspodelu uzme neku od vrednosti u intervalu od $z_{0.025}$ do $z_{0.975}$, tj. od -1.96 do +1.96 jednaka je 0.95 ili 95%.

Verovatnoća da slučajna varijabla koja ima standardizovanu normalnu raspodelu uzme neku od vrednosti u intervalu od $z_{0.005}$ do $z_{0.995}$, tj. od -2.58 do +2.58 jednaka je 0.99 ili 99%.

Verovatnoća da slučajna varijabla koja ima standardizovanu normalnu raspodelu uzme neku od vrednosti manjih od $z_{0.025}$ (tj. manjih od -1.96) ili neku od vrednosti većih od $z_{0.975}$ (tj. većih od +1.96) jednaka je 1 - 0.95, tj. 0.05 ili 5%.

Verovatnoća da slučajna varijabla koja ima standardizovanu normalnu raspodelu uzme neku od vrednosti manjih od $z_{0.005}$ (tj. manjih od -2.58) ili neku od vrednosti većih od $z_{0.995}$ (tj. većih od +2.58) jednaka je 1 - 0.99, tj. 0.01 ili 5%.

Svaka slučajna varijabla X koja ima normalnu raspodelu sa parametrima μ i σ može se prevesti u varijablu Z koja ima standardizovanu normalnu raspodelu veoma jednostavnom transformacijom:

$$Z=\frac{X-\mu}{\sigma}$$

Ovo sledi na osnovu pravila iz teorije verovatnoće o distribuciji slučajnih varijabli pri određenim transformacijama slučajnih varijabli, što ovde nećemo podrobnije

obrazlagati.[25]

Zapamtite: Ako slučajna varijabla X ima normalnu raspodelu sa parametrima μ i σ, onda slučajna varijabla Z, pri čemu je $Z = (X - \mu)/\sigma$, ima standardizovanu normalnu raspodelu, tj. normalnu raspodelu sa parametrima $\mu = 0$ i $\sigma = 1$.

Uočimo da ukoliko varijabla X ima normalnu raspodelu sa parametrima μ i σ, tada vrednosti na toj varijabli koje su definisane sa $\mu \pm z^*\sigma$ uvek odgovaraju onim vrednostima jednakim $\pm z$ na varijabli Z koja ima standardizovanu normalnu raspodelu. U kojem smislu „odgovaraju“? U smislu udaljenosti od odgovarajuće aritmetičke sredine kada se ta udaljenost iskazuje u jedinicama standardne devijacije. Na primer, ako varijabla X ima normalnu raspodelu sa parametrima $\mu = 100$ i $\sigma = 15$ tada vrednost na varijabli X jednaka 100 + 1.96*15 (vrednost 129.4) odgovara vrednosti +1.96 na varijabli Z. Isto tako, vrednost na varijabli X jednaka 100 - 1.96*15 (vrednost 70.6) odgovara vrednosti -1.96 na varijabli Z. Vrednost 129.4 na varijabli X podjednako je udaljena od aritmetičke sredine varijable X kao i vrednost +1.96 na varijabli Z od aritmetičke sredine varijable Z, ukoliko ovu udaljenost iskazujemo u jedinicama odgovarajuće standardne devijacije. Dakle, vrednost 129.4 na varijabli X, i njoj odgovarajuća vrednost +1.96 na varijabli Z veće su od “svoje” aritmetičke sredine za 1.96 odgovarajućih standardnih devijacija.
Prema tome, možemo reći da vrednosti $\mu \pm z^*\sigma$ na varijabli X i odgovarajuće vrednosti z na varijabli Z predstavljaju iste kvantile. U konkretnom primeru koji smo naveli vrednost 129.4 na varijabli X i vrednost +1.96 na varijabli Z predstavljaju kvantil 0.975 u „svojoj“ raspodeli.

Familija normalnih raspodela predstavlja teorijski matematički model kojima se, kako se često veruje, može adekvatno opisati „ponašanje“ mnogih varijabli s kojima se sreću istraživači u psihologiji i srodnim oblastima. To znači da istraživači u tim oblastima polaze od pretpostavke da se mnoge ljudske osobine, ne samo fizičke (npr.,visina, težina) već i psihološke (npr., inteligencija, pojedine crte ličnosti, znanje) raspodeljuju u populaciji prema normalnoj raspodeli. Najzaslužniji za „prevođenje“ normalne funkcije gustine iz modela koji dobro opisuje raspodelu grešaka u astronomskim merenjima u matematički model populacionih raspodela mnogih bioloških i psiholoških karakteristika je belgijski astronom Ketelet.
Mereći pojedine fizičke karakteristike velikog broja vojnika Ketelet uočava veliku sličnost empirijske raspodele učestalosti mera sa normalnom raspodelom. Ketelet počinje da veruje u ideju da je ideal Prirode da kreira prosečnog čoveka a da su individualne razlike među ljudima zapravo posledica grešaka Prirode koja „promašuje“ u „nastojanju“ da stvori prosečnog (po Keteletovom shvatanju idealnog) čoveka. Otuda, po Keteletu, jedan isti matematički model dobro opisuje dve vrste grešaka, kako one u astronomskim i drugim merenjima, tako i onih koje pravi „majka Priroda“ u nastojanju da stvori idealnog čoveka. Vera u sveprisutnost ove raspodele kada su u pitanju biološke i psihološke karakteristike ljudi kumovala je i imenu raspodele – „zakon grešaka“ postaje *normalna* raspodela.[26]

[25] Podrobnije o ovome može se videti u knjigama iz teorije verovatnoće (npr., Dekking et al., 2005, str.106).
[26] Podrobnije o ovome može se pročitati u Cowles, 2001.

Većina statističkih paketa ima ugrađene funkcije kojima se mogu izračunati verovatnoće da slučajna varijabla koja ima određenu raspodelu (npr., binomnu ili normalnu) uzme neku od vrednosti u određenom intervalu. U programu SPSS ove funkcije se nalaze u okviru komande COMPUTE i počinju skraćenicom Cdf (npr., za familiju binomnih raspodela je to funkcija Cdf.Binom a za familiju normalnih raspodela funkcija Cdf.Normal. U većini paketa postoje i funkcije za određivanje verovatnoće da diskretna slučajna varijabla uzme određenu vrednost ili za određivanje vrednosti funkcije gustine za kontinuiranu slučajnu varijablu. U programu SPSS nazivi ovih funkcija počinju skraćenicom Pdf. Isto tako, postoje i funkcije za određivanje kvantila koji odgovaraju određenoj vrednosti funkcije distribucije. Imena ovih funkcija u programu SPSS počinju skraćenicom Idf (od engleskog Inverse Distribution function).

Sažetak: najvažnije ideje iz teorije verovatnoće za razumevanje statistike

- Verovatnoća može biti bilo koji realni broj od 0 do 1.
- Verovatnoću nekog događaja možemo oceniti na osnovu relativne učestalosti (relativne frekvencije) tog događaja u velikom broju pokušaja ili posmatranja.
- Šanse za neki događaj predstavljaju količnik verovatnoće dešavanja i verovatnoće nedešavanja tog događaja.
- Zbir verovatnoća nekog događaja i komplementa tog događaja jednak je 1.
- Uslovna verovatnoća događaja *A*, pod uslovom da se desio događaj *B*, jednaka je količniku verovatnoće zajedničkog dešavanja ova dva događaja i verovatnoće događaja *B*.
- Uslovna verovatnoća događaja *A* pod uslovom da se desio događaj *B*, nije, u opštem slučaju, jednaka uslovnoj verovatnoći događaja *B* pod uslovom da se desio događaj *A*.
- Ako su događaji *A* i *B* statistički nezavisni onda su uslovne verovatnoće ovih događaja jednake običnim verovatnoćama tih događaja.
- Verovatnoća zajedničkog dešavanja uzajamno isključivih događaja jednaka je nuli.
- Verovatnoća dešavanja jednog ili drugog događaja, ako su događaji uzajamno isključivi, jednaka je zbiru dešavanja pojedinačnih događaja.
- Verovatnoća zajedničkog dešavanja statistički nezavisnih događaja jednaka je proizvodu verovatnoća svakog od događaja.
- Slučajna varijabla je varijabla koja na slučaj uzima neku od svojih mogućih vrednosti tako da zbir verovatnoća za sve moguće vrednosti varijable bude jednak jedinici.
- Statistici, tj. statističke mere slučajnih uzoraka predstavljaju slučajne varijable.
- Distribucija verovatnoća za diskretnu slučajnu varijablu sadrži parove mogućih vrednosti varijable i verovatnoća koje su pridružene tim vrednostima.
- Funkcija gustine za kontinuirane varijable pokazuje relativnu učestalost ("gustinu") vrednosti varijable u nekom intervalu. Celokupna površina pod funkcijom gustine jednaka je 1 (a ako se iskazuje procentualno jednaka je 100%).
- Za diskretne slučajne varijable može se odrediti verovatnoća pojedinih vrednosti varijable, a za kontinuirane slučajne varijable moguće je odrediti samo verovatnoću da varijabla uzme vrednosti u određenom intervalu.

- Funkcija distribucije za datu vrednost varijable pokazuje verovatnoću da varijabla uzme neku vrednost manju od date vrednosti ili jednaku datoj vrednosti.
- Kvantil q_p za slučajnu varijablu je najmanja vrednost varijable za koju važi sledeće: verovatnoća da varijabla uzme neku vrednost od najniže moguće vrednosti do vrednosti kvantila jednaka je p, tj. indeksu iz oznake kvantila.
- Očekivana vrednost ili aritmetička sredina slučajne varijable jednaka je zbiru proizvoda mogućih vrednosti varijable i odgovarajućih verovatnoća za te vrednosti.
- Varijansa slučajne varijable jednaka je zbiru proizvoda kvadriranih odstupanja vrednosti na varijabli od očekivane vrednosti varijable i odgovarajućih verovatnoća za te vrednosti.
- Familija normalnih distribucija definisana je parametrima μ i σ, tj. aritmetičkom sredinom i standardnom devijacijom varijable.
- Za mnoge varijable koje se koriste u psihologiji i srodnim oblastima pretpostavlja se da se normalno distribuiraju u populaciji.
- Standardizovana normalna raspodela definisana je parametrima $\mu = 0$ i $\sigma = 1$.
- Verovatnoća da slučajna varijabla koja ima standardizovanu normalnu raspodelu uzme neku od vrednosti u intervalu od -1.96 do +1.96 jednaka je 0.95 ili 95%.
- Verovatnoća da slučajna varijabla koja ima standardizovanu normalnu raspodelu uzme neku od vrednosti u intervalu od -2.58 do +2.58 jednaka je 0.99 ili 99%.

III. 4. Odnos matematičkog modela i prirode pojave na koju se on primenjuje

Između matematičkih pojmova ili modela i prirodnih pojava postoji izomorfizam (sličnost oblika). Primenjujući neki matematički model na pojave (usled dovoljnog slaganja oblika pojava i oblika matematičkih izraza) možemo doći do vrlo korisnih pretpostavki o prirodi pojave sledeći logiku njenog matematičkog modela. Naravno, pretpostavke do kojih dolazimo na ovaj način potrebno je podvrgnuti empirijskoj proveri.

III. 5. Odnos teorijskih distribucija verovatnoće i empirijskih distribucija učestalosti

Teorijske distribucije verovatnoća i funkcije gustine verovatnoće vezane su za populaciju ili proces koji je generisao podatke. Distribucija verovatnoća za diskretnu slučajnu varijablu pokazuje verovatnoću da slučajno izabrani član populacije ima određenu vrednost u pogledu varijable koja se meri.
Za kontinuiranu slučajnu varijablu funkcija gustine verovatnoće dopušta da se odredi verovatnoća da član populacije slučajno izabran ima vrednost koja će se nalaziti između određenih granica.
Kada je reč o uzorku, tabuliranje podataka da bi se pokazala učestalost sa kojima se svaka vrednost varijable pojavljuje u uzorku daje funkciju frekvencije ili distribuciju frekvencija. Distribucija frekvencija u uzorku određena je prirodom distribuije verovatnoća date varijable u populaciji i varijabilnošću procesa izbora

reprezentativnog uzorka. Distribucija frekvencija će, prema tome, odstupati slučajno manje ili više od distribucije populacije.

Literatura na koju se poziva u ovom tekstu

Arbuthnott, J. (1710–1712). An Argument for Divine Providence, Taken from the Constant Regularity Observ'd in the Births of Both Sexes. *Philosophical Transactions (1683–1775)*, *27*, 186–190. Preuzeto sa http://uk.jstor.org 11.02.2014.

Bartoszynski, R., & Niewiadomska-Bugaj, M. (2008). *Probability and statistical inference, Second Edition*. Hoboken, NJ: John Wiley & Sons, Inc.

Cowles, M. (2001). *Statistics in psychology: an historical perspective, Second Edition*. London: Lawrence Erlbaum Associates, Inc.

Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). *A modern introduction to Probability and Statistics*. London: Springer-Verlag.

Edvards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, *70*(3), 193–242.

Ellis, L., Hershberger, S., Field, E., Wersinger, S., Pellis, S., Geary, D., Palmer, C., Hoyenga, K., Hetsroni, A., & Karadi, K. (2008). *Sex differences: summarizing more than a century of scientific research*. New York: Psychology Press.

Falk, R., & Lann, A. L. (2013). Numbers defy the law of large numbers. *Teaching statistics*, *37*(2), 54–60.

Fulton, L. V., Mendez, F. A., Bastian, N. D., & Musal, R. M. (2012). Confusion Between Odds and Probability, a Pandemic? *Journal of Statistics Education*, 20(3), skinuto 20.10.2013 sa URL adrese: www.amstat.org/publications/jse/v20n3/fulton.pdf

Hogg, R., V., & Craig, A. T. (1978). *Introduction to mathematical statistics, Fourth Edition*. New York: Macmillan Publishing Co., Inc., London: *Collier Macmillan Publishers.*

Mukhopadhyay, N. (2000). *Probability and statistical inference*. Basel: Marcel Dekker, Inc.

Stigler, S. M. (1983). Who discovered Bayes's theorem? *The American Statistician*, *37*(4, P1), 290–296.

Wheeler, D. J. (2011a). Problems with Skewness and Kurtosis, Part One. Skinuto 7. novembra 2013. godine sa URL adrese http://www.qualitydigest.com/inside/qualityinsiderarticle/problemsskewnessandkurtosispartone.html