Binomna i Poissonova razdioba diskretne slučajne varijable

Tadić, Ema

SVEUČILIŠTE U RIJECI

TEHNIČKI FAKULTET

Prijediplomski sveučilišni studij elektrotehnike

Završni rad

BINOMNA I POISSONOVA RAZDIOBA DISKRETNE SLUČAJNE VARIJABLE

Rijeka, rujan 2023.

Ema Tadić
0069086652

SVEUČILIŠTE U RIJECI

TEHNIČKI FAKULTET

Prijediplomski sveučilišni studij elektrotehnike

Završni rad

BINOMNA I POISSONOVA RAZDIOBA DISKRETNE SLUČAJNE VARIJABLE

Mentor: izv. prof. dr. sc. Ivan Dražić

Komentor: doc. dr. sc. Angela Bašić-Šiško

Rijeka, rujan 2023.

Ema Tadić
0069086652

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
POVJERENSTVO ZA ZAVRŠNE ISPITE

Rijeka, 13. ožujka 2023.

Zavod: **Zavod za matematiku, fiziku i strane jezike**
Predmet: **Inženjerska matematika ET**
Grana: **1.01.07 primijenjena matematika i matematičko modeliranje**

ZADATAK ZA ZAVRŠNI RAD

Pristupnik: **Ema Tadić (0069086652)**
Studij: Sveučilišni prijediplomski studij elektrotehnike

Zadatak: **Binomna i Poissonova razdioba diskretne slučajne varijable**

Opis zadatka:

U radu je potrebno navesti temeljne pojmove teorije vjerojatnosti i precizno definirati slučajnu varijablu. Potrebno je objasniti pojam razdiobe diskretne slučajne varijable i načine njihova zadavanja, pri čemu se posebno treba osvrnuti na Binomnu i Poissonovu razdiobu. Također je potrebno navesti temeljna svojstva ove dvije razdiobe u vidu njihovih numeričkih pokazatelja.
U praktičnom dijelu rada potrebno je navedene razdiobe staviti u kontekst primjene u inženjerstvu s posebnim naglaskom na primjene u elektrotehnici. Također se trebaju objasniti temeljni principi simuliranja opisanih slučajnih varijabli.

Rad mora biti napisan prema Uputama za pisanje diplomskih / završnih radova koje su objavljene na mrežnim stranicama studija.

Zadatak uručen pristupniku: 20. ožujka 2023.

Mentor:

Izv. prof. dr. sc. Ivan Dražić

Doc. dr. sc. Angela Bašić-Šiško (komentor)

Predsjednik povjerenstva za završni ispit:

Prof. dr. sc. Dubravko Franković

IZJAVA

Sukladno članku 7. stavku 1. Pravilnika o završnom radu, završnom ispitu i završetku sveučilišnih prijediplomskih studija Tehničkog fakulteta Sveučilišta u Rijeci od 4. travnja 2023., izjavljujem da sam samostalno izradila završni rad prema zadatku preuzetom dana 20. ožujka 2023.

Rijeka, 11. rujna 2023.

Ema Tadić

Ovim putem zahvaljujem se obitelji, osobito roditeljima i bratu, te prijateljima koji su bili uz mene i davali mi podršku kroz studentske dane.

Veliko hvala profesoru i mentoru izv. prof. dr. sc. Ivanu Dražiću te doc. dr. sc. Angeli Bašić-Šiško na mentorstvu i strpljenju tijekom izrade ovog završnog rada.

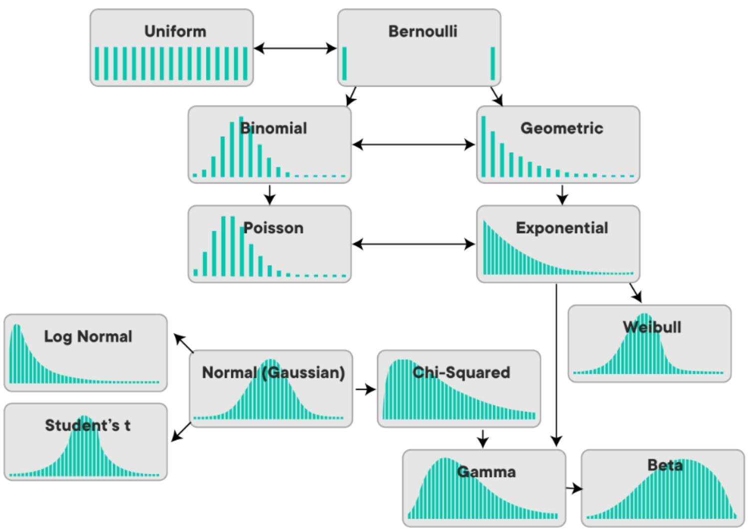
Sadržaj

1. Uvod

Slučajni događaj, u kontekstu matematike i teorije vjerojatnosti, odnosi se na događaj ili ishod koji je podložan slučaju i neizvjesnosti. Ti se događaji obično opisuju u smislu vjerojatnosti njihovog ostvarivanja. Samim time se uvodi i pojam vjerojatnosti koji odgovara na pitanje: "Kolika je vjerojatnost da će se određeni ishod ili događaj dogoditi?". Vjerojatnost različitih ishoda ili događaja opisuje se matematičkom funkcijom razdiobe koja određuje kako je ukupna vjerojatnost raspoređena među svim mogućim ishodima. Razdiobe je bitno razumjeti jer nam pomažu shvatiti i upravljati neizvjesnošću u različitim poljima i situacijama, uključujući znanost, inženjerstvo, financije i društvene znanosti.

Ovaj rad se posebno osvrće na dvije vrste razdioba: binomnu i Poissonovu. Binomna razdioba je diskretna raspodjela vjerojatnosti koja opisuje broj uspjeha u fiksnom broju neovisnih i identičnih Bernoullijevih pokušaja, gdje svaki pokušaj može rezultirati jednim od dva ishoda: uspjeh ili neuspjeh. Poissonova razdioba pak opisuje broj događaja koji se dogode unutar fiksnog vremenskog ili prostornog intervala, s obzirom na poznatu prosječnu stopu pojavljivanja. Nazvana je po francuskom matematičaru Siméonu Denisu Poissonu i posebno je korisna za modeliranje rijetkih događaja koji se događaju nasumično i neovisno.

U ovom radu ćemo prvo detaljno objasniti slučajne događaje te navesti njihova svojstva. Nadalje ćemo objasniti razne osnovne koncepte vjerojatnosti kako bi se mogao shvatiti pojam razdiobe. Uvest ćemo funkciju vjerojatnosti kao alat za predviđanje ishoda u raznim područjima te slučajnu varijablu koja omogućuje da događaje opišemo na strukturiran način. Zatim ćemo objasniti nekoliko numeričkih karakteristika razdiobe vjerojatnosti koje daju smisao složenim informacijama kao što su matematičko očekivanje, varijanca, moment, koeficijent asimetrije i koeficijent spljošte-

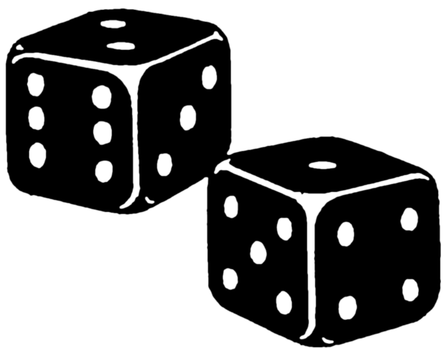


Slika 1.1. Prikaz nekih razdioba, Izvor: [2]

nosti. Nakon što smo objasnili sve osnovne koncepte vjerojatnosti posvetit ćemo se razdiobama. Navest ćemo neke njihove vrste od kojih ćemo se više osvrnuti na binomnu i Poissonovu. Svaku od njih ćemo objasniti, navesti njihove ključne karakteristike te objasniti njihovu primjenu u stvarnom životu tj. na koji način nam pomažu bolje shvatiti razne svakodnevne događaje.

2. Slučajni događaji

Ishod koji se pojavi unutar nekog eksperimenta naziva se događaj. Događaj je temeljni koncept u teoriji vjerojatnosti te se koristi za opisivanje različitih mogućih rezultata slučajne pojave. Događaji se u teoriji vjerojatnosti označavaju velikim slovima abecede. Ako se u svakom rezultatu pokusa pojavi događaj U, tada U nazivamo sigurnim događajem. Na primjer, u pokusu bacanja igraće kockice događaj "pao je broj između 1 i 6" je siguran događaj. Isto vrijedi i za bacanje novčića jer je događaj "palo je pismo ili glava" siguran događaj. Suprotno tome postoji i nemoguć događaj V. Na primjer, nemoguće je da se bacanjem kockice dobije broj 7 jer kao što se ustanovilo u prethodnom primjeru skala je od 1 do 6. Nemogući događaji su važni u teoriji vjerojatnosti jer pomažu u definiranju raspona mogućih ishoda i daju kontrast događajima koji su mogući. Oni usporedbom doprinose razumijevanju vjerojatnosti. Elementaran događaj je onaj koji se realizira na samo jedan način te se ne može raščlaniti na manje događaje, odnosno B predstavlja najosnovniju jedinicu promatranja u okviru eksperimenta. Elementaran događaj se obično označava s ω, a skup svih elementarnih događaja s Ω. Za bolje shvaćanje može se navesti da su pismo i glava u pokusu bacanja novčića elementarni događaji.



Slika 2.1. Igraće kockice, Izvor: [3]

Unija se odnosi na događaj kada se ostvari bar jedan od navedenih događaja. Ako postoje događaji A i B njihova unija se definira kao skup ishoda koji pripadaju ili događaju A ili B ili oboje:

$$A \cup B = \{\omega : \omega \in A \,\text{ili}\, \omega \in B\}. \tag{2.1}$$

Unija omogućuje analizu situacija u kojima smo zainteresirani za pojavu barem jednog od niza događaja. Ona je ključni alat za izračunavanje vjerojatnosti u složenim scenarijima koji uključuju više događaja.

Presjek događaja se odnosi na ostvarenje svih navedenih događaja. Ako postoje događaji A i B presjek se tada definira kao novi događaj koji uključuje sve ishode koji pripadaju svim izvornim događajima:

$$A \cap B = \{\omega : \omega \in A \text{ i } \omega \in B\}. \tag{2.2}$$

Poddogađaj je događaj koji je podskup nekog drugog događaja. Drugim riječima, ako je A događaj i B je podskup od A, onda je B poddogađaj od A $(B \subset A)$. Na primjer, neka je u pokusu bacanja kockice događaj A "pao je paran broj", a događaj B "pao je broj 2 ili broj 4". Tada je B poddogađaj događaja A jer su svi ishodi poddogađaja B također i ishodi od A. Poddogađaji su bitni u kontekstu uvjetne vjerojatnosti jer ona uključuje izračun vjerojatnosti da će se neki događaj dogoditi s obzirom da se neki drugi već dogodio. Ekvivalentni događaji su oni koji se sastoje od istih ishoda, odnosno, kažemo da su događaji A i B ekvivalentni ako vrijedi $A \subset B$ i $B \subset A$ (u tom slučaju je $A = B$). Na primjer, u pokusu bacanja kockice događaji $A =$ "pao je paran broj veći od broja 1" i $B =$ "pao je neparan broj" su ekvivalentni jer imaju istu vjerojatnost da će se dogoditi.

Disjunktni događaji, također poznati kao "međusobno isključivi događaji", su oni koji se ne mogu dogoditi u isto vrijeme. Ako se dogodi jedan od događaja, onda se drugi događaj ne može dogoditi. Događaji A i B su disjunktni ako im je presjek prazan skup: $A \cap B = \emptyset$. U tom slučaju ne postoji ishod koji pripada i događaju A i događaju B. Ako u pokusu bacanja kockice događaj A znači dobivanje parnog broja, a događaj B dobivanje neparnog broja tada su A i B disjunktni jer ne postoji ishod u kojemu će dobiveni broj biti istovremeno paran i neparan. Suprotan događaji ili "negacija" je događaj koji se sastoji od svih ishoda koji nisu uključeni u događaj A. Predstavlja suprotnost navedenom događaju. Ako je A "dobivanje parnog broja" na kockici onda je $\bar{A}$ "dobivanje neparnog broja". Dodatni primjer je kada bi događaj B bio "dobivanje crvene karte" u pokusu izvlačenja jedne karte iz špila igraćih karata, $\bar{B}$ bi bilo "dobivanje karte koja nije crvena". Relacija događaja i njemu suprotnog događaja glasi:

$$A + \bar{A} = \Omega. \tag{2.3}$$

Ova jednadžba pojašnjava da je vjerojatnost da se događaj A dogodi i događaj $\bar{A}$ ne dogodi jednaka 1 što također znači da se ili A ili $\bar{A}$ mora dogoditi. Potpuni sistem odnosi se na skup događaja koji pokrivaju sve moguće ishode, međusobno su disjunktni te u uniji daju siguran događaj.

2.1. Svojstva događaja $\cup$ i $\cap$ na skupu događaja

Neka su A i B događaji. Operacija unije ima sljedeća svojstva:

1. Uključenost

   Ako su A i B događaji, tada je $A \cup B$ događaj koji sadrži sve elemente koji pripadaju A ili B ili oba:

$$A \subset A \cup B \text{ i } B \subset A \cup B. \tag{2.4}$$

2. Komutativnost

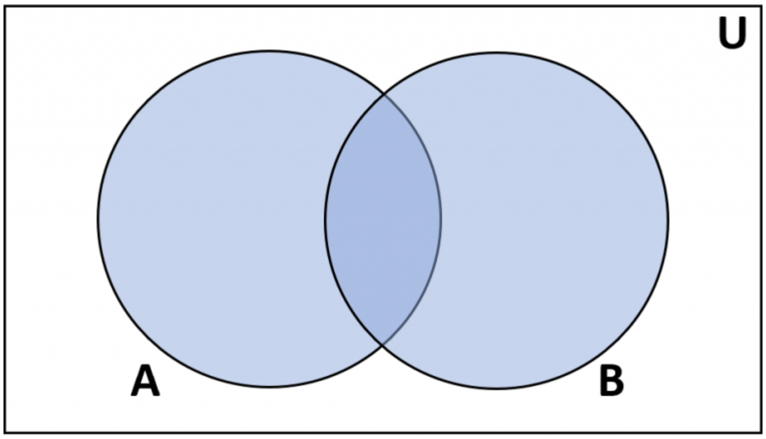
$$A \cup B = B \cup A. \tag{2.5}$$

3. Asocijativnost

$$(A \cup B) \cup C = A \cup (B \cup C). \tag{2.6}$$

4. Idempotentnost Unija događaja sa samim sobom daje isti taj događaj.

$$A \cup A = A. \tag{2.7}$$



Slika 2.2. Unija događaja A i B, Izvor: [4]

Operacija presjeka ima sljedeća svojstva:

1. Uključenost

   Ako su A i B događaji, tada je $A \cap B$ događaj koji sadrži samo one elemente koji pripadaju i A i B:

$$A \cap B \subset A \text{ i } A \cap B \subset B. \tag{2.8}$$

2. Komutativnost

$$A \cap B = B \cap A. \tag{2.9}$$

3. Asocijativnost

$$(A \cap B) \cap C = A \cap (B \cap C). \tag{2.10}$$

4. Idempotentnost

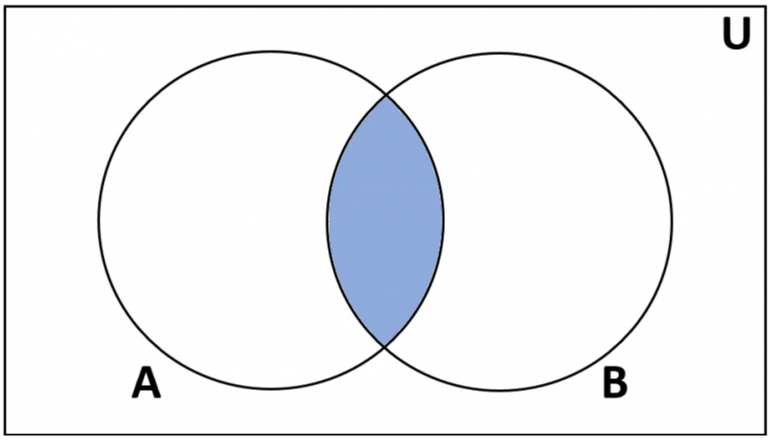
   Događaj presječen sa samim sobom jednak je samom događaju.

$$A \cap A = A. \tag{2.11}$$

5. Distributivnost

   Operacija presjeka distribuira se preko operacije unije:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C). \tag{2.12}$$



Slika 2.3. Presjek događaja A i B, Izvor: [4]

2.2. σ-algebra događaja

Za shvaćanje različitih događaja i ishoda potreban je strukturiran način za postići isto. Tu se javlja ideja o σ-algebri. σ-algebra je poput alata koji pomaže organizirati i analizirati događaje unutar zadanog skupa. To je skup podskupova kojima se može manipulirati na određene načine, a da se pritom zadrže određena važna svojstva. Upravo ta svojstva osiguravaju rad s vjerojatnostima na dosljedan i strog način. Neka je U skup objekata, a S neprazan skup podskupova skupa U koji ispunjava sljedeća tri uvjeta [6]:

1. Sadržava puni skup i prazan skup

   Skup S i prazan skup $\emptyset$ su oboje članovi σ-algebre:

$$S \in U \text{ i } \emptyset \in U. \tag{2.13}$$

2. Komplementi su u σ-algebri

   Ako je A u sigma-polju, onda je i $\bar{A}$ također u sigma-polju:

$$A \in U \Rightarrow \bar{A} \in U. \tag{2.14}$$

3. Prebrojiv niz skupova je u σ-algebri

   Ako je $A_1, A_2, \ldots$ prebrojiv niz skupova σ-algebre onda je i njihova unija također u σ-algebri:

$$A_i \in U \text{ za } i = 1, 2, \cdots \Rightarrow \bigcup_{i=1}^{\infty} A_1 \in U. \tag{2.15}$$

Slijedeći ova pravila, σ-algebra omogućuje da se definiraju događaji za koje se mogu dodijeliti vjerojatnosti te omogućuje računanje vjerojatnosti različitih događaja. Bilo da se baca novčić, kockica ili predviđaju vremenski uvjeti, σ-algebra pomaže u kretanju svijetom neizvjesnosti i vjerojatnosti s dobro definiranom strukturom.

2.2.1. σ-algebra događaja u svakodenvnom životu

Zadatak meteorologa je predvidjeti hoće li sutra padati kiša ili ne. Vremenske prilike su prilično teško predvidive, pa je potrebno imati sustavan način rješavanja svih mogućnosti i neizvjesnosti. Ovdje se javlja ideja ideja o σ-algebri. U svijetu meteorologije, "prostor uzoraka" bi bili različiti vremenski uvjeti kao što su "kišovito", "sunčano", "oblačno" i tako dalje. Svaki od ovih uvjeta može se smatrati "događajem". Osim u meteorologiji može se vidjeti njegova primjena i u raznim drugim područjima:

1. Eksperiment bacanja novčića

   Prostor uzoraka sastoji se od dva ishoda: glava (G) i pismo (P). σ-algebra može uključivati skupove $\{G\}$, $\{P\}$ i $\{G, P\}$ koji predstavljaju pojedinačne ishode odnosno cijeli prostor uzoraka.

Slika 2.4. Eksperiment bacanja novčića, Izvor: [5]

2. Eksperiment bacanja kockice

   Za eksperiment bacanja kockice σ-algebra može uključivati podskupove koji odgovaraju određenim ishodima (na primjer: 1, 2, zatim kombinacije 1,2) i cijeli skup ishoda.

3. Kontrola kvalitete u prizvodnji

   Kada se prati kvaliteta proizvoda u procesu proizvodnje, σ-algebra se može koristiti za predstavljanje različitih razina neispravnosti. Događaji u σ-algebri mogu odgovarati prihvatljivim i neprihvatljivim razinama kvalitete.

4. Sportska analitika

   U sportskoj analitici može se koristiti σ-algebra za predstavljanje različitih situacija u igri ili performansu igrača. To može uključiti definiranje događaja povezanih s bodovanjem, statistikom igrača i ishodom igre.

5. Medicinska dijagnostika

   U medicinskoj dijagnostici može se definirati σ-algebra koja predstavlja prisutnost ili odsutnost određenih simptoma ili stanja čime pomaže u izračunavanju vjerojatnosti povezanih s dijagnozama i ishodima liječenja.

Ovi primjeri ilustriraju kako se σ-algebra može primijeniti na situacije iz stvarnog života za modeliranje neizvjesnosti, analizu vjerojatnosti i donošenje informiranih odluka na temelju različitih događaja ili ishoda.

3. Vjerojatnost

Vjerojatnost je temeljni koncept u matematici i statistici koji analizira šansu da se događaj dogodi. To je mjera nesigurnosti i koristi se za opisivanje i predviđanje ishoda u raznim područjima, uključujući matematiku, znanost, ekonomiju, inženjerstvo i društvene znanosti.

Vjerojatnost je funkcija koja dodjeljuje nenegativan realni broj svakom događaju u prostoru uzorka. Ti se brojevi nazivaju vjerojatnostima i u većini slučajeva se označavaju s $P(A)$, gdje je A događaj. Da bi funkcija bila vjerojatnost mora zadovoljavati sljedeće uvjete:

- Vjerojatnosti se nalaze u intervalu između 0 i 1: $0 \leq P(A) \leq 1$ za bilo koji događaj A.
- Vjerojatnost cijelog prostora uzorka je 1: $P(\Omega) = 1$.
- σ-aditivnost: Ako su događaji $A_1, A_2, A_3, \dots$ međusobno disjunktni, tada vrijedi

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \tag{3.1}$$

Vjerojatnost ima i neka dodatna svojstva koja se lako izvode iz definicijskih uvjeta, a koja olakšavaju račun s vjerojatnosti:

1. Konačna aditivnost

   Za događaje koji se međusobno isključuju (događaji koji se ne mogu dogoditi istovremeno), vjerojatnost njihove unije je zbroj njihovih vjerojatnosti: $P(A \cup B) = P(A) + P(B)$ ako su A i B međusobno isključivi tj. $A \cap B = \emptyset$.

2. Komplementarna vjerojatnost

   Vjerojatnost komplementa događaja A je $P(\bar{A}) = 1 - P(A)$.

Dva važna pojma vezana za vjerojatnost su uvjetna vjerojatnost te nezavisnost događaja:

1. Uvjetna vjerojatnost

   Uvjetna vjerojatnost je vjerojatnost da će se jedan događaj dogoditi s obzirom da se dogodio drugi događaj [7]. Označava se kao $P(A|B)$, što je vjerojatnost od A uz uvjet da se dogodio događaj B te se računa prema izrazu

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{3.2}$$

2. Neovisni događaji

   Dva događaja, A i B su neovisni ako pojavljivanje jednog (ili nepojavljivanje) ne utječe na vjerojatnost drugog. Spomenuti uvjet se matematički izražava kao:

$$P(A \cap B) = P(A) \cdot P(B). \tag{3.3}$$

3.1. Zakon totalne vjerojatnosti

Zakon totalne vjerojatnosti je koncept u teoriji vjerojatnosti koji omogućuje računanje vjerojatnosti događaja B razmatranjem svih mogućih načina da se B može dogoditi kroz različite međusobno isključive događaje $A_1, A_2, \ldots, A_n$. Zakon totalne vjerojatnosti glasi:

$$P(B) = \sum_i P(B|A_i) \cdot P(A_i), \tag{3.4}$$

za koji vrijedi:

- $A_1, A_2, \ldots, A_n$ su međusobno isključivi događaji.
- $P(A_i) > 0$ za svaki i, što znači da svaki događaj ima vjerojatnost veću od nule da će se dogoditi.
- $P(B|A_i)$ je uvjetna vjerojatnost da će se dogoditi događaj B ako se događaj A_i već dogodio.

Ovaj se zakon često koristi u situacijama kada je lakše izračunati $P(B|A_i)$ za svaki događaj te gdje ti događaji pokrivaju sve moguće slučajeve. Koristi se za rastavljanje složenih izračuna vjerojatnosti u jednostavnije izračune uvjetne vjerojatnosti. Bitna primjena zakona potpune vjerojatnosti je Bayesovo zaključivanje, gdje se ažuriraju uvjerenja o događaju na temelju novih informacija.

Primjer 3.1 (Vjerojatnost padanja kiše)**.** *Želimo izračunati vjerojatnost kiše sutra* (K)*, a poznato je da vjerojatnost kiše ovisi o tome je li radni dan* (R) *ili vikend* (V)*. Također je poznata vjerojatnost svakog scenarija:*

- *Radnim danom vjerojatnost kiše je* $P(K|R) = 0.3$.
- *Vikendom vjerojatnost kiše je* $P(K|V) = 0.6$.

Također su poznate vjerojatnosti da li je sutra radni dan ili vikend:

- *Vjerojatnost da je sutra radni dan je* $P(R) = 0.7$.
- *Vjerojatnost da je sutra vikend je* $P(V) = 0.3$.

Sada se može koristiti zakon totalne vjerojatnosti kako bi se odredila ukupna vjerojantost kiše sutra:

$$\begin{aligned} P(K) &= P(K|R) \cdot P(R) + P(K|V) \cdot P(V) \\ &= 0.3 \cdot 0.7 + 0.6 \cdot 0.3 = 0.21 + 0.18 = 0.39. \end{aligned} \tag{3.5}$$

Dakle, vjerojatnost kiše sutra je 0.39, uzimajući u obzir i scenarij radnog dana i vikenda. Ovaj primjer pokazuje kako se zakon totalne vjerojatnosti može primjeniti za izračun ukupne vjerojatnosti događaja uzimajući u obzir sve moguće scenarije i podjele.

4. Slučajna varijabla

Slučajna varijabla je funkcija koja pridružuje realni broj svakom ishodu slučajnog eksperimenta. To je ključni koncept u teoriji vjerojatnosti i statistici, koji omogućuje da neizvjesne događaje predstavimo na strukturiran i kvantificiran način.

Neka je X slučajna varijabla povezana sa slučajnim eksperimentom. Za svaki ishod eksperimenta ω, $X(\omega)$ je realni broj koji odgovara vrijednosti slučajne varijable za taj određeni ishod.

Slučajne varijable se klasificiraju u dvije glavne vrste:

- diskretna slučajna varijabla,
- neprekidna slučajna varijabla.

4.1. Diskretna slučajna varijabla

Diskretna slučajna varijabla je vrsta slučajne varijable u teoriji vjerojatnosti i statistici koja može poprimiti prebrojivo mnogo različitih vrijednosti. Predstavlja ishode eksperimenata gdje se rezultat može nabrojati ili navesti. Slučajna varijabla X je diskretna ako postoji prebrojiv skup $R_X = \{X_1, X_2, \ldots, X_n, \ldots\}$ za koji vrijedi:

$$P(X \in R_X) = 1 \text{ i } P(X \in R_X^C) = 0. \tag{4.1}$$

Varijabla X je definirana za niz vrijednosti $x_1, x_2, \ldots, x_n$ i niz pripadnh vjerojatnosti $p_1, p_2, \ldots,$ p_n pri čemu je $P(X = x_i) = p_i$ uz zadovoljen uvjet $\sum_{i=1}^{n} p_i = 1$. Zakon po kojem je ukupna vjerojatnost raspoređena na vrijednosti x_i zove se zakon vjerojatnosti diskretne slučajne varijable. Neke od najpoznatijih diskretnih slučajnih varijabli su Bernoullijeva, binomna, Poissonova i geometrijska slučajna varijabla. Više o nekima od njih će biti rečeno u idućim poglavljima.

4.2. Neprekidna slučajna varijabla

Neprekidna slučajna varijabla predstavlja numeričku veličinu koja može poprimiti beskonačno neprebrojivo mnogo mogućih vrijednosti unutar određenog raspona ili intervala [8]. Za razliku od diskretnih slučajnih varijabli, koje mogu poprimiti samo određene različite vrijednosti, neprekidne slučajne varijable mogu poprimiti bilo koju vrijednost unutar neprekidnog raspona.

Ključne karakteristike neprekidne slučajne varijable su:

1. Beskonačno moguće vrijednosti

Neprekidna slučajna varijabla može poprimiti beskonačan broj vrijednosti unutar zadanog intervala. Taj interval može biti konačan ili beskonačan.

2. Neprebrojiv skup vrijednosti

   Skup mogućih vrijednosti za neprekidnu slučajnu varijablu je neprebrojiv. To znači da postoji više vrijednosti nego što se može navesti u nizu.

3. Površina ispod krivulje

   Površina ispod krivulje u određenom intervalu predstavlja vjerojatnost da se neprekidna slučajna varijabla nalazi unutar tog intervala. Drugim riječima, vjerojatnosti su predstavljene područjima, a ne određenim vrijednostima.

4. Funkcija gustoće vjerojatnosti

   Neprekidnu slučajnu varijablu karakterizira funkcija gustoće vjerojatnosti. Ona opisuje relativnu vjerojatnost da varijabla bude unutar određenog raspona vrijednosti

Neke od najpoznatijih neprekidnih slučajnih varijabli su normalna, eksponencijalna, uniformna, ali i mnoge druge.

4.2.1. Funkcija razdiobe i funkcija gustoće vjerojatnosti

Neprekidna slučajna varijabla određena je svojom funkcijom razdiobe F:

$$F(x) = P(X \leq x), \tag{4.2}$$

dok je pripadna funkcija gustoće slučajne varijable dana je s:

$$f(x) = \frac{dF(x)}{dx}. \tag{4.3}$$

Svojstva funkcije gustoće su:

1. Ne-negativnost

   Funkcija vjerojatnosti je uvijek ne-negativna za sve vrijednosti slučajne varijable:

$$f(x) \geq 0, \forall x. \tag{4.4}$$

2. Ukupna površina ispod krivulje

   Ukupna površina ispod krivulje funkcije gustoće u cijelom rasponu mogućih vrijednosti jednaka je 1:

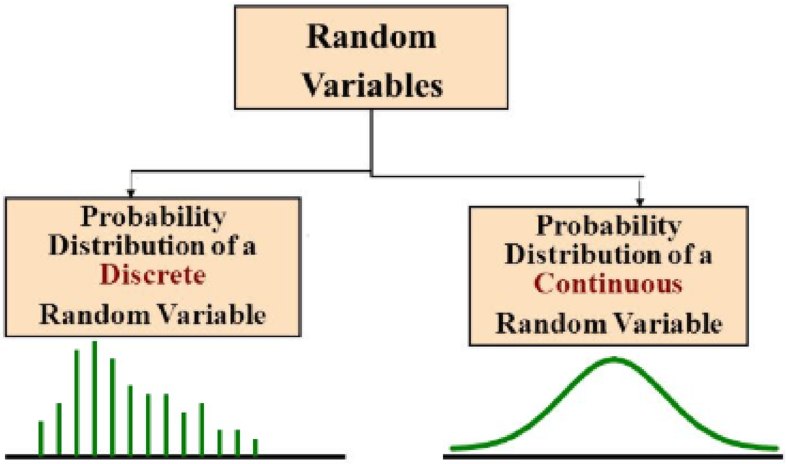
$$\int_{-\infty}^{\infty} f(x)\,dx = 1. \tag{4.5}$$

   Ovo svojstvo osigurava da je zbroj vjerojatnosti svih mogućih ishoda jednak 1.

3. Vjerojatnosti kao područja

   Vjerojatnost da se slučajna varijabla nalazi u određenom intervalu $[x_1, x_2]$ zadana je integralom funkcije gustoće preko tog intervala:

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)\,dx,\ \forall x_1 < x_2. \tag{4.6}$$



Slika 4.1. Prikaz razdioba diskretne i neprekidne slučajne varijable, Izvor: [9]

4.3. Nezavisne slučajne varijable

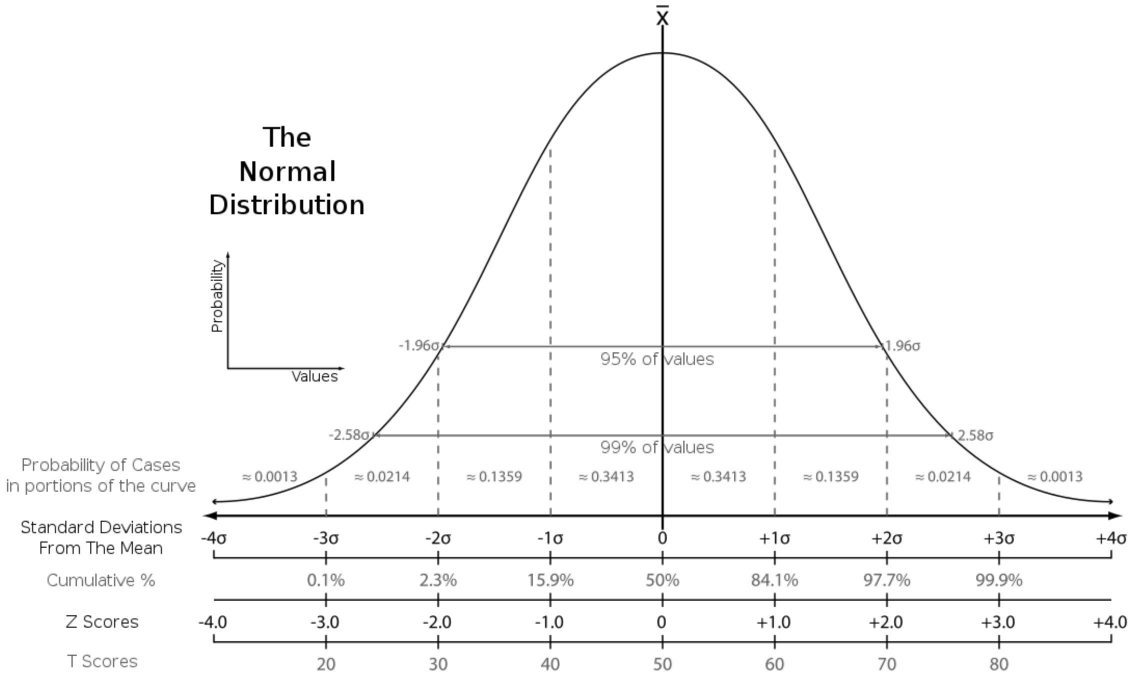
Kada se dvije ili više slučajnih varijabli smatraju nezavisnima, to znači da pojavljivanje ili vrijednost jedne varijable ne utječe niti pruža informacije o pojavljivanju ili vrijednosti drugih. Drugim riječima, na ponašanje jedne slučajne varijable ne utječe ponašanje druge. Zajednička razdioba vjerojatnosti nezavisnih slučajnih varijabli jednaka je umnošku njihovih pojedinačnih razdioba vjerojatnosti. Matematički, za dvije nezavisne slučajne varijable X i Y vrijedi:

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y). \tag{4.7}$$

Nezavisne slučajne varijable omogućuju pojednostavljenje modeliranja složenih sustava. Mnoge statističke metode i razdiobe vjerojatnosti pretpostavljaju neovisnost među varijablama, što pojednostavljuje izračune i analize. Povrede pretpostavki o neovisnosti mogu dovesti do pristranih i netočnih rezultata u statističkim analizama.

4.4. Funkcija slučajne varijable

Funkcija slučajne varijable je nova varijabla koja je izvedena primjenom matematičke funkcije na vrijednosti izvorne slučajne varijable. Ako je zadana slučajna varijabla X, funkcija od X,



Slika 4.2. Funkcija gustoće za jediničnu normalnu razdiobu, Izvor: [10]

označena kao $Y = g(X)$, formira se primjenom funkcije g na svaku moguću vrijednost od X. Rezultirajuće vrijednosti od Y će ovisiti o vrijednostima X, čineći Y slučajnom varijablom.

Neki načini razumijevanja funkcije slučajne varijable:

1. Transformacija

   Funkcija g transformira vrijednosti izvorne slučajne varijable X u vrijednosti nove slučajne varijable Y.

2. Transformacija funkcije gustoće vjerojatnosti

   Ako je poznata funkcija gustoće vjerojatnosti od X, često se može izračunati funkcija gustoće vjerojatnosti od $Y = g(X)$ koristeći se metodom transformacije slučajnih varijabli. To uključuje izračun funkcije gustoće vjerojatnosti od Y na temelju funkcije g i funkcije gustoće vjerojatnosti od X.

3. Transformacija numeričkih karakteristika razdiobe

   Numeričke transformacije nove slučajne varijable poput očekivanja i varijance mogu se izračunati pomoću svojstava funkcije g i svojstava originalne slučajne varijable X.

4. Važne razdiobe

   Mnoge uobičajene razdiobe povezane su jedna s drugom kroz funkcije slučajnih varijabli. Na primjer, ako X slijedi normalnu razdiobu, onda $Y = aX + b$ (gdje su a i b konstante) također slijedi normalnu razdiobu.

Recimo da X predstavlja visinu jedinki u centimetrima i Y je definiran kao $Y = 0.0328 \cdot X$ za pretvorbu visine u stope. U ovom slučaju, Y je funkcija od X te ako je poznata funkcija gustoće vjerojatnosti od X može se pronaći i funkcija gustoće vjerojatnosti od Y metodom transformacije. Funkcije slučajnih varijabli ključne su za modeliranje složenih odnosa u statistici, stvaranje predviđanja na temelju transformiranih podataka i razumijevanje kako se slučajne varijable ponašaju pod različitim transformacijama.

5. Numeričke karakteristike razdiobe vjerojatnosti

Numeričke karakteristike koriste se za dobivanje uvida iz podataka. Pomažu sažeti, analizirati i razumjeti skupove podataka, dajući smisao često složenim i raznolikim informacijama koje sadrže. Ove karakteristike daju numerički opis razdioba podataka, središnjih tendencija, varijabilnosti i odnosa između varijabli.

5.1. Matematičko očekivanje

Matematičko očekivanje predstavlja dugoročni prosjek ili srednju vrijednost slučajne varijable. Očekivana vrijednost daje način da se sažme središnja tendencija razdiobe slučajne varijable. Za diskretnu slučajnu varijablu X s funkcijom vjerojatnosti $P(X = x_i)$ i odgovarajuće vrijednosti x_i, očekivana vrijednost $E(X)$ računa se na sljedeći način:

$$E(X) = \sum_i x_i \cdot P(X = x_i). \tag{5.1}$$

Za neprekidnu slučajnu varijablu X s funkcijom gustoće vjerojatnosti $f(x)$ očekivana vrijednost se računa integriranjem:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx. \tag{5.2}$$

Ključne stavke o očekivanoj vrijednosti:

1. Tumačenje

   Očekivana vrijednost predstavlja "prosječnu" vrijednost koja se očekuje dobiti ako se opetovano uzima uzorak iz slučajne varijable X veliki broj puta.

2. Linearnost očekivanja

   Očekivana vrijednost je linearni operator, što znači da za konstante a i b i slučajne varijable X i Y vrijedi:

$$E(aX + bY) = aE(X) + bE(Y). \tag{5.3}$$

3. Uvjetno očekivanje

   Uvjetno očekivanje $E(X|Y)$ je proširenje koncepta, gdje se očekivana vrijednost od X računa prema određenim informacijama ili uvjetima koje daje druga slučajna varijabla Y.

4. Zakon velikih brojeva

   Zakon velikih brojeva navodi da se, kako se veličina uzorka povećava, srednja vrijednost uzorka približava očekivanoj vrijednosti populacije [11].

Primjer 5.1 (Igra na sreću). *Pretpostavimo da se igra na sreću odvija pod sljedećim uvjetima:*

- *Šansa da se osvoji jackpot koji iznosi \$1000000 je 1 prema 1000000.*
- *Kupljeni listić iznosi \$11.*

Za izračun očekivane vrijednosti sudjelovanja koristi se formula za očekivanu vrijednost. U ovom slučaju vrijedi:

- *Dobivanje jackpota* $(X = \$1,000,000)$ *ima vjerojatnost* $P(X = \$1,000,000) = \frac{1}{1{,}000{,}000}$.
- *Ne dobivanje jackpota* $(X = -\$11)$, *što znači gubitak u visini troška listića, ima vjerojatnost* $P(X = -\$11) = 1 - P(X = \$1,000,000) = 1 - \frac{1}{1{,}000{,}000}$.

Nadalje se s poznatim vrijednostima može izračunati očekivana vrijednost:

$$\begin{aligned} E(X) &= \$1,000,000 \cdot \frac{1}{1,000,000} + (-\$11) \cdot (1 - \frac{1}{1,000,000}) \\ &= \$1 - \$11 + \$11(1 - \frac{1}{1,000,000}) \\ &= -\$10 + \$11(\frac{999,999}{1,000,000}) \\ &= -\$10 + \$10.999989 \\ &= \$0.999989. \end{aligned}$$

Dakle, očekivana vrijednost iznosi \$0.999989 što znači da se, tijekom velikog broja proba, očekuje gubitak od \$0.999989 po listiću.

Očekivana vrijednost pomaže da se predvide i procijene rizici te donesu odluke na temelju vjerojatnosnih informacija. Omogućuje način da na koji se može sažeti središnja tendencija slučajnih varijabli.

5.2. Varijanca i standardna devijacija

Varijanca je statistička mjera koja kvantificira širenje ili disperziju skupa podatkovnih točaka. Omogućuje uvid u to koliko pojedinačne podatkovne točke odstupaju od srednje vrijednosti (prosjeka, očekivanja) podataka [12]. Drugim riječima, pokazuje koliko su vrijednosti podataka "rasprostranjene" od središta razdiobe. Varijancu slučajne varijable X, $\text{Var}(X)$, definiramo kao kvadrat odstupanja od srednje vrijednosti slučajne varijable te kao kvadrat standardne devijacije σ:

$$\text{Var}(X) = \sigma^2 = E[(X - \mu)^2]. \tag{5.4}$$

U smislu razdiobe vjerojatnosti, varijanca pokazuje koliko vrijednosti slučajne varijable u toj razdiobi variraju oko srednje (očekivane vrijednosti) razdiobe.

Ključne stavke o varijanci:

1. Kvadrat razlike

   Varijanca se računa uzimanjem prosjeka kvadrata razlike između svake podatkovne točke i srednje vrijednosti. Kvadriranje ovih razlika osigurava da se negativna i pozitivna odstupanja od srednje vrijednosti međusobno ne poništavaju.

2. Jedinice

   Jedinica varijance je kvadrat izvorne jedinice podataka. Kako bi se moglo interpretirati, često se koristi standardna devijacija (kvadratni korijen varijance), jer ima istu jedinicu kao izvorni podatci:

   $$\sigma = \sqrt{V(X)}. \tag{5.5}$$

   Standardna devijacija slučajne varijable X je manja što se vrijednosti x_i više grupiraju oko matematičkog očekivanja μ.

3. Mjera disperzije

   Veća varijanca ukazuje na veće širenje ili disperziju podatkovnih točaka od srednje vrijednosti, dok manja varijanca ukazuje da su podatkovne točke bliže srednjoj vrijednosti.

Varijanca je koncept koji se koristi u statistici, analizi podataka i kontroli kvalitete za kvantificiranje širenja te za procjenu varijabilnosti i dosljednosti podataka. Često se označava sa σ^2 za varijancu populacije i sa s^2 za varijancu uzorka.

5.3. Momenti

U statistici, momenti su matematičke veličine koje daju informacije o obliku, središnjoj tendenciji i širenju razdiobe vjerojatnosti ili skupa podataka. Momenti se koriste za opisivanje karakteristika razdiobe i obično se izračunavaju iz podataka ili funkcije gustoće vjerojatnosti (ili mase). Postoji nekoliko momenata koji se obično koriste u statistici, a prva dva su najosnovnija. Za diskretnu slučajnu varijablu X moment se računa na sljedeći način:

$$\mu_k = E[(X-\mu)^k] = \sum_i (x_i - \mu)^k p(x). \tag{5.6}$$

Za neprekidnu slučajnu varijablu koristi se integral:

$$\int_{-\infty}^{\infty} (x-\mu)^k f(x) dx. \tag{5.7}$$

1. Nulti moment ($k = 0$)

$$\mu_0 = E[(X-\mu)^0] = \sum_i (x_i - \mu)^0 p(x_i) = 1. \tag{5.8}$$

Nulti moment je najjednostavniji moment koji odgovara ukupnoj težini razdiobe ili skupa podataka. U praksi, nulti moment govori koliko je podatkovnih točaka ili opažanja u skupu podataka ili koliku "masu" ili "težinu" nosi cijela razdioba.

2. Prvi moment ($k = 1$)

$$\begin{aligned}\mu_1 = E[(X-\mu)^1] &= \sum_i (x_i - \mu)^1 p(x_i) \\ &= \sum_i x_i p(x_i) = \mu \sum_i p(x_i) = \mu - \mu \\ &= 0.\end{aligned} \tag{5.9}$$

Prvi moment je ubiti srednja vrijednost skupa podataka. Predstavlja središnju ili prosječnu vrijednost razdiobe.

3. Drugi moment ($k = 2$)

$$\sigma^2 = E[(X-\mu)^2] = \text{Var}(X) \tag{5.10}$$

Drugi moment je varijanca skupa podataka. Kvantificira disperziju podatkovnih točaka oko srednje vrijednosti.

Momenti omogućuju sažimanje i razumijevanje karakteristika podataka i razdioba vjerojatnosti. Nulti i prvi moment su konstantni za sve razdiobe te ne pružaju nikakve informacije o svojstvima razdiobe. Treći i četvrti moment će se detaljnije objasniti u nastavku, a momenti viši od četvrtog pružaju dodatne informacije o razdiobi, ali se rjeđe koriste u praksi.

5.4. Koeficijent asimetrije

Treći moment (koeficijent asimetrije) obično se odnosi na moment koji mjeri asimetriju razdiobe vjerojatnosti ili skupa podataka. Kvantificira koliko je razdioba iskrivljena u odnosu na središnju vrijednost (srednja vrijednost ili medijan). Koeficijent asimetrije je dan izrazom:

$$\text{Koeficijent asimetrije} = \frac{E[(X-\mu)^3]}{\sigma^3}. \tag{5.11}$$

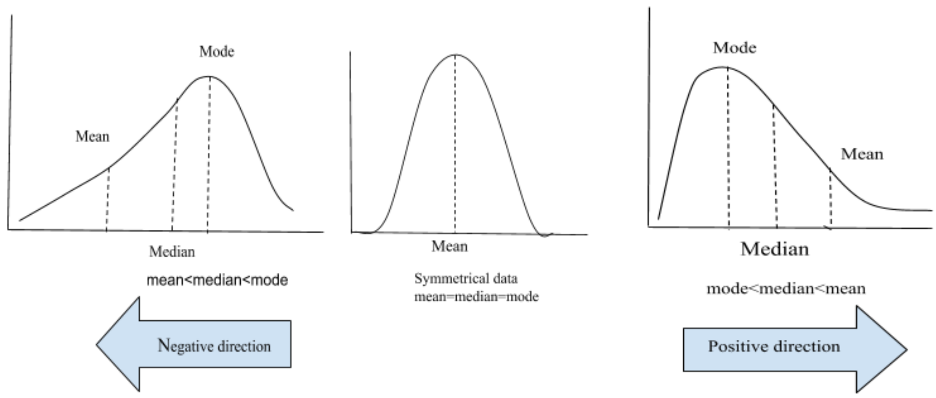
Koeficijent asimetrije govori o smjeru i stupnju asimetrije u podatcima:

- Ako je asimetrija pozitivna, to znači da je rep razdiobe duži na desnoj strani (desno nakošen ili pozitivno nakošen) [13]. Ovo sugerira da postoji nekoliko ekstremnih vrijednosti većih od srednje vrijednosti, koje povlače srednju vrijednost udesno.

- Ako je asimetrija negativna, to znači da je rep razdiobe duži na lijevoj strani (lijevo nakošen ili negativno nakošen). Ovo sugerira da postoji nekoliko ekstremnih vrijednosti manjih od srednje vrijednosti, koje povlače srednju vrijednost ulijevo.

- Ako je asimetrija blizu nule, to sugerira da je razdioba relativno simetrična, to jest da nema značajne asimetrije.

U praktičnom smislu, asimetrija nam pomaže razumjeti oblik razdiobe. Na primjer, u financijama, asimetrija povrata dionica pruža uvid u rizik i potencijal za ekstremne događaje na tržištu. U kontroli kvalitete, asimetrija pokazuje da li proizvodni proces proizvodi dosljedno ili postoje povremeni nedostatci koji uzrokuju varijacije. Asimetrija je statistički alat za razumijevanje razdiobe podataka u stvarnom svijetu.



Slika 5.1. Prikaz pozitivno asimetrične, simetrične i negativno asimetrične razdiobe, Izvor: [14]

5.5. Koeficijent spljoštenosti

Četvrti moment (koeficijent spljoštenosti ili kurtozis) mjeri vrhove razdiobe vjerojatnosti ili skupa podataka te pokazuje ima li razdioba teže repove i oštriji vrh (leptokurtična razdioba) ili lakše repove i ravniji vrh (platikurtična razdioba) u usporedbi s normalnom razdiobom. Formula za spljoštenost glasi:

$$\text{Spljoštenost} = \frac{E[(X-\mu)^4]}{\sigma^4} - 3 \tag{5.12}$$

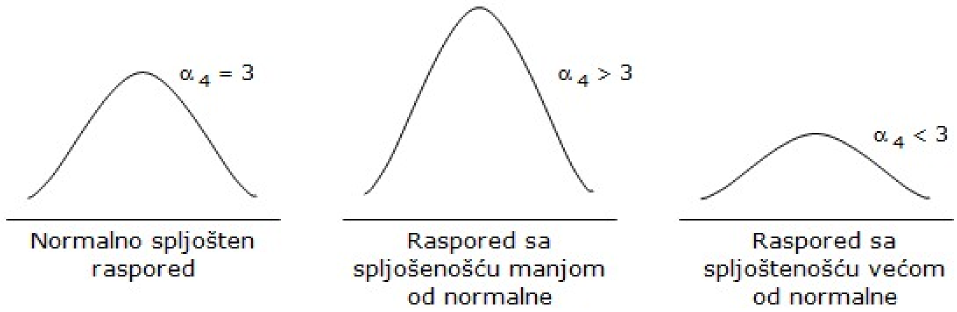
Spljoštenost može poprimiti razne vrijednosti, no bilo koje odstupanje od 0 ukazuje na odstupanje od normalnosti:

- Ako je spljoštenost pozitivna, to znači da je razdioba leptokurtična; ima teže repove i oštriji vrh u usporedbi s normalnom razdiobom. Označava prisutnost ekstremnijih vrijednosti.

- Ako je spljoštenost jednaka nuli, onda je razdioba mezokurtična; slična je normalnoj razdiobi u smislu repa i vrha.

- Ako je spljoštenost manja od nule, to znači da je razdioba platikurtična; ima lake repove i ravniji vrh u usporedbi s normalnom razdiobom. Označava prisutnost manje ekstremnih vrijednosti od normalne razdiobe.

Koeficijent spljoštenosti se pojavljuje u raznim područjima, uključujući financije, gdje pomaže u procjeni rizika povezanog s povratom ulaganja te u kontroli kvalitete, gdje pomaže identificirati odstupanja od normalne razdiobe.



Slika 5.2. Razdioba za različite koeficijente spljoštenosti, Izvor: [15]

6. Binomna razdioba

6.1. Bernoullijeva slučajna varijabla

Bernoullijeva slučajna varijabla je diskretna slučajna varijabla koja može imati jedan od dva moguća ishoda: 0 ili 1 [16]. Ona predstavlja samo jedan Bernoullijev pokušaj ili eksperiment koji ima samo dva moguća ishoda, često se nazivaju "uspjeh" i "neuspjeh". Bernoullijeva slučajna varijabla označava se s X i definirana je funkcijom mase vjerojatnosti kako slijedi:

$$P(X = 1) = p \text{ i } P(X = 0) = q = 1 - p, \tag{6.1}$$

za koju vrijedi:

- X: Bernoullijeva slučajna varijabla koja ima vrijednost 0 ili 1.
- p: Vjerojatnost uspjeha $(0 \leq p \leq 1)$. Predstavlja vjerojatnost da je slučajna varijabla jednaka 1.

Neke karakteristike Bernoullijeve slučajne varijable:

1. Dva ishoda

   Modelira situacije u kojima postoje točno dva moguća ishoda, kao što su uspjeh/neuspjeh, da/ne, glava/rep, itd.

2. Neovisnost

   Pretpostavlja se da je svaki Bernoullijev pokus neovisan o prethodnim pokusima. To znači da ishod jednog pokusa ne utječe na ishod drugog.

3. Konstantna vjerojatnost

   Vjerojatnost uspjeha p ostaje ista za svaki pokušaj.

U binomnoj razdiobi broj uspjeha se može modelirati u fiksnom broju Bernoullijevih pokušaja.

6.1.1. Numeričke karakteristike Bernoullijeve slučajne varijable

Očekivanje Bernoullijeve slučajne varijable se lako računa i iznosi:

$$E[x] = p, \tag{6.2}$$

jer za Bernoullijevu slučajnu varijablu vrijedi:

$$P(X = 1) = p \text{ i } P(X = 0) = q, \tag{6.3}$$

iz čega slijedi:

$$E[x] = P(X = 1) \cdot 1 + P(X = 0) \cdot 0 = p \cdot 1 + q \cdot 0 = p. \tag{6.4}$$

Varijanca se također može lako odrediti na način da se prvo izračuna očekivana vrijednost:

$$E[X^2] = P(X = 1) \cdot 1^2 + P(X = 0) \cdot 0^2 = p \cdot 1^2 + q \cdot 0^2 = p = E[X], \tag{6.5}$$

iz čega slijedi izraz za varijancu:

$$\text{Var}[X] = E[X^2] - E[X]^2 = E[X] - E[X]^2 = p - p^2 = p(1-p) = pq. \tag{6.6}$$

Momenti se računaju s obzirom na poznat moment reda k:

$$\mu_k = (1-p)(-p)^k + p(1-p)^k, \tag{6.7}$$

iz čega se dobivaju prvih šest momenata:

$$\begin{aligned}
\mu_1 &= 0, \\
\mu_2 &= p(1-p), \\
\mu_3 &= p(1-p)(1-2p), \\
\mu_4 &= p(1-p)(1-3p(1-p)), \\
\mu_5 &= p(1-p)(1-2p)(1-2p(1-p)), \\
\mu_6 &= p(1-p)(1-5p(1-p)(1-p(1-p))).
\end{aligned}$$

6.2. Definicija binomne razdiobe

Binomna razdioba je diskretna vjerojatnost koja modelira broj uspjeha (obično označen s k) u fiksnom broju nezavisnih Bernoullijevih pokusa, gdje svaki pokus ima samo dva moguća ishoda: uspjeh ili neuspjeh [17]. Pretpostavlja se da su ta ispitivanja identična i neovisna, što znači da ishod jednog ispitivanja ne utječe na ishod drugog. Neke ključne karakteristike binomne razdiobe su:

1. Parametri

   Binomna razdioba se definira pomoću dva parametra:

   - n: ukupan broj pokušaja,
   - p: vjerojatnost uspjeha u bilo kojem pokušaju, $p \in (0, 1)$.

2. Funkcija gustoće vjerojatnosti

   Funkcija gustoće vjerojatnosti binome razdiobe daje vjerojatnost točno k uspjeha u n broj pokušaja. Označava se s $P(X = k)$ te se računa koristeći formulu:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}, \tag{6.8}$$

   za koju vrijedi:

- $\binom{n}{k}$ predstavlja binomni koeficijent, što je broj načina odabira k uspjeha od n pokušaja te se računa kao: $\frac{n!}{k!(n-k)!}$
- p^k predstavlja vjerojatnost k uspjeha.
- $(1-p)^{n-k}$ predstavlja vjerojatnost $(n-k)$ neuspjeha.

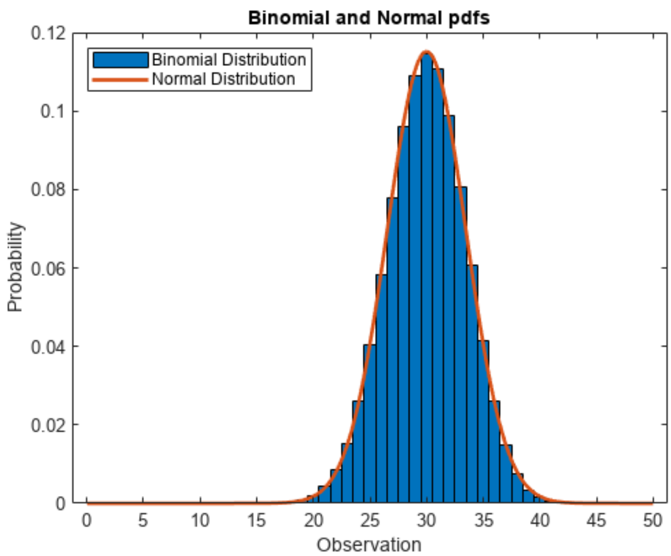
3. Oblik razdiobe

   Oblik binomne razdiobe ovisi o vrijednostima n i p. Kako se n povećava i/ili p približava 0.5, razdioba postaje više simetrična i zvonolikija, nalik normalnoj razdiobi.

4. Primjena

   Binomna razdioba obično se koristi u raznim scenarijima stvarnog svijeta, kao što su:

   - Modeliranje broja uspješnih pokušaja u fiksnom broju pokušaja (na primjer, broj glava u 10 bacanja novčića).
   - Analiza procesa kontrole kvalitete (na primjer, broj neispravnih artikala u određenom uzorku).
   - Predviđanje vjerojatnosti dobitka u nizu pokušaja (na primjer u sportu ili igrama na sreću).



Slika 6.1. Usporedba binomne i normalne razdiobe, Izvor: [18]

Primjer 6.1 (Analiza procesa kontrole kvalitete)**.** *Recimo da neka proizvođačka tvrtka proizvodi žarulje, a poznato je da je 5% proizvedenih žarulja neispravno. Kako bi se osigurala kontrola kvalitete, iz svake proizvodne serije nasumično se odabire uzorak od 10 žarulja te se bilježi broj neispravnih žarulja u uzorku. Postavljaju se pitanja:*

1. *Kolika je vjerojatnost da su točno 2 od 10 žarulja iz uzorka neispravne?*

2. *Kolika je vjerojatnost da su najmanje 3 od 10 žarulja iz uzorka neispravne?*

Rješenje:

1. *Vjerojatnost točno 2 neispravne žarulje* $(k = 2)$*:*

   - n *(broj pokušaja)= 10 (budući da je u uzorku 10 žarulja).*
   - p *(vjerojatnost uspjeha)=0.05* $(5\%$ *žarulja je neispravno).*
   - k *(željeni rezultat)= 2.*

   Koristeći binomnu funkciju gustoće vjerojatnosti dobije se:

$$P(X = 2) = \binom{10}{2} \cdot (0.05)^2 \cdot (1 - 0.05)^{10-2} = 45 \cdot 0.0025 \cdot 0.9512 = 0.10701. \tag{6.9}$$

   Dakle, vjerojatnost da su točno 2 od 10 žarulja iz uzorka neispravne je 0.10701, odnosno 10.63%.

2. *Vjerojatnost najmanje 3 neispravne žarulje* $(k \geq 3)$*:*

   Kako bi se izračunala vjerojatnost najmanje 3 neispravne žarulje potrebno je zbrojiti vjerojatnosti za $k = 3, 4, 5, \ldots, 10$*:*

$$P(X \geq 3) = P(X = 3) + P(X = 4) + \cdots + P(X = 10). \tag{6.10}$$

   Zbroj vjerojatnosti svih članova će dati odgovor na pitanje, međutim korištenjem svojstava vjerojatnosti, $P(X \geq 3)$ *možemo izračunati malo lakše:*

$$\begin{aligned} P(X \geq 3) &= P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) \\ &= \binom{10}{0} \cdot 0.05^0 \cdot (1 - 0.05)^{10-0} + \binom{10}{1} \cdot 0.05^1 \cdot (1 - 0.05)^{10-1} + 0.07463 \\ &= 0.59874 + 0.31512 + 0.07463 \\ &= 0.98849. \end{aligned} \tag{6.11}$$

   Iz računa se zaključuje da je vjerojatnost da je najmanje 3 od 10 žarulja neispravno jednaka 0.98849, odnosno 98.85%.

Primjer 6.2 (Kartaška igra)**.** *Kao primjer se može navesti i kartaška igra "blackjack". U igri blackjack igraču se dodijele dvije karte iz standardnog špila od 52 karte. Cilj je u ruci imati vrijednost karata što bližu 21, bez da se ta vrijednost prekorači. Karte s licem (kralj, dama, dečko) računaju se kao 10 bodova, numerirane karte se računaju kao njihova nominalna vrijednost, a as se može računati kao 1 ili 11 bodova, ovisno o tome koja vrijednost ide u korist igraču. Recimo da igrač želi znati kolika je vjerojatnost da dobije točno jedan as u početne dvije karte. U tom slučaju uvjeti su:*

- $n = 2$ *(igrač dobiva dvije karte),*
- p *(vjerojatnost da dobije jedan as)=* $\frac{4}{52} = \frac{1}{13}$ *(jer su četiri asa u špilu od 52 karte),*
- k *(željeni uspjeh)= 1.*

$$P(X = 1) = \binom{2}{1} \cdot \left(\frac{1}{13}\right)^1 \cdot \left(1 - \frac{1}{13}\right)^{2-1} = 2 \cdot \frac{1}{13} \cdot \frac{12}{13} \approx 0.14201. \tag{6.12}$$

Vjerojatnost da se igraču na početku igre dodijeli točno jedan as je približno 0.14201 ili 14.2%.

6.3. Očekivanje binomne razdiobe

U kontekstu binomne razdiobe, očekivanje (očekivana vrijednost) predstavlja srednju ili prosječnu vrijednost razdiobe. Daje ideju o središnjoj tendenciji razdiobe i pomaže razumjeti koja vrijednost se može očekivati, u prosjeku, za slučajnu varijablu nakon binomne razdiobe. Formula glasi:

$$E(X) = np. \tag{6.13}$$

Taj izraz lako slijedi iz linearnosti očekivane vrijednosti zajedno s činjenicom da je X zbroj n identičnih Bernoullijevih slučajnih varijabli, svaka s očekivanom vrijednošću p. Drugim riječima, ako su $X_1, \ldots, X_n$ identične (i neovisne) Bernoullijeve slučajne varijable s očekivanom vrijednošću p, tada je $X = X_1 + \cdots + X_n$ i vrijedi:

$$E[X] = E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n] = p + \cdots + p = np. \tag{6.14}$$

Drugim riječima, ako bi se ponovio niz od n neovisnih Bernoullijevih pokusa, svaki s vjerojatnošću p uspjeha, očekivana vrijednost predstavlja prosječan broj uspjeha koji bi se očekivao postići tijeko n pokusa. Ovo svojstvo se koristi za predviđanje i informiranje donošenja odluka u scenarijima gdje je binomna razdioba primjenjiva, kao što je kontrola kvalitete, testiranje hipoteza i razne primjene u stvarnom svijetu gdje se pojavljuju eksperimenti uspjeha i neuspjeha.

Primjer 6.3 (Kontrola kvalitete)**.** *Recimo da neka tvrtka koja proizvodi elektroničke komponente proizvede 10% neispravnih komponenti. Tim za kontrolu kvalitete nasumično odabere 50 komponenti za inspekciju te želi izračuanti očekivani broj neispravnih komponenti među tih 50. U ovom slučaju može se koristiti binomna razdioba za modeliranje broja neispravnih komponenti u uzorku. Svaka komponenta koja se ispituje je Bernoullijev pokus koja ima 10% šanse da je neispravna. Uvjeti za navedeni pokus glase:*

- $n = 50$ *(50 uzoraka koji se ispituju),*
- $p = 0.1$ *(vjerojatnost da je komponenta neispravna),*
- *slučajna varijabla* X*: broj neispravnih komponenti u uzorku od 50 komponenti.*

Sada kada je svaka stavka objašnjena može se koristiti formula za očekivanu vrijednost:

$$E(X) = np = 50 \cdot 0.1 = 5. \tag{6.15}$$

Dakle, u prosjeku se može očekivati 5 neispravnih komponenti u uzorku od 50 pregledanih iz ovog proizvodnog procesa.

6.4. Disperzija binomne razdiobe

U kontekstu binomne razdiobe, disperzija se obično odnosi na varijabilnost ili širenje razdiobe, a standardna devijacija je jednostavno kvadratni korijen varijance koja u slučaju binomne razdiobe iznosi:

$$\text{Var}(X) = \sigma^2 = np(1-p) = npq. \tag{6.16}$$

Varijanca i devijacija pomažu u procjeni razine neizvjesnosti povezane s brojem uspjeha u fiksnom broju pokusa s danom vjerojatnošću uspjeha. U praktičnom smislu, manja varijanca i standardna devijacija pokazuju da su ishodi gušće grupirani oko srednje vrijednosti, dok veća varijanca i standardna devijacija sugeriraju da su ishodi više rašireni. Razumijevanje ovog širenja vrijedi za različite primjene, kao što je kontrola kvalitete, procjena rizika i statističko zaključivanje koje uključuje binomno distribuirane podatke.

Primjer 6.4 (Ispit s višestrukim izborom)**.** *Recimo da u ispitu s višestrukim izborom svako pitanje ima dva moguća odgovora: točan ili netočan. Student polaže ispit od 20 pitanja te ima 70% šanse da točno odgovori na svako pitanje i 30% šanse da odgovori netočno. Želimo izračunati varijancu kako bismo razumjeli disperziju u broju točnih odgovora koje će student vjerojatno postići. U ovom slučaju binomna razdioba se može koristiti za modeliranje broja točnih odgovora koje će student dobiti od ukupnih 20. Iz navedenog vrijede idući uvjeti:*

- $n = 20$ *(ukupno 20 pitanja na ispitu),*
- $p = 0.7$ *(vjerojatnost da će student točno odgovoriti na pitanje),*
- $q = 0.3$ *(vjerojatnost da će student netočno odgovoriti na pitanje),*
- *slučajna varijabla X: broj točnih odgovora od ukupnih 20.*

Za dodatno shvaćanje ćemo izračunati i očekivanu vrijednost:

$$E(X) = 20 \cdot 0.7 = 14. \tag{6.17}$$

Sada se može izračunati varijanca:

$$\textit{Var}(X) = npq = 20 \cdot 0.7 \cdot 0.3 = 4.2. \tag{6.18}$$

Ova vrijednost varijance označava stupanj disperzije u broju točnih odgovora koje student očekuje da postigne. U ovom slučaju, varijanca od 4.2 znači da iako je očekivana vrijednost 14 točnih odgovora, stvarni broj točnih odgovora vjerojatno će varirati oko ove srednje vrijednosti. Neki studenti mogu postići više od 14 točnih odgovora, dok drugi mogu postići manje. Upravo to odražava inherentnu varijabilnost rezultata ispita zbog vjerojatnosne prirode pitanja.

6.5. Stabilnost binomne razdiobe

Stabilnost razdiobe vjerojatnosti odnosi se na ponašanje te razdiobe dok se slučajne varijable kombiniraju ili zbrajaju.

1. Stabilnost pri zbrajanju

   Ako su $X_1 \sim B(n_1, p)$ i $X_2 \sim B(n_2, p)$ nezavisne binomne slučajne varijable, onda je i $X_1 + X_2$ binomna slučajna varijabla s parametrima $n_1 + n_2$ i p tj. $X_1 + X_2 \sim B(n_1 + n_2, p)$. Promatrani pokus je podijeljen u dvije cjeline. U prvoj cjelini pokus je ponovljen n_1 puta, a u drugoj cjelini n_2 puta. Zbroj realizacija u pojedinim cjelinama daje broj realizacija nekog događaja u cijelom pokusu.

2. Centralni granični teorem

   Binomna razdioba postaje približno normalna (Gaussova) kada je broj pokušaja n velik i vjerojatnost uspjeha p nije preblizu 0, ni 1. Ovo dokazuje da binomna razdioba konvergira normalnoj razdiobi kako se n povećava. Centralni granični teorem tvrdi da će zbroj (ili prosjek) velikog broja neovisnih, identično distribuiranih slučajnih varijabli imati približno normalnu razdiobu, bez obzira na izvornu razdiobu.

3. Stabilnost u prisutnosti drugih razdioba

   Kada se binomna razdioba kombinira s drugim razdiobama kroz matematičke operacije (npr. množenje, konvolucija), rezultirajuća razdioba možda neće biti binomna, ali se često može opisati korištenjem poznatih razdioba, kao što je Poissonova ili normalna. Ovo ukazuje na stabilnost s obzirom na način kakvu interakciju razdioba ima s drugima.

7. Poissonova razdioba

Poissonova razdioba je diskretna razdioba vjerojatnosti koja modelira broj događaja ili pojava koji se događaju unutar fiksnog intervala vremena ili prostora [19]. Ime je dobila po francuskom matematičaru Siméonu Denisu Poissonu, koji ju je predstavio početkom 19. stoljeća. Neke ključne karakteristike Poissonove razdiobe su:

1. Diskretnost

   Poissonova razdioba se bavi diskretnim slučajnim varijablama, što znači da se broj događaja računa cijelim brojevima $(0, 1, 2, 3, \dots)$.

2. Neovisnost

   Pretpostavlja se da se događaji odvijaju neovisno unutar fiksnog intervala. Pojava jednog događaja ne utječe na vjerojatnost da će se dogoditi drugi događaj.

3. Konstantna stopa

   Poissonova razdioba je parametrizirana samo jednom pozitivnom vrijednošću, označenom s λ, koja predstavlja prosječan broj događaja koji se pojavljuje u fiksnom intervalu. U matematičkom smislu, λ je očekivana vrijednost (srednja vrijednost) razdiobe.

Funkcija gustoće vjerojatnosti Poissonove razdiobe, označena kao $P(X = k)$ daje vjerojatnost opažanja točno k događaja u nekom intervalu:

$$P(X = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!} \tag{7.1}$$

Za navedenu funkciju gustoće vrijedi:

- X: slučajna varijabla Poissonove razdiobe koja predstavlja broj događaja,
- k: određeni broj događaja za koje se želi izračunati vjerojatnost,
- λ: prosječna stopa događaja unutar određenog intervala,
- $e \approx 2.71828$: Eulerov broj,
- $k!$: faktorijel od k.

Ako slučajna varijabla X slijedi Poissonovu razdiobu s parametrom λ, pišemo $X \sim P(\lambda)$.

Primjer 7.1 (Pojava potresa)**.** *Pretpostavimo da je potrebno izračunati pojavu rijetkih potresa u određenoj regiji. Povijesno gledano, potresi su se u promatranoj regiji događali s prosječnom stopom od 2 potresa godišnje. Poissonova razdioba se može koristiti za izračun vjerojatnosti povezanih s različitim scenarijima potresa te se može modelirati broj potresa u godini:*

1. *Vjerojatnost da se ne desi potres* $(P(X=0))$*:*

$$P(X=0)=\frac{e^{-2}\cdot 2^{0}}{0!}=\frac{e^{-2}}{1}\approx 0.1353. \tag{7.2}$$

2. *Vjerojatnost da se desi jedan potres* $(P(X=1))$*:*

$$P(X=1)=\frac{e^{-2}\cdot 2^{1}}{1!}=2e^{-2}\approx 0.2707. \tag{7.3}$$

3. *Vjerojatnost da se dese barem dva potresa* $(P(X\geq 2))$*: Za izračunati ovu vjerojatnost potrebno je pronaći komplement "najviše jednog potresa":*

$$\begin{aligned} P(X\geq 2) &= 1-P(X\leq 1)=1-(P(X=0)+P(X=1)) \\ &= 1-(0.1353+0.2707) \\ &\approx 0.5940. \end{aligned} \tag{7.4}$$

Koristeći Poissonovu razdiobu za modeliranje rijetke pojave potresa izračunate su vjerojatnosti povezane s različitim scenarijima potresa. Postoji približno 13.53% šanse da u godini nema potresa, 27.07% šanse jednog potresa i 59.40% šanse za najmanje dva potresa.

Primjer 7.2 (Broj kupaca u trgovini)**.** *U idućem primjeru ćemo razmotriti maloprodajnu trgovinu koja želi razumjeti obrazac dolazaka kupaca tijekom radnog dana. Trgovina je primijetila da u prosjeku dolazi 30 kupaca svakog sata što predstavlja prosječnu stopu. Međutim, žele analizirati dolaske kupaca u manjim vremenskim intervalima, poput intervala od 15 minuta i izračunati vjerojatnosti povezane s različitim stopama dolazaka. Upravo zbog toga će se prosječna stopa* λ *raščlaniti u manje intervale. Jedan sat se sastoji od četiri intervala od 15 minuta pa će prosječna stopa za svaki interval biti* $\frac{30}{4}=7.5$ *klijenata po intervalu.*

1. *Vjerojatnost da neće doći nijedan kupac u intervalu od 15 minuta* $(P(X=0))$*:*

$$P(X=0)=\frac{e^{-7.5}\cdot 7.5^{0}}{0!}=e^{-7.5}\approx 0.00056. \tag{7.5}$$

2. *Vjerojatnost da će u intervalu od 15 minuta doći točno 5 kupaca* $(P(X=5))$*:*

$$P(X=5)=\frac{e^{-7.5}\cdot 7.5^{5}}{5!}\approx 0.10679. \tag{7.6}$$

3. *Vjerojatnost da će doći više od 10 kupaca u intervalu od 30 minuta* $(P(X>10))$*:*

Za interval od 30 minuta trebaju se uzeti u obzir dva uzastopna intervala od 15 minuta stoga će se izračunati vjerojatnost primanja 10 ili manje kupaca u svakom intervalu te potom pronaći komplement:

$$P(X > 0) = 1 - (P(X \leq 10) \cdot P(X \leq 10)) \approx 0.05975. \tag{7.7}$$

Vrlo je mala vjerojatnost (otprilike 0,056%) da trgovina neće primiti nijednog kupca u intervalu od 15 minuta, 10.679% vjerojatnosti da će primiti točno 5 kupaca i 5.975% vjerojatnosti da će primiti više od 10 kupaca u intervalu od 30 minuta. Ova analiza može pomoći maloprodajnoj trgovini da bolje razumije i planira dolaske kupaca tijekom različitih doba dana.

7.1. Očekivanje Poissonove razdiobe

Očekivanje (očekivana vrijednost ili srednja vrijednost) slučajne varijable Poissonove razdiobe jednostavno je prosječna ili središnja vrijednost te razdiobe. Za Poissonovu razdiobu očekivanje je jednako parametru λ koji predstavlja prosječnu stopu ili prosječan broj događaja koji se pojavljuju u fiksnom intervalu:

$$E(X) = \lambda. \tag{7.8}$$

Ova činjenica se može lako dokazati iz definicije očekivane vrijednosti:

$$E(X) = \sum_{x \in Im(X)} xP(X = x), \tag{7.9}$$

jer prema definiciji Poissonove razdiobe imamo:

$$E(X) = \sum_{k \geq 0} k \frac{1}{k!} \lambda^k e^{-\lambda}. \tag{7.10}$$

Iz navedenog slijedi:

$$\begin{aligned} E(X) &= \lambda e^{-\lambda} \sum_{k \geq 1} \frac{1}{(k-1)!} \lambda^{k-1}, \text{ za } j = k-1 : \\ &= \lambda e^{-\lambda} \sum_{j \geq 0} \frac{\lambda^j}{j!} \\ &= \lambda e^{-\lambda} e^{\lambda} \\ &= \lambda \end{aligned} \tag{7.11}$$

Primjer 7.3 (Taksi stajalište)**.** *Vlasnik taksi stajališta u prometnoj zračnoj luci želi izračunati prosječan broj taksija koji stignu na njegovo stajalište tijekom tipičnog intervala od 30 minuta. Na temelju prijašnjih podataka, zna da u prosjeku dolazi 8 taksija kroz jedan sat. Dakle, prosječna stopa je 8 taksija na sat, no mi ćemo računati broj taksija koji stižu u intervalu od 30 minuta pa je potrebno preračunati prosječnu stopu:*

$$\lambda = \frac{8}{2} = 4 \textit{ taksija u 30 minuta}. \tag{7.12}$$

Dakle, za očekivati je da će četiri taksija stići na taksi stajalište u intervalu od 30 minuta. Međutim, stvarni broj taksija koji stižu u bilo kojem intervalu od 30 minuta varirat će od jednog intervala do drugog zbog slučajne prirode Poissonove razdiobe.

7.2. Disperzija Poissonove razdiobe

Disperzija razdiobe je mjera koliko su raširene ili promjenjive vrijednosti slučajne varijable oko svoje srednje vrijednosti (očekivanja). U slučaju Poissonove razdiobe, disperzija je jednaka parametru λ, što je također srednja vrijednost razdiobe:

$$\mathrm{Var}(X) = \lambda. \tag{7.13}$$

Standardna devijacija Poissonove razdiobe, koja je mjera koliko vrijednosti odstupaju od srednje vrijednosti, kvadratni je korijen varijance:

$$\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{\lambda}. \tag{7.14}$$

U praktičnom smislu, to govori o tome koliko stvarni broj događaja (kao što su dolasci, događaji ili nesreće) ima tendenciju da varira oko prosječne stope λ. Veća λ označava veću varijaciju, i obrnuto.

Primjer 7.4 (E-trgovina)**.** *Primjerice, voditelja e-trgovine može zanimati dnevni broj prodajnih naloga. Poznati su povijesni podatci koji pokazuju prosječnu stopu od 50 narudžbi dnevno te je potrebno izračunati varijancu kako bi se razumio stupanj varijabilnosti u dnevnoj prodaji. Dakle, prosječna stopa je 50 narudžbi dnevno te se koristeći formulu Poissonove razdiobe Var$(X) = \lambda$ dobiva varijanca dnevnog broja prodajnih naloga:*

$$\mathit{Var}(X) = 50. \tag{7.15}$$

Voditelj e-trgovine zaključuje da je varijanca broja dnevnih prodajnih naloga 50. S druge strane, standardna devijacija jednaka je korijenu varijance tj.

$$\sigma(X) = \sqrt{\mathit{Var}(X)} = \sqrt{50} \approx 7.07. \tag{7.16}$$

7.3. Stabilnost Poissonove razdiobe

Ako su $X_1 \sim P(\lambda_1)$ i $X_2 \sim P(\lambda_2)$ nezavisne Poissonove slučajne varijable, onda je i $X_1 + X_2$ Poissonova slučajna varijabla s parametrima $\lambda_1 + \lambda_2$, odnosno $X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$. Poissonova razdioba ima neka značajna svojstva stabilnosti:

1. Stabilnost pri zbrajanju

Ako se zbroji veliki broj nezavisnih Poissonovih slučajnih varijabli, zbroj teži slijediti drugu Poissonovu razdiobu. Točnije, ako su $X_1, X_2, \ldots, X_n$ nezavisne slučajne varijable s parametrima $\lambda_1, \lambda_2, \ldots, \lambda_n$ onda suma $X = X_1 + X_2 + \cdots + X_n$ prati Poissonovu razdiobu s parametrima koji su jednaki sumi zasebnih parametara:

$$X \sim P(\lambda_1 + \lambda_2 + \cdots + \lambda_n). \tag{7.17}$$

Ovo je svojstvo korisno kod modeliranja broja događaja koji se pojavljuju u fiksnom vremenskom ili prostornom intervalu.

2. Stabilnost pri skaliranju

   Ako se Poissonova slučajna varijabla skalira s konstantnim faktorom, rezultat je i dalje unutar Poissonove razdiobe. Ako je X slučajna varijabla s parametrom λ te je a pozitivna konstanta, onda je $Y = aX$ u Poissonovoj razdiobi s parametrom $\lambda' = a\lambda$:

$$Y \sim P(a\lambda). \tag{7.18}$$

   Ovo je svojstvo korisno kada se radi o situaciji u kojoj se događaji odvijaju različitom brzinom ili intenzitetom.

7.4. Prilagođavanje Poissonove razdiobe empiričkim podatcima

Prilagodba Poissonove razdiobe empiričkim podatcima uključuje usporedbu promatranih podataka s očekivanom razdiobom i izmjene ako promatrani podatci značajno odstupaju od Poissonovog modela. Ovaj postupak prilagodbe pomaže u poboljšanju modela kako bi bolje odgovarao podatcima. Opći koraci za prilagođavanje Poissonove razdiobe empiričkim podatcima su:

1. Prikupiti i organizirati podatke

   Prvo je potrebno prikupiti i organizirati empiričke podatke koji se žele analizirati. Podatci trebaju predstavljati događaje koji bi potencijalno mogli slijediti Poissonovu razdiobu.

2. Odrediti Poissonov parametar λ

   Računa se srednja vrijednost uzorka λ iz empiričkih podataka. Ova srednja vrijednost uzorka poslužit će kao procjena parametra za Poissonovu razdiobu.

3. Provesti test prikladnosti

   Potrebno je provesti statističke testove za procjenu koliko dobro Poissonova razdioba odgovara empiričkim podatcima. Uobičajeni testovi usklađenosti za podatke su χ^2 test i Kolmogorov-Smirnov test.

4. Procijeniti prikladnost

   Ispituju se rezultati testa prilagodbe. Ako je p vrijednost visoka (što ukazuje na dobro uklapanje), a promatrani podatci vrlo dobro odgovaraju Poissonovoj razdiobi, možda nije potrebno vršiti daljnje prilagodbe.

5. Provjeriti prekomjernu disperziju

   Ponekad podatci empiričkog brojanja mogu pokazivati veću varijabilnost (prekomjernu disperziju) nego što to može objasniti Poissonova razdioba. U takvim slučajevima potrebno je razmotriti alternativne razdiobe poput negativne binomne razdiobe ili Poisson-inverzne Gaussove razdiobe, koje mogu modelirati prekomjerno disperzirane podatke brojanja.

6. Modifikacije modela (po potrebi)

   Ako provedeni test prilagodbe sugerira da Poissonova razdioba nije dobra prilagodba, moguće je modificirati model.

7. Procijeniti odgovaranje modela

   Nakon modifikacija modela, test se ponavlja kako bi se procijenilo koliko dobro prilagođeni model odgovara podatcima.

Izbor razdiobe i modifikacija bi se trebao temeljiti na karakteristikama podataka i ciljevima analize, stoga je poželjno konzultirati se sa statističarom za smjernice o radu sa složenim zadatcima modeliranja podataka.

8. Aproksimacija binomne razdiobe Poissonovom

Poissonova razdioba može se koristiti kao aproksimacija za binomnu razdiobu pod određenim uvjetima. Ovaj slučaj je poznat kao Poissonova aproksimacija binomne razdiobe. Uvjeti za navedenu aproksimaciju su [20]:

1. Mala vjerojatnost uspjeha p

   Matematički se ovaj uvjet može izraziti kao $np \leq 10$ ili čak $np \leq 5$ za konzervativniju aproksimaciju.

2. Veliki uzorak n

   Naravno, zbog subjektivnosti ne postoji strogo pravilo što se smatra "velikim" uzorkom, ali obično n treba biti najmanje 20.

Koraci za provođenje navedene aproksimacije:

1. Izračunati srednju vrijednost Poissonove razdiobe ($\lambda = np$).

2. Aproksimirati binomnu razdiobu Poissonovom s istom srednjom vrijednosti. Drugim riječima, X je binomna slučajna varijabla te ju se aproksimira kao:

$$Y \sim P(\lambda). \tag{8.1}$$

Primjer 8.1 (Širenje rijetke bolesti)**.** *Pretpostavimo da se provodi studija kako bi se razumjela pojava rijetke bolesti u populaciji. Poznato je da je u prosjeku samo jedan na svakih 1000 pojedinaca u populaciji pogođen bolešću* ($p = 0.001$)*. No, nas zanima populacija od 10000 pojedinaca* ($n = 10000$)*. Koraci aproksimacije binomne razdiobe Poissonovom su:*

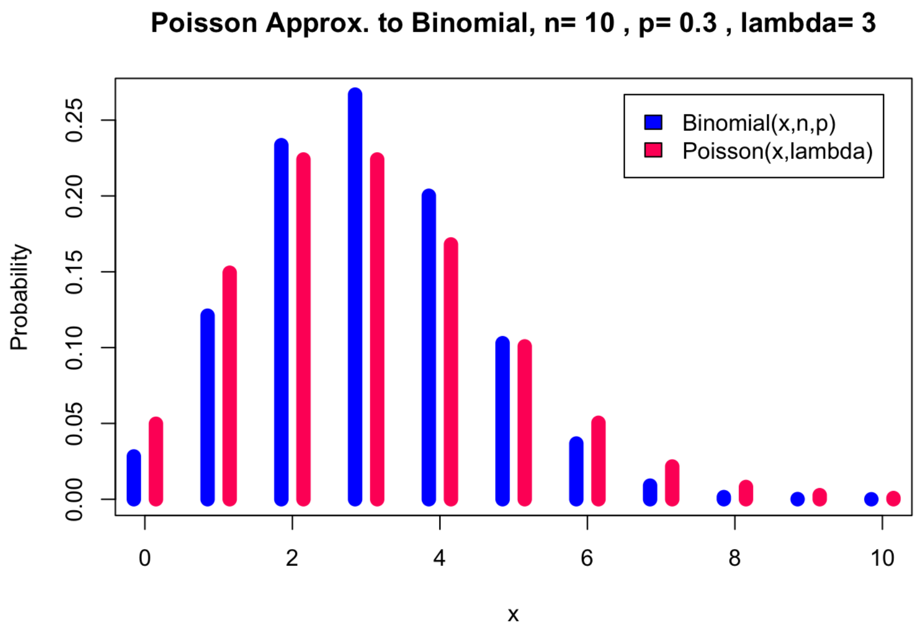
1. *Računanje srednje vrijednosti Poissonove razdiobe:* $\lambda = np = 10000 \cdot 0.001 = 10.$

2. *Aproksimacija binomne razdiobe Poissonovom s* $\lambda = 10$*. Dakle, broj zaraženih pojedinaca se modelira kao:*

$$Y \sim P(10). \tag{8.2}$$

Sada recimo da nas zanima vjerojatnost da je u ovoj populaciji zaraženo manje od 5 osoba:

$$\begin{aligned} P(Y < 5) &= P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3) + P(Y = 4) \\ &= \frac{e^{-10} \cdot 10^0}{0!} + \frac{e^{-10} \cdot 10^1}{1!} + \frac{e^{-10} \cdot 10^2}{2!} + \frac{e^{-10} \cdot 10^3}{3!} + \frac{e^{-10} \cdot 10^4}{4!} \\ &= 4.54 \cdot 10^{-5} + 4.54 \cdot 10^{-4} + 2.27 \cdot 10^{-3} + 7.57 \cdot 10^{-3} + 0.01891 \\ &= 0.02925. \end{aligned} \tag{8.3}$$



Slika 8.1. Aproksimacija binomne razdiobe Poissonovom za $n = 10, p = 0.3$ i $\lambda = 3$, Izvor: [21]

Iz rezultata računa se zaključuje da je vjerojatnost da je u populaciji zaraženo manje od 5 osoba jednaka 0.02925, odnosno 2.93%.

Svaka od ovih vjerojatnosti se može izračunati pomoću Poissonove funkcije gustoće vjerojatnosti. Ova aproksimacija je brži i lakši način računanja te se koristi kada nije dobro pozanto koliki je n (iako je poznato da je velik). Aproksimacija daje i razumno točnu procjenu vjerojatnosti promatranja malog broja slučajeva rijetke bolesti u velikoj populaciji. Osobito je korisna u studijama epidemiologije i javnog zdravlja kada se radi o rijetkim bolestima tj. kada se radi o situacijama u kojima je broj pokušaja velik, a vjerojatnost uspjeha niska, što čini rad s Poissonovom razdiobom računalno učinkovitijom. Međutim, bitno je napomenuti da je Poissonova aproksimacija binomne razdiobe samo aproksimacija i možda neće biti točna ako uvjeti nisu ispunjeni.

9. Zaključak

Ovaj rad posvećen je binomnoj i Poissonovoj diskretnoj razdiobi slučajne varijable. U teoriji vjerojatnosti i statistici, razdioba, koja se često naziva i razdioba vjerojatnosti ili statistička razdioba, matematička je funkcija koja opisuje kako se vrijednosti slučajne varijable šire ili raspodjeljuju po mogućim ishodima. Razdioba je bitna matematička funkcija jer pruža informacije o vjerojatnosti promatranja svake moguće vrijednosti slučajne varijable.

Binomna razdioba modelira broj uspjeha u fiksnom broju neovisnih Bernoullijevih pokusa, gdje svaki pokus ima samo dva moguća ishoda. U ovom radu objasnili smo relevantne pokazatelje kao što su matematičko očekivanje, te disperziju i stabilnost. S druge strane, Poissonova razdioba koja je dobila ime po matematičaru Siméonu Denisu Poissonu, modelira broj događaja koji se dogode unutar fiksnog intervala vremena ili prostora. Kao i kod binomne razdiobe i u ovom slučaju dan je pregled osnovnih numeričkih karakteristika razdiobe te je objašnjeno kako se Poissonova razdioba može prilagoditi empiričkim podatcima. Važna poveznica ove dvije razdiobe je mogućnost aproksimacije binomne razdiobe korištenjem Poissonove, što uvelike može olakšati razne proračune u primjeni.

Kroz rad smo dokazali da je primjena binomne i Poissonove razdiobe uistina široka. Binomna razdioba se koristi u raznim slučajevima kao što je kontrola kvalitete, biološke studije, provođenje anketa, bacanje novčića i igraćih kockica, rezultati izbora i mnogi drugi, tj. bilo koji slučaj u kojem postoje dva moguća ishoda u nizu pokusa. Neke od navedenih slučajeva smo objasnili i u radu. Poissonova razdioba se pak koristi na primjer za modeliranje podataka sportske statistike, financija, zdravstva, dolaznih poziva u pozivnom centru te za situacije koje uključuju rijetke događaje kao što su prometne nesreće, potresi, poplave i mnoge druge nesvakodnevne događaje.

Bibliografija

[1] Benšić, M.; Šuvak, N.: "Uvod u vjerojatnost i statistiku", s Interneta, https://www.mathos.unios.hr/images/uploads/31.pdf, 1. rujna 2023.

[2] Akshay, S.; "Understanding Different Types of Distributions You Will Encounter As A Data Scientist", s Interneta, https://medium.com/mytake/understanding-different-types-of-distributions-you-will-encounter-as-a-data-scientist-27ea4c375eec, 1. rujna 2023.

[3] "Dice Roll Probability: 6 Sided Dice", s Interneta, https://www.statisticshowto.com/probability-and-statistics/probability-main-index/dice-roll-probability-6-sided-dice/, 10. rujna 2023.

[4] "Set Operations: Union, Intersection, Complement, and Difference", s Interneta, https://www.statology.org/set-operations/, 7. rujna 2023.

[5] "Coin Toss Probability Formula & Solved Examples", s Interneta, https://collegedunia.com/exams/coin-toss-probability-formula-and-solved-examples-articleid-4538, 10. rujna 2023.

[6] Tomašić, L; "Matematika IV", Sveučilište u Rijeci, Rijeka, 1993.

[7] "Uvjetna vjerojatnost", s Interneta, https://www.grad.hr/vera/webnastava/vjerojatnostistatistika/html/VISch4.html, 7. rujna 2023.

[8] "Kontinuirane slučajne varijable", s Interneta, https://www.pmf.unizg.hr/_download/repository/PREDAVANJE9.pdf, 8. rujna 2023.

[9] "Random Variable Tutorial", s Interneta, https://prwatech.in/blog/data-science/data-science-modules/random-variable-tutorial/, 8. rujna 2023.

[10] "Normalna raspodjela", s Interneta, https://hr.wikipedia.org/wiki/Normalna_raspodjela, 8. rujna 2023.

[11] "Zakon velikih brojeva", s Interneta, https://www.enciklopedija.hr/natuknica.aspx?ID=69867, 8. rujna 2023.

[12] "Varijanca i standardna devijacija", s Interneta, https://stedy.hr/opisivanje-podataka/varijanca-i-standardna-devijacija, 8. rujna 2023.

[13] Tomašić, L; "Matematika IV", Sveučilište u Rijeci, Rijeka, 1993.

[14] Dodos, S; "Skewness", s Interneta, https://alevelmaths.co.uk/statistics/skewness/, 8. rujna 2023.

[15] "Statističke metode u marketingu", s Interneta, http://www.link-university.com/lekcija/Statisti%C4%8Dke-metode-u-marketingu/2923, 8. rujna 2023.

[16] Elezović, N; "Diskretna vjerojatnost", Element, Zagreb, 2013.

[17] "Binomna distribucija", s Interneta, https://stedy.hr/distribucije/binomna-distribucija, 8. rujna 2023.

[18] Berezovsky, O.; "The Difference Between Normal and Binomial Distributions - Issue 73", s Interneta, https://dataanalysis.substack.com/p/the-difference-between-normal-and, 8. rujna 2023.

[19] "Poissonova raspodjela", s Interneta, https://www.definebusinessterms.com/hr/poissonova-raspodjela/, 7. rujna 2023.

[20] Elezović, N; "Diskretna vjerojatnost", Element, Zagreb, 2013.

[21] "Approximating Binomial with Poisson", s Interneta, https://www.solon-karapanagiotis.com/post/approx_binomial/approximating-binomial-with-poisson/, 7. rujna 2023.

Sažetak i ključne riječi

U ovom radu analizirane su osnove vjerojatnosti koje su potrebne za shvaćanje pojma razdiobe. Navedene su razne vrste razdioba i njihova primjena u svakodnevnom životu. Posebno je stavljen naglasak na binomnu i Poissonovu diskretnu razdiobu. Objašnjeno je da binomna razdioba modelira broj uspjeha u fiksnom broju neovisnih Bernoullijevih pokusa, gdje svaki pokus ima samo dva moguća ishoda, a da Poissonova razdioba modelira broj događaja koji se dogode unutar fiksnog intervala vremena ili prostora. Dokazano je i da se binomna razdioba može aproksimirati Poissonovom radi lakšeg računanja.

Ključne riječi: slučajni događaj, σ-algebra, slučajna varijabla, binomna razdioba, Poissonova razdioba

Summary and key words

This paper analyzes the basics of probability, which are necessary for understanding the concept of distribution. Various types of distribution and their application in everyday life are listed. Emphasis is placed on the binomial and Poisson discrete distributions. It is explained that the binomial distribution models the number of successes in a fixed number of independent Bernoulli trials, where each trial has only two possible outcomes and that the Poisson distribution models the number of events that occur within a fixed interval of time or space. It is also proven that the binomial distribution can be approximated by the Poisson distribution for easier calculations.

Keywords: random event, σ-field, random variable, binomial distribution, Poisson distribution