Vjerojatnost i statistika u nastavi matematike

Barać, Ana

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Ana Barać

VJEROJATNOST I STATISTIKA U NASTAVI MATEMATIKE

Diplomski rad

Voditelj rada:
doc. dr. sc. Matija Bašić

Zagreb, rujan 2024.

Ovaj diplomski rad obranjen je dana ____________________ pred ispitnim povjerenstvom u sastavu:

1. ____________________________, predsjednik
2. ____________________________, član
3. ____________________________, član

Povjerenstvo je rad ocijenilo ocjenom ________________.

Potpisi članova povjerenstva:

1. ____________________________
2. ____________________________
3. ____________________________

Za moju baku ♡
Zahvaljujem mentoru, doc. dr. sc. Matiji Bašiću, čija podrška, stručnost i strpljenje su bili ključni za uspješno dovršavanje ovog diplomskog rada.
Posebnu zahvalnost upućujem svojoj obitelji i prijateljima. Hvala Vam na podršci, motivaciji, ohrabrenju i razumijevanju tijekom cijelog mog akademskog puta.

Sadržaj

Uvod

Dugi niz godina u hrvatskom se obrazovnom sustavu premalo pažnje pridavalo učenju vjerojatnosti i statistike. Vjerojatnost, bila ona u školskom programu ili ne, sastavni je dio svakodnevnog života. Često čujemo: šanse za dobitak su 50%, svaki treći listić lutrije je dobitan, vrlo vjerojatno će sutra padati kiša, nema šanse da položimo ovaj ispit i sl. Stoga je važno da tim svakodnevnim tvrdnjama damo preciznije značenje unutar matematičke teorije. Matematika je egzaktna znanost koja se kroz povijest razvijala u skladu s razvojem čovječanstva. Među prvim granama razvile su se algebra i geometrija, dok je u novije vrijeme sve veća važnost primijenjenih grana među kojima su se smjestile vjerojatnost i statistika. Teorija vjerojatnosti je temelj statistike, pa se ove dvije grane često zajedno proučavaju te primijenjuju u različitim područjima fizike, biologije, ekonomije, medicine i dr.

U ovom diplomskom radu analizirat ćemo elemente teorije vjerojatnosti i statistike u novom kurikulumu matematike i predložiti aktivnosti kroz koje se ta znanja mogu učenicima zornije približiti, produbiti te prilagoditi za istraživački usmjerene oblike nastave. Slučajne varijable predstavljaju ključni koncept u statistici jer omogućuju modeliranje i analizu nesigurnosti i varijacije u podacima. Među različitim vrstama slučajnih varijabli, normalna razdioba igra posebnu ulogu zbog svoje široke primjene i značajnih matematičkih svojstava. Kao "središnji alat" u statistici, normalna razdioba pomaže u razumijevanju i predviđanju ponašanja podataka, a također služi kao osnovna komponenta mnogih statističkih tehnika i testova. Poseban naglasak stavit ćemo na odnos diskretnih i neprekidnih slučajnih varijabli, te primjenu Zakona velikih brojeva i Centralnog graničnog teorema kao teorijske podloge za rješavanje statističkih problema u nastavi matematike.

Uz tradicionalne metode istraživanja i analize, korišten je i ChatGPT, napredni jezični model razvijen od strane OpenAI. ChatGPT je korišten za pomoć u formuliranju ideja, strukturiranju teksta i usavršavanju izraza. Korišten je za generiranje i preoblikovanje tekstualnih dijelova, pružanje prijedloga za poboljšanje i osiguranje dosljednosti u pisanju. Ovaj alat je pružio vrijednu podršku tijekom procesa izrade rada, doprinoseći njegovom konačnom izgledu.

Poglavlje 1

Pojam i osnovna svojstva vjerojatnosti

Razumijevanje osnovnih koncepata vjerojatnosti ključno je za uspješno savladavanje izazova koji se pojavljuju već na početku učenja ovog područja matematike. Za uvođenje osnovnog jezika vjerojatnosti u nastavi matematike, potrebno je krenuti od temeljnog pojma – elementarnog događaja. Nakon što se učenici upoznaju sa skupom elementarnih događaja, nastavnik može uvesti pojam vjerojatnosnog prostora, čime se postavlja okvir za daljnje učenje. Uvođenje pojma vjerojatnosnog prostora često počinje s jednostavnim i razumljivim primjerima iz svakodnevnog života nakon čega učenici uče kako definirati različite događaje i izračunati njihove vjerojatnosti. Na kraju, dolaze do formalnih pojmova i uče da je vjerojatnosni prostor sastavljen od triju komponenti. Ovaj metodički pristup osigurava čvrste temelje za daljnje proučavanje i primjenu vjerojatnosti u nastavi matematike. Formalno uvođenje pojma vjerojatnosti, koje uključuje Kolmogorovljeve aksiome, obično se provodi tek tijekom studija.

1.1 Skup elementarnih događanja

Vjerojatnost je mjera kojom opisujemo neizvjesnost ili mogućnost pojavljivanja određenog ishoda u situaciji u kojima postoji više mogućih ishoda. Kako bismo bili precizniji, u matematici takve situacije s više mogućih ishoda nazivamo **slučajni pokus**, a svaki mogući ishod nazivamo još i **elementarni događaj**. Najčešći primjeri slučajnog pokusa su bacanje simetričnog novčića ili bacanje kocke. Prije izvođenja samog pokusa, dobro je s učenicima usmeno prokomentirati što sve možemo očekivati da će se dogoditi ukoliko bacamo novčić ili kocku odnosno koji su to elementarni događaji koje možemo očekivati. Novčić ima dvije strane, koje zovemo pismo i glava. Pri bacanju novčića jednako često će se pojavljivati obje strane, pa kažemo da je novčić simetričan. Slično, kocka ima šest strana na kojima se nalazi od 1 do 6 točkica i pretpostavljamo da će se svaka strana jednako često pojaviti. Stoga, kažemo da je kocka također simetrična.

U osnovnoj školi koriste se konkretni primjeri i jednostavni nazivi koji su intuitivni i lako razumljivi za učenike. Stoga, nije potrebno uvoditi formalnu oznaku za elementarne događaje nego je dovoljno koristiti konkretne opise. Na primjeru bacanje kocke konkretni opisi bili bi "past će broj 1", "past će broj 2"...

U nastavku ovog rada, elementarni događaj označavat ćemo s ω, a ako se radi o više njih, onda ćemo elementarne događaje označavati s ω_1, ω_2.... Osnovna pretpostavka slučajnog pokusa glasi da svako izvođenje pokusa mora dati ishod koji odgovara jednom i samo jednom elementarnom događaju. Ipak, pretpostavlja se da će rezultat takvog pokusa slijediti određene zakonitosti ako ga ponovimo dovoljno mnogo puta.

Svi elementarni događaji čine **skup elementarnih događaja** koji označavamo s Ω, a on može biti konačan, prebrojivo beskonačan ili neprebrojiv. S obzirom na temu ovog diplomskog rada i naglasak na srednjoškolsku razinu obrazovanja, definicija skupa elementarnih događaja Ω često uključuje dodatan uvjet na konačnost skupa Ω. Ova specifikacija pomaže učenicima u razumijevanju osnovnih principa vjerojatnosti kroz primjere koji su konkretni i jasni za njihovu razinu obrazovanja. Grupiranjem elementarnih događaja možemo kreirati nove, **složene događaje** ili kraće samo **događaje**. Događaj je podskup skupa elementarnih dogadaja Ω. Označimo neki događaj s D. Ako je ishod slučajnog pokusa $\omega \in \Omega$ i vrijedi $\omega \in D$, reći ćemo da se događaj D dogodio.

Primjer 1.1.1. *Odredimo skup elementarnih događaja pri bacanju novčića.*

Rješenje: Ukoliko bacamo jedan novčić, imamo dva elementarna događaja koja ćemo kratko označavati s P (pismo) i G (glava). Stoga, skup elementarnih događaja pri bacanju novčića je $\Omega = \{P, G\}$.

Primjer 1.1.2. *Odredimo skup elementarnih događaja pri bacanju dvaju novčića.*

Rješenje: Ukoliko bacamo dva novčića, imamo četiri elementarna događaja: PP, PG, GP, GG. Stoga, skup elementarnih događaja pri bacanju dvaju novčića je $\Omega = \{PP, PG, GP, GG\}$.

Primjer 1.1.3. *Odredimo skup elementarnih događaja pri bacanju kocke.*

Rješenje: Pri bacanju kocke imamo šest mogućih ishoda (pala je jedinica, pala je dvojka, pala je trojka...), što čini skup elementarnih događaja $\Omega = \{1, 2, 3, 4, 5, 6\}$.

1.2 Definicije vjerojatnosti

Sljedeći temeljni pojam koji je potrebno uvesti je frekvencija događaja, a uvodimo ga pomoću broja ponavljanja pokusa. Označimo sa n broj ponavljanja nekog pokusa, a s n_A

frekvenciju (tj. broj) ostvarivanja događaja A. Broj n_A nazivamo **frekvencija** događaja A, a omjer $\frac{n_A}{n}$ nazivamo **relativna frekvencija** događaja A u danih n ponavljanja pokusa. Iz definicije slijedi da je relativna frekvencija realni broj između 0 i 1. Razlikujemo dva ekstremna slučaja, a to su siguran događaj i nemoguć događaj. **Siguran događaj** ostvaruje se pri svakom ponavljanju pokusa i relativna frekvencija mu je jednaka 1, dok se **nemoguć događaj** ne može nikada ostvariti te mu je relativna frekvencija jednaka 0. Uz pojam frekvencije veže se jedna od klasičnih definicija vjerojatnosti odnosno vjerojatnost *a posteriori*.

Vjerojatnost *a posteriori*

Vjerojatnost *a posteriori* (lat. "iz onoga što dolazi poslije") odnosi se na znanje koje se temelji na iskustvu ili empirijskim dokazima. Drugim riječima, to je znanje do kojeg dolazimo nakon što smo obavili promatranje ili eksperiment. Zbog toga vjerojatnost *a posteriori* još nazivamo i eksperimentalnom vjerojatnosti. Vjerojatnost *a posteriori* ima smisla jer nam omogućuje da testiramo naše teorije i hipoteze u stvarnom svijetu. Bez ovakvih empirijskih potvrda, ne bismo mogli biti sigurni u točnost naših pretpostavki. Znanost i tehnologija uvelike se oslanjaju na vjerojatnost *a posteriori* metode za prikupljanje i potvrđivanje informacija. Također, važno svojstvo koje vežemo uz ovu definiciju je **svojstvo statističke stabilnosti relativnih frekvencija** koje govori da ukoliko pokus ponavljamo dovoljno velik broj puta, relativne frekvencije će se grupirati oko nekog fiksnog realnog broja.

Definicija 1.2.1. *[3] [**Vjerojatnost a posteriori**] Ako slučajni pokus zadovoljava uvjet statističke stabilnosti relativnih frekvencija, tada se vjerojatnost a posteriori poizvoljnog događaja A vezanog uz taj pokus definira kao realan broj* $\mathbb{P}(A), 0 \leq \mathbb{P}(A) \leq 1$, *oko kojeg se grupiraju, odnosno kojemu teže, relativne frekvencije tog događaja.*

Pokažimo to na primjeru:

Primjer 1.2.2. *[6] Profesorica u ruci ima 10 čavlića i baca ih tako da svi učenici mogu vidjeti ishode bacanja. Poslije svakog bacanja učenici zapisuju koliko vrhova gleda gore (oznaka G), a koliko dolje (oznaka D). Profesorica bacanje ponavlja 10 puta. Učenici su zapisali rezultate svih 10 bacanja. Odredimo približne vjerojatnosti ishoda G i D.*

bacanje	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
vrh gore (G)	4	3	5	6	5	5	2	3	4	4
vrh dolje (D)	6	7	5	4	5	5	8	7	6	6

Rješenje: Ova dva ishoda nisu simetrična, pa ne možemo očekivati da su jednako vjerojatni. Da bismo otkrili kolika je vjerojatnost ishoda G odnosno D, morali smo provesti pokus bacanja čavlića više puta i odrediti kolika je relativna frekvencija ishoda G i D. Rezultati bacanja prikazani su u tablici. Da bismo odredili približne vjerojatnosti ishoda G (vrh gore) i D (vrh dolje), analizirajmo dane podatke. Imamo 10 bacanja, a u svakom bacanju 10 čavlića. To znači da ukupno imamo 100 pojedinačnih ishoda. Iz tablice možemo izračunati ukupni broj pojavljivanja ishoda G i D. Tada, prema definiciji vjerojatnosti *a posteriori* zaključujemo da približnu vjerojatnost svakog ishoda možemo izračunati kao omjer broja pojavljivanja tog ishoda i ukupnog broja bacanja, pa je $\mathbb{P}(G) \approx 0.41$ i $\mathbb{P}(D) \approx 0.59$.

	frekvencija	relativna frekvencija
G	41	41/100
D	59	59/100

U prethodnom primjeru, profesorica je izvela 10 skupova bacanja po 10 čavlića, umjesto 100 pojedinačnih bacanja jednog čavlića. Ako pretpostavimo da su ishodi svakog čavlića neovisni jedan o drugome, 100 ishoda možemo tretirati kao skup individualnih slučajnih pokusa. Dodatno, učenici su prvo promatrali i bilježili rezultate bacanja čavlića, iz kojih su potom izračunali relativne frekvencije i procijenili vjerojatnosti ishoda G i D. Nakon deset bacanja, relativne frekvencije za ishode G i D približile su se vrijednostima 0.41 i 0.59. Da smo nastavili bacati čavliće mnogo puta, očekivali bismo da bi se te relativne frekvencije dodatno stabilizirale oko ovih vrijednosti, pod uvjetom da nema drugih varijabli koje bi utjecale na ishod. Ovo je formalizirano *Zakonom velikih brojeva*, koji kaže da će, uz dovoljno veliki broj nezavisnih i identično distribuiranih slučajnih varijabli, njihova aritmetička sredina konvergirati prema očekivanoj vrijednosti, čime pruža temelj za predviđanja i dugoročne procjene u statistici.

Primjer 1.2.3. *[6] Ante je promašio koš u 120 od 300 ponovljenih pokušaja. Kolika je vjerojatnost da će u sljedećem pokušaju:*

a) promašiti,

b) pogoditi?

Rješenje:

a) Označimo s A događaj "Ante je promašio". Tada je $\mathbb{P}(A) = \frac{120}{300} = 0.4$. Vjerojatnost da će Ante promašiti u sljedećem pokušaju jednaka je 0.4.

b) Označimo s B događaj "Ante je pogodio". Događaj B suprotan je događaju A. Tada je $\mathbb{P}(B) = 1 - 0.4 = 0.6$. Vjerojatnost da će Ante pogoditi u sljedećem pokušaju jednaka je 0.6.

U ovom primjeru, vjerojatnost da će Ante promašiti ili pogoditi koš u sljedećem pokušaju temelji se na prethodnim rezultatima njegovih pokušaja (120 promašaja od 300 pokušaja) i služe nam kao empirijski dokazi koji nam omogućuju procjenu vjerojatnosti budućih ishoda. Koncept statističke stabilnosti relativnih frekvencija je ključan za razumijevanje ovog primjera. Promatranjem velikog broja pokušaja (300 pokušaja), relativne frekvencije promašaja (0.4) i pogodaka (0.6) stabilizirale su se dovoljno da ih možemo smatrati pouzdanim procjenama stvarnih vjerojatnosti odnosno iako su pojedinačni ishodi pokušaja podložni varijacijama, ukupni obrazac koji se pojavljuje nakon velikog broja promatranja odražava prirodu ishoda.

Vjerojatnost *a priori*

Vjerojatnost *a priori* predstavlja drugi pristup računanja vjerojatnosti, koji se oslanja na prethodno poznate informacije i logičku analizu, umjesto na eksperimentalne podatke. Ovim načinom vjerojatnost određujemo temeljem osnovnih pravila i simetrija, bez potrebe za provođenjem pokusa.

Definicija 1.2.4 (Vjerojatnost *a priori*). *Ako su svi elementarni događaji jednako izgledni, vjerojatnost događaja A jednaka je omjeru broja povoljnih ishoda za A i broja svih mogućih ishoda*

$$\mathbb{P}(A) = \frac{k(A)}{k(\Omega)} = \frac{\textit{broj povoljnih elementarnih događaja } A}{\textit{broj svih mogućih elementarnih događaja}}.$$

Vjerojatnost *a priori* (lat. ”od ranije”) temelji se na pravilnosti koju uočavamo iz uvjeta pokusa prije njegovog izvođenja (bez iskustva), a određuje se kao omjer broja mogućnosti koje ostvaruju neki događaj i broja svih jednako mogućih ishoda. Ovaj pristup je posebno koristan u situacijama kada su eksperimenti skupi, nemogući ili kada imamo dovoljno teorijskog znanja koje nam omogućava pouzdane procjene. Najjednostavniji primjer vjerojatnosti *a priori* je bacanje simetričnog novčića. Simetrija novčića se odnosi na vjerojatnost ishoda odnosno kažemo da je novčić simetričan ako je vjerojatnost da će pasti pismo jednaka vjerojatnosti da će pasti glava. Bez ikakvih eksperimentalnih podataka, možemo zaključiti da su šanse za ishod ”glava” ili ”pismo” jednake i iznose 0.5 na temelju simetrije novčića. Pogledajmo neke primjere:

Primjer 1.2.5. *[7] Snop karata sadrži 52 karte u četiri boje: karo, srce, pik i tref. Svaka boja ima po 13 karata. Kolika je vjerojatnost da izvučena karta bude karo ili srce?*

Rješenje: Karata karo ima 13, karata srce ima 13. To je 26 povoljnih elementarnih događaja. Označimo događaj s $KS = \{$izvučena je karta karo ili srce$\}$. Pišemo:

$$\mathbb{P}(KS) = \frac{26}{52} = \frac{1}{2}.$$

Primjer 1.2.6. *[6] Koji su elementarni ishodi kod bacanja dviju kocki? Kolika je vjerojatnost svakoga od elementarnih ishoda? Kolika je vjerojatnost da zbroj na obje kocke bude 9?*

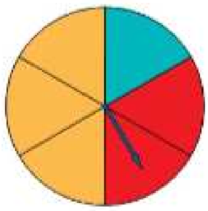
Rješenje: Mogući ishodi bacanja dviju kocki jesu svi mogući parovi brojeva od 1 do 6. Prostor elementarnih događaja Ω određen je s uređenim parovima (a, b), gdje a predstavlja broj koji je pao na prvoj kocki, a b broj koji je pao na drugoj kocki. Na primjer, elementarni događaj ”na prvoj kocki je 5, a na drugoj je 6” predstavljen je parom $(5, 6)$. Elementarnih događaja je 36 i svaki od njih je jednako vjerojatan, pa je vjerojatnost svakog od njih jednaka $\frac{1}{36}$. Elementarni događaji povoljni da zbroj na obje kocke bude 9 su: $(3, 6), (4, 5), (5, 4), (6, 3)$. Dakle, vjerojatnost tog događaja je $\frac{4}{36} = \frac{1}{9}$ tj. $\mathbb{P}(\text{zbroj je } 9) = \frac{1}{9}$.

	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Slika 1.1: Prostor elementarnih događaja bacanja dviju kocki s označenim parovima koji daju zbroj 9

Laplaceov model vjerojatnosti zasniva se na pretpostavci da su svi elementarni događaji jednako izgledni, odnosno da imaju istu vjerojatnost pojavljivanja. Ovo je ključno kod definicije vjerojatnosti *a priori* gdje svaki elementarni događaj dobiva istu šansu za realizaciju. Takav pristup je jednostavan i efektivan u mnogim situacijama, naročito kada imamo posla s jasnim, simetričnim scenarijima, poput bacanja novčića ili kocke. Međutim, važno je naglasiti da ovaj model nije primjenjiv na sve slučajeve. Postoje situacije gdje nije realno očekivati da su svi elementarni događaji jednako vjerojatni. Primjeri iz stvarnog života često sadrže asimetrije ili različite uvjete koji utječu na izglede pojedinih ishoda, zbog čega je Laplaceov model u takvim slučajevima nedovoljan. Ovaj nedostatak najbolje je ilustrirati kroz sljedeći primjer.

Primjer 1.2.7. *[17] Na slučajan način zavrtimo kolo sreće prikazano na slici 1.2. Odredimo skup elementarnih događaja i vjerojatnosti elementarnih događaja. Odredimo vjerojatnost događaja $A = \{$strelica se nije zaustavila na plavom polju$\}$.*



Slika 1.2: Kolo sreće

Rješenje: Strelica će se zaustaviti na žutom, plavom ili crvenom polju. Skup svih elementarnih događaja je $\Omega = \{\text{Ž}, \text{P}, \text{C}\}$, a $A = \Omega \backslash \{\text{P}\} = \{\text{Ž}, \text{C}\}$. Ako bi sva polja imala istu površinu (jednaka polja), tada bi koristili Laplaceov model, gdje su vjerojatnosti svih elementarnih događaja jednake. Međutim, u ovom slučaju imamo polja različitih veličina, što znači da vjerojatnosti nisu jednake, što predstavlja odmak od Laplaceovog modela gdje je svaki ishod podjednako vjerojatan. Žuta boja prekriva polovinu, plava šestinu, a crvena trećinu kruga. Stoga, vrijedi:

$$\mathbb{P}(\text{Ž}) = \frac{1}{2}, \quad \mathbb{P}(\text{P}) = \frac{1}{6}, \quad \mathbb{P}(\text{C}) = \frac{1}{3}.$$

Žutom ili crvenom bojom prekriveno je pet šestina površine, pa je $\mathbb{P}(A) = \frac{5}{6}$.
Uočimo da smo morali računati ovako:

$$\mathbb{P}(A) = \mathbb{P}(\{\text{Ž}, \text{C}\}) = \mathbb{P}(\text{Ž}) + \mathbb{P}(\text{C}) = \frac{3}{6} + \frac{2}{6} = \frac{5}{6}.$$

Geometrijska vjerojatnost

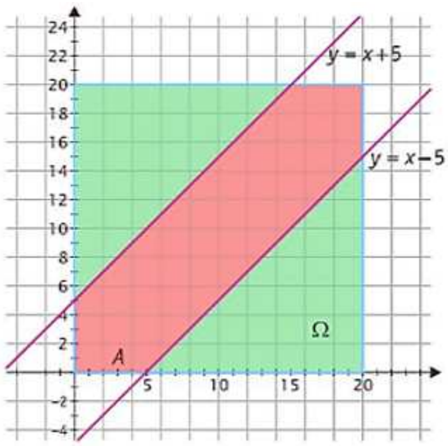
Geometrijska vjerojatnost pojavila se kao prirodni nastavak klasične teorije vjerojatnosti kako bi se odgovorilo na probleme i situacije gdje se vjerojatnost ne može lako odrediti brojanjem povoljnih i mogućih ishoda. ”Mjeri” vjerojatnost događaja pomoću omjera geometrijskih mjera, kao što su duljina, površina ili volumen. Pri računanju geometrijske vjerojatnosti, skup elementarnih događaja biramo kao skup točaka na pravcu, u ravnini ili prostoru.

Definicija 1.2.8. *Neka je skup elementarnih događaja Ω ograničen podskup od $\mathbb{R}^n$ koji je izmjeriv. Geometrijska vjerojatnost proizvoljnog izmjerivog podskupa $A \subseteq \Omega$ je dana formulom:*

$$\mathbb{P}(A) = \frac{m(A)}{m(\Omega)},$$

gdje je m označava geometrijsku mjeru (duljina, površina ili volumen).

Primjer 1.2.9. *[6] Dva prijatelja, Vedran i Bojan, idu svako jutro na posao s iste tramvajske stanice. Svaki od njih stiže na stanicu nasumice u vremenu između 7.00 i 7.20. Voljni su se pričekati 5 minuta, nakon čega idu na tramvaj ili zajedno ili sami. Kolika je vjerojatnost da se prijatelji sretnu na stanici?*



Slika 1.3: Primjer 1.2.9.

Rješenje: Neka je x vrijeme Vedranova, a y vrijeme Bojanova dolaska na stanicu, izraženo u minutama nakon 7 sati. Dakle, $x, y \in [0, 20]$. Skup elementarnih događaja Ω predstavljaju sve točke (x, y) unutar "zelenog" kvadrata prikazanog na slici 1.3. Slijedi: $m(\Omega) = 20 \cdot 20 = 400$. Vedran i Bojan srest će se na stanici ako je razlika u vremenima njhova dolaska manja od 5 minuta odnosno ako je $|x-y| \leq 5$. Dakle, $x-y \leq 5$ i $x-y \geq -5$ tj. $y \geq x-5$ i $y \leq x+5$. Pravci $y = x+5$ i $y = x-5$ također su prikazani na slici 1.3. "Crveni" presjek područja između pravaca označimo s A. Skup A predstavlja povoljne ishode za događaj *"prijatelji će se sresti na stanici"*. Površina skupa A razlika je površine skupa Ω i površine dvaju pravokutnih jednakokračnih trokuta. Slijedi: $m(A) = 400 - 2 \cdot \frac{15 \cdot 15}{2} = 175$. Dakle,

$$\mathbb{P}(A) = \frac{m(A)}{m(\Omega)} = \frac{175}{400} = 0.4375.$$

Vjerojatnost da će se prijatelji sresti na stanici je 43.75%.

Važno je napomenuti da u prostoru $\mathbb{R}^n$ postoje podskupovi koji nisu izmjerivi, odnosno za koje nije moguće definirati odgovarajuću geometrijsku mjeru. Jedan od poznatih primjera takvih skupova je Vitalijev skup. Neizmjerivi skupovi su ključni razlog zašto nije dovoljno definirati samo skup Ω i promatrati sve njegove podskupove. Zbog tih ograničenja, u aksiomatskom pristupu vjerojatnosti uvodi se pojam σ-algebre, o čemu će biti više riječi u sljedećem poglavlju.

1.3 Vjerojatnosni prostor

Nakon uvođenja temeljnih pojmova iz područja vjerojatnosti, nastavnik dolazi do uvođenja pojma vjerojatnosnog prostora. U udžbenicima za srednju školu, definicija vjerojatnosti je poprilično općenita, ali često se nadopunjava pomoću svojstava koja za nju vrijede. Ovakav pristup opravdavaju zahtjevi praktičnosti i ekonomičnosti nastave, no s matematičkog gledišta, ovakav pristup ima određenih nedostataka koji stvaraju poteškoće u daljnjem nadograđivanju i proširivanju gradiva. Pojednostavljena definicija može dovesti do površnog razumijevanja pojma vjerojatnosti, što kasnije može otežati razumijevanje složenijih koncepata u naprednijem matematičkom obrazovanju. Neki od nedostataka su:

- **Nedostatak formalnosti:** Srednjoškolske definicije vjerojatnosti ne uključuju formalne matematičke pojmove poput $\sigma-$algebre.
- **Veze s drugim matematičkim područjima:** Definicije i svojstva vjerojatnosti koje se uče u srednjoj školi često su izolirane i nisu povezane s drugim matematičkim područjima poput statistike, teorije skupova ili matematičke analize.
- **Problemi s interpretacijom i primjenom:** U srednjoj školi, vjerojatnost se često predaje kroz konkretne primjere poput bacanja kockica ili izvlačenja karata, što može dovesti do pogrešnih interpretacija u složenijim situacijama. Na primjer, razumijevanje uvjetne vjerojatnosti i nezavisnosti može biti teško bez čvrstog temelja u formalnim definicijama.
- **Ignoriranje distribucija i funkcija gustoće:** U srednjoj školi često se izbjegavaju detalji o raspodjelama i funkcijama gustoće, što može otežati razumijevanje koncepata normalne distribucije.

Stoga, važno je da nastavnici prilikom objašnjavanja vjerojatnosnog prostora i vjerojatnosti naglase i dublje matematičke aspekte i logiku koja stoji iza tih pojmova. Trebalo bi s učenicima ponoviti i naglasiti razliku između vjerojatnosti *a priori* i vjerojatnosti *a posteriori*. Također, izgradnja razumijevanja može se postići kroz simetriju, kao što je primjer ravnomjernog rasporeda na kocki, nakon čega se mogu uvesti složeniji primjeri i primjeri iz stvarnog života, poput procjene rizika u medicinskim testovima ili predviđanja vremenskih uvjeta.

Definicija 1.3.1. *[3] Familija $\mathcal{F}$ podskupova od Ω ($\mathcal{F} \subset \mathcal{P}(\Omega)$) jest $\sigma-$* ***algebra skupova*** *(na Ω) ako je*

F1. $\emptyset \in \mathcal{F}$,

F2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$,

F3. $A_i \in \mathcal{F},\ i \in \mathbb{N} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Definicija 1.3.2. *[3] Neka je $\mathcal{F}$ $\sigma-$algebra na skupu Ω. Uređen par $(\Omega, \mathcal{F})$ zove se* ***izmjeriv prostor****.*

Definicija 1.3.3. *[3] Neka je $(\Omega, \mathcal{F})$ izmjeriv prostor. Funkcija $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ jest* ***vjerojatnost*** *(na $\mathcal{F}$, na Ω) ako vrijedi*

P1. $\mathbb{P}(A) \geq 0,\ A \in \mathcal{F}$,

P2. $\mathbb{P}(\Omega) = 1$,

P3. $A \in \mathcal{F},\ i \in \mathbb{N}$ *i* $A_i \cap A_j = \emptyset$ *za* $i \neq j$.
(A_i - međusobno disjunktni) $\Rightarrow \mathbb{P}\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Svojstvo $\mathbb{P}(A) \geq 0,\ A \in \mathcal{F}$ jest svojstvo **nenegativnosti vjerojatnosti**, a svojstvo $\mathbb{P}(\Omega) = 1$ jest svojstvo **normiranosti vjerojatnosti**. Svojstvo iz aksioma $P3$. jest svojstvo **prebrojive** ili $\sigma-$**aditivnosti** vjerojatnosti. Ukoliko promatramo situaciju u kojoj je zadan konačan ili prebrojiv skup elementarnih događaja $\Omega = \{\omega_1, \omega_2, ..., \omega_i, ...\}, i \in \mathbb{N}$ i vjerojatnost $p_i = \mathbb{P}(\{\omega_i\}) \in [0, 1]$ za svaki element $\omega_i \in \Omega$ za koje vrijedi

$$\sum_{i=1}^{\infty} \mathbb{P}(\omega_i) = 1,$$

tada za svaki podskup $A \subseteq \Omega$ možemo definirati vjerojatnost kao:

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega).$$

Definicija 1.3.4. *[3] Uređena trojka $(\Omega, \mathcal{F}, \mathbb{P})$ gdje je $\mathcal{F}$ $\sigma-$algebra na Ω i $\mathbb{P}$ vjerojatnost na $\mathcal{F}$, zove se* ***vjerojatnosni prostor****.*

Skup Ω naziva se skup elementarnih događaja. Elementarni događaj je svaki jednočlani podskup skupa Ω, a svaki podskup A od Ω zovemo događajem. Kada je Ω konačan ili prebrojiv, a odabrana $\sigma-$algebra je skup svih podskupova od Ω, uključujući prazan skup $\emptyset$ i sam skup Ω , u oznaci $\mathcal{P}(\Omega)$, kažemo da imamo diskretan vjerojatnosni prostor. Drugim riječima, ako je Ω skup ishoda, tada $\mathcal{P}(\Omega)$ predstavlja sve moguće događaje koje možemo definirati s obzirom na taj skup ishoda. Također, korištenjem partitivnog skupa osiguravamo da možemo pravilno definirati operacije unije, presjeka i komplementa događaja, koje su ključne za analizu složenijih događaja. Na primjer, za skup $\Omega = \{glava, pismo\}$ partitivni skup $\mathcal{P}(\Omega)$ bi bio $\{\emptyset, \{glava\}, \{pismo\}, \{glava, pismo\}\}$. Broj $\mathbb{P}(A)$ zovemo vjerojatnost događaja A, za svaki A koji je podskup skupa Ω.

Vjerojatnost aksiomatski zasnivamo pomoću aksioma Kolmogorova koji na precizan i strogo matematički način opisuju funkciju vjerojatnosti. Prije Kolmogorova, kao što smo vidjeli, vjerojatnost se uglavnom temeljila na intuiciji i primjerima, a ne na strogoj matematičkoj definiciji. Postojali su različiti pristupi, ali ti pristupi nisu bili dovoljno općeniti da bi pokrili sve slučajeve ili su bili matematički problematični u određenim situacijama, posebno kada se radilo o beskonačnim skupovima ili složenijim događajima. S obzirom na sve veće zahtjeve različitih grana matematike postalo je očito da je potrebno razviti formalnu teoriju vjerojatnosti koja bi bila konzistentna i matematički rigorozna. Aksiomatski pristup osigurava da su svi rezultati u teoriji vjerojatnosti (poput Zakona velikih brojeva, Centralnog graničnog teorema, itd.) logički dosljedni i izvedeni iz temeljnih aksioma kojima se izbjegavaju kontradikcije i omogućuje pouzdanost rezultata. Kolmogorov je aksiomatizirao vjerojatnost koristeći samo tri jednostavna aksioma.

Neka svojstva vjerojatnosti su dio same definicije i ne treba ih posebno naglašavati, dok neka svojstva proizlaze iz same definicije i često se ističu zbog njihove korisnosti u rješavanju zadataka. Nastavnik bi trebao učenicima objasniti ta svojstva, a ako je moguće i vremenski izvedivo, pružiti i jednostavne primjere ili skice dokaza kako bi učenici stekli dublje razumijevanje ovih pojmova.

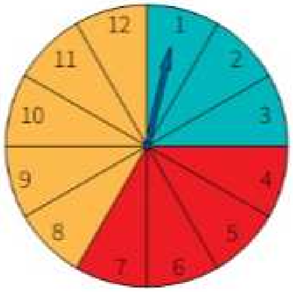
Svojstva vjerojatnosti:

1. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
2. $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$

Primjer 1.3.5. *[17] Na slučajan način zavrtimo kolo sreće prikazano na slici 1.4. Kolika je vjerojatnost događaja $T = \{strelica\ se\ zaustavila\ na\ žutome\ ili\ na\ neparnom\ polju\}$?*

Rješenje: Neka je $A = \{\text{strelica se zaustavila na žutom polju}\}$ i $B = \{\text{strelica se zaustavila na neparnom polju}\}$. Budući da je $T = A \cup B$, vrijedi:

$$\mathbb{P}(T) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \frac{5}{12} + \frac{6}{12} - \frac{2}{12} = \frac{9}{12} = \frac{3}{4} = 0.75.$$



Slika 1.4: Primjer 1.3.5. - Kolo sreće

Poglavlje 2

Slučajne varijable

Teorija bazirana samo na definiciji vjerotnosnog prostora nije dovoljno moćan alat jer ne može adekvatno rješavati složene probleme koji uključuju uvjetne vjerojatnosti, slučajne procese i napredne statističke analize. Potreba za uvođenjem slučajnih varijabli javila se jer one omogućuju kvantifikaciju i analizu različitih svojstava slučajnih fenomena, poput očekivanja, varijance i distribucija. Slučajne varijable pružaju most između apstraktnih događaja i konkretnih numeričkih vrijednosti, čineći analizu složenijih situacija, u statistici, fizici, ekonomiji i drugim disciplinama, praktičnijom i matematički preciznijom. Iz tog razloga, definiramo funkciju koju nazivamo slučajna varijabla. Općenito, slučajna varijabla je funkcija koja elementarnim događajima pridružuje realne brojeve.

Slučajne varijable obzirom na njihovu sliku dijelimo na diskretne i neprekidne slučajne varijable. Diskretna slučajna varijabla X za sliku ima konačan ili prebrojiv skup. Na primjer, X može biti:

- broj kuća u predgrađu koje imaju dvije garaže,
- broj novih bicikala koje svake godine proda specijalizirana trgovina,
- broj neispravnih žarulja u narudžbenici trgovine...

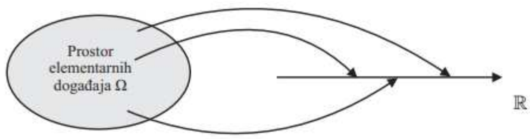
Neprekidna slučajna varijabla X za sliku ima neprebrojiv skup, stoga često govorimo o vrijednosti takve slučajne varijable u nekom intervalu. Na primjer, X može biti:

- visina jednog muškaraca koji je visok između 185 cm i 186 cm,
- volumen vode u spremniku za kišnicu koji može sadržavati do 100 m^3...

Za razliku od diskretnih slučajnih varijabli, kod neprekidnih slučajnih varijabli, svaka pojedinačna vrijednost iz tog intervala ima vjerojatnost 0. Zbog toga, vjerojatnost kod neprekidnih slučajnih varijabli izračunava se za intervale. Na primjer, možemo izračunati vjerojatnost da slučajna varijabla X padne unutar određenog intervala $[a, b]$ tj. $\mathbb{P}(a \leq X \leq b)$. Oznaka $(a \leq X \leq b)$ označava skup svih elementarnih događaja za koje X poprima vrijednosti između a i b, tj. to je skup $\{\omega \in \Omega : a \leq X(\omega) \leq b\} = X^{-1}([a, b])$. Za takve skupove želimo imati definiranu vjerojatnost, tj. oni bi trebali biti elementi σ-algebre vjerojatnosnog prostora na kojem je definirana slučajna varijabla X. Ovaj pristup temelji se na funkciji raspodjele i gustoći vjerojatnosti, pri čemu vjerojatnost bilo kojeg intervala nije nula, iako su vjerojatnosti pojedinačnih točaka u tom intervalu nula. Više o tome bit će objašnjeno u nastavku rada.

U slučaju kada je Ω diskretan, u smislu da je Ω konačan ili prebrojiv, a $\mathcal{F} = \mathcal{P}(\Omega)$, X zaista može biti bilo koja funkcija sa Ω u $\mathbb{R}$. S druge strane, kada je Ω neprebrojiv pokazuje se da svih mogućih funkcija sa Ω u $\mathbb{R}$ ima "previše", što je usko povezano s činjenicom da u slučaju neprebrojivog Ω za prostor događaja ne možemo uzeti $\mathcal{P}(\Omega)$. Da bi razrješili ovaj problem, postavljamo dodatan uvjet na X, a to je "izmjerivost" obzirom na $\mathcal{F}$ ($\mathcal{F}$ je manji od $\mathcal{P}(\Omega)$). [11]

Definicija 2.0.1. *[11] Neka je $(\Omega, \mathcal{F}, \mathbb{P})$ vjerojatnosni prostor.* ***Slučajna varijabla*** *je izmjeriva funkcija $X : \Omega \to \mathbb{R}$, tj. takva funkcija da je $X^{-1}(\langle -\infty, a \rangle) \in \mathcal{F}$ za svaki realni broj a.*



Slika 2.1: Prikaz djelovanja slučajne varijable [12]

2.1 Primjeri diskretnih slučajnih varijabli

U ovom dijelu prikazujemo nekoliko primjera diskretnih slučajnih varijabli.

Primjer 2.1.1. *[11] U mjestu s 2000 obitelji ispitano je koliko koja obitelj posjeduje mobitela. Dobiveni su sljedeći rezultati (prikazani u Tablici 2.1). Odredimo vjerojatnost da slučjano odabrana obitelj ima točno dva mobitela.*

Rješenje: Neka slučajna varijabla U svakom elementarnom ishodu pridružuje broj mobitela u obitelji. Tada, slučajna varijabla U može poprimiti vrijednosti 0, 1, 2, 3 ili 4. Traženi podatak možemo iščitati iz priložene tablice. Zapisujemo:

$$\mathbb{P}(U = 2) = 0.425.$$

broj mobitela	frekvencija	relativna frekvencija
0	30	0.015
1	470	0.235
2	850	0.425
3	490	0.245
4	160	0.08

Tablica 2.1: Frekvencije obitelji s različitim brojem mobitela - primjer 2.1.1.

Primjer 2.1.2. *Promatrajmo pokus bacanja dva novčića. Slučajna varijabla X svakom elementarnom ishodu pridružuje broj pisama. Odredimo moguće vrijednosti slučajne varijable X i vjerojatnosti da slučajna varijabla poprimi te vrijednosti.*

Rješenje: Razmotrimo slučajnu varijablu X koja predstavlja broj pisama koja se pojavljuju u dva bacanja novčića. Elementarni događaji pri bacanju novčića, gdje novčić može pasti na glavu G ili pismo P, su svi mogući događaji koji se mogu dogoditi u svakom bacanju. Ako razmatramo dva bacanja novčića, elementarni događaji su svi mogući parovi rezultata oba bacanja:

- (G, G) - oba bacanja daju glavu,
- (G, P) - prvo bacanje daje glavu, a drugo pismo,
- (P, G) - prvo bacanje daje pismo, a drugo glavu,
- (P, P) - oba bacanja daju pismo.

Dakle, u ovom slučaju, elementarni događaji su $(G, G), (G, P), (P, G), (P, P)$.
Slučajna varijabla X može poprimiti sljedeće vrijednosti: 0, 1 ili 2. Vrijednost 0 poprimit će ako padnu dvije glave i nijedno pismo, slučaj (G, G). To znači da nijedno bacanje ne rezultira pismom. S obzirom na to da postoje ukupno četiri moguće kombinacije elementarnih događaja, a samo jedan od njih odgovara situaciji kada $X = 0$ odnosno da nije palo nijedno pismo, vjerojatnost da X poprimiti vrijednost 0 je $\frac{1}{4}$. Zapisujemo $\mathbb{P}(X = 0) = \frac{1}{4}$ i čitamo *"vjerojatnost da X poprimi vrijednost 0 je $\frac{1}{4}$"*. Vrijednost 1 poprimit će ako pade jedna glava i jedno pismo, odnosno jedno pismo i jedna glava. To uključuje kombinacije (G, P) ili (P, G). Od ukupno četiri moguće kombinacije, dva slučaja odgovaraju situaciji kada $X = 1$ odnosno kada će se u jednom bacanju pojaviti pismo. Stoga, vjerojatnost da X poprimiti vrijednost 1 je $\frac{2}{4} = \frac{1}{2}$. Zapisujemo $\mathbb{P}(X = 1) = \frac{1}{2}$ i čitamo *"vjerojatnost da X poprimi vrijednost 1 je $\frac{1}{2}$"*. Analogno, za vrijednost 2. Vrijednost 2 poprimit će ako padnu dva pisma i nijedna glava, slučaj (P, P). To znači da su oba bacanja rezultirala pismom. S obzirom na to da postoje ukupno četiri moguće kombinacije ishodnih događaja, a samo

jedan od njih odgovara situaciji kada $X = 2$ odnosno da je palo pismo na oba novčića, vjerojatnost da se to dogodi je $\frac{1}{4}$. Zapisujemo $\mathbb{P}(X = 2) = \frac{1}{4}$ i čitamo *"vjerojatnost da X poprimi vrijednost 2 je* $\frac{1}{4}$*"*.

Primjer 2.1.3. *Promatrajmo pokus bacanja dvije simetrične kocke. Slučajna varijabla Y svakom elementarnom događaju pridružuje broj pojavljivanja broja 5. Odredimo vjerojatnosti slučajne varijable Y i vjerojatnosti da slučajna varijabla poprimi te vrijednosti.*

Rješenje: Razmotrimo pokus bacanja simetrične kocke dva puta i definirajmo slučajnu varijablu Y koja svakom elementarnom ishodu pridružuje broj koliko puta se na kocki pojavljuje broj 5. Želimo odrediti vjerojatnosti za različite vrijednosti slučajne varijable Y. U ovom pokusu, bacamo kocku dva puta, pa su svi mogući elementarni događaji kombinacije rezultata oba bacanja, ukupno $6 \cdot 6 = 36$ elemenata. Slučajna varijabla Y može poprimiti vrijednosti 0, 1 ili 2.
Slučajna varijabla Y pridružuje vrijednost 0 svakom elementarnom ishodu u kojem se broj 5 ne pojavljuje niti jednom u oba bacanja. To znači da niti jedan od oba bacanja ne rezultira brojem 5. Svaki od 6 mogućih rezultata bacanja kocke može biti bilo koji od brojeva 1, 2, 3, 4, 5 ili 6. Ako ne želimo da se na kocki pojavi broj 5, onda moramo izabrati rezultate između 1 i 4 ili 6 za oba bacanja. Dakle, broj mogućih rezultata za jedno bacanje, bez broja 5, je 5 (brojevi 1, 2, 3, 4, 6). Za dva bacanja, broj mogućih kombinacija koje ne uključuju broj 5 je: $5 \cdot 5 = 25$. Stoga, vjerojatnost da se broj 5 ne pojavi niti jednom u oba bacanja tj. za $Y = 0$ je omjer broja povoljnih slučajeva i ukupnog broja slučajeva:

$$\mathbb{P}(Y = 0) = \frac{\text{broj povoljnih slučajeva}}{\text{ukupan broj slučajeva}} = \frac{25}{36}.$$

Vrijednost 1 može se dogoditi ako u prvom bacanju padne broj 5, a u drugom bacanju bilo koji drugi broj ili obratno odnosno da bismo dobili $Y = 1$, trebamo odabrati jedno od dva bacanja koje će rezultirati brojem 5, a drugo bacanje ne smije dati broj 5. Izračunajmo vjerojatnosti za oba slučaja. Vjerojatnost da se broj 5 pojavljuje na prvom bacanju, a ne pojavljuje na drugom jednaka je $\frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36}$, dok je vjerojatnost da se broj 5 pojavljuje na drugom bacanju, a ne pojavljuje na prvom jednaka $\frac{5}{6} \cdot \frac{1}{6} = \frac{5}{36}$.

Ova dva načina (broj 5 u prvom bacanju ili broj 5 u drugom bacanju) su disjunktni događaji. Ako se broj 5 pojavi u prvom bacanju, ne može se pojaviti u drugom i obrnuto. Prema pravilu zbrajanja vjerojatnosti za disjunktne događaje, ako imamo dva ili više disjunktnih događaja A i B, vjerojatnost da se dogodi barem jedan od njih je zbroj njihovih vjerojatnosti. Stoga, ukupna vjerojatnost da se broj 5 pojavi točno jednom je

$$\mathbb{P}(Y = 1) = \frac{5}{36} + \frac{5}{36} = \frac{10}{36} = \frac{5}{18}.$$

Slučajna varijabla Y će poprimiti vrijednost 2 ako u oba bacanja padne broj 5. Postoji samo jedan ishod koji zadovoljava uvjet da broj 5 bude prisutan u oba bacanja, a to je slučaj $(5, 5)$. Dakle, broj povoljnih slučajeva je 1. Vjerojatnost da se to dogodi u jednom bacanju je $\frac{1}{6}$, pa je vjerojatnost da se broj 5 pojavi u oba bacanja jednaka

$$\mathbb{P}(Y = 2) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}.$$

Razdioba diskretne slučajne varijable

Distribucija diskretne slučajne varijable može biti prikazana:

- u obliku tablice,
- u grafičkom obliku,
- u funkcijskom obliku kao funkcija distribucije vjerojatnosti ili funkcija raspodjele vjerojatnosti $\mathbb{P}(x) = \mathbb{P}(X = x)$.

Svaku diskretnu slučajnu varijablu tablično možemo zapisati s

$$\begin{pmatrix} x_1 & x_2 & \dots \\ p_1 & p_2 & \dots \end{pmatrix}$$

i tu tablicu nazivamo raspodjela ili distribucija od X i pišemo

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots \\ p_1 & p_2 & \dots \end{pmatrix}.$$

Zapišimo raspodjele (distribucije) za prethodno navedene primjere:

Primjer 2.1.1: [11] $\quad U \sim \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 0.015 & 0.235 & 0.425 & 0.245 & 0.08 \end{pmatrix}.$

Primjer 2.1.2: $\quad X \sim \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}.$

Primjer 2.1.3: $\quad Y \sim \begin{pmatrix} 0 & 1 & 2 \\ \frac{25}{36} & \frac{5}{18} & \frac{1}{36} \end{pmatrix}.$

Za svaku slučajnu varijablu postoji odgovarajuća razdioba ili distribucija vjerojatnosti koja opisuje vjerojatnost da će varijabla poprimiti bilo koju vrijednost. Promatrajući prethodne primjera možemo uočiti da pomoću slučajnih varijabli možemo opisati vjerojatnosti za postizanje svih mogućih vrijednosti slučajne varijable vezane uz slučajan pokus. Dakle, za svaku diskretnu slučajnu varijablu možemo istaknuti dva bitna niza brojeva. Jedan čine sve vrijednosti koje slučajna varijabla X može primiti, tj. skup $R(X) = \{x_i : i \in \mathbb{N}\}$ kojega zovemo slika slučajne varijable, a drugi je niz pripadnih vjerojatnosti za koje vrijedi $p_i = \mathbb{P}(X = x_i) = \mathbb{P}(\omega \in \Omega : X(\omega) = x_i)$, za $x_i \in R(X)$. Dodatno, možemo uočimo da za elemente od $R(X)$ vrijedi $x_i \neq x_j$ za $i \neq j$ s obzirom da je to skup, dok u nizu $(p_i, i \in \mathbb{N})$ može biti istih elemenata, ali on ima druga dva bitna svojstva:

1. $0 \leq p_i \leq 1$, za svaki $i \in \mathbb{N}$,

2. $\sum_{i=1}^{\infty} p_i = 1$.

Kada razmatramo diskretne slučajne varijable, često promatramo intervale u kojima varijabla može poprimiti vrijednost. Stoga, potrebno je biti pažljiv oko toga jesu li krajnje točke intervala uključene. Razmotrimo neke primjere:

Notacija	Opis
$\mathbb{P}(X \leq 3)$	vjerojatnost da je X najviše 3
$\mathbb{P}(3 < X \leq 7)$	vjerojatnost da je X više od 3, ali najviše 7

Distribuciju diskretne slučajne varijable grafički prikazujemo grafom funkcije koji se sastoji od točaka (x_i, p_i). Također, za grafički prikaz distribucije diskretne slučajne varijable često koristimo stupčasti dijagram. Stupčasti dijagram sastoji se od niza pravokutnika iste širine, svaki za jednu vrijednost $x_i \in R(X)$ visine p_i.

Primjer 2.1.4. *[10] Trgovina časopisima zabilježila je broj kupljenih časopisa od strane svojih kupaca u jednom tjednu. 23% kupaca kupilo je jedan časopis, 38% kupilo je dva časopisa, 21% kupilo je tri časopisa, 13% kupilo je četiri časopisa, a 5% kupilo je pet časopisa. Slučajna varijabla X predstavlja broj kupljenih časopisa po kupcu.*

a. Prikažimo razdiobu X tablicom.

b. Prikažimo razdiobu X grafički.

c. Odredimo $\mathbb{P}(1 \leq X < 3)$.

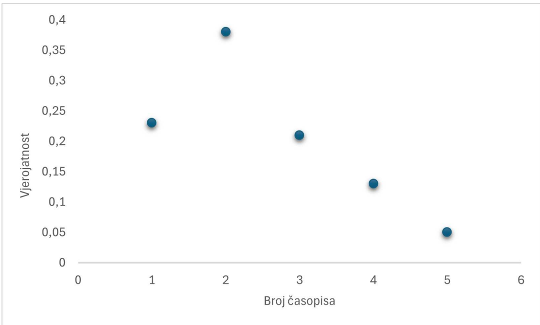
Rješenje:

a. Razdioba tablicom: Tablica 2.2.

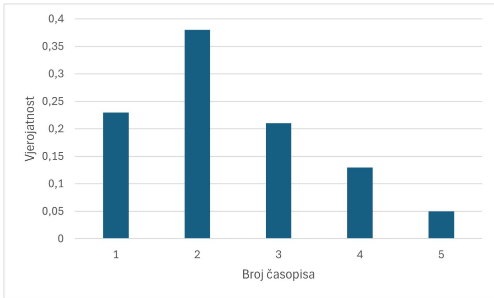
b. Razdioba grafički: Slika 2.2., Slika 2.3.

Broj časopisa (X)	1	2	3	4	5
Vjerojatnost ($\mathbb{P}(X)$)	0.23	0.38	0.21	0.13	0.05

Tablica 2.2: Tablica vjerojatnosti broja kupljenih časopisa



Slika 2.2: Grafički prikaz razdiobe



Slika 2.3: Grafički prikaz razdiobe - stupčasti dijagram

c. Da bismo odredili $\mathbb{P}(1 \leq X < 3)$ trebamo pronaći vjerojatnost da slučajna varijabla X bude u intervalu od 1 do 3, uključujući 1, ali ne uključujući 3. Prema prethodno navedenoj tablici vjerojatnosti:

- vjerojatnost da $X = 1$ je 0.23,
- vjerojatnost da $X = 2$ je 0.38.

S obzirom da X može imati vrijednosti 1 ili 2 za interval $1 \leq X < 3$, trebamo zbrojiti vjerojatnosti za ove vrijednosti:

$$\mathbb{P}(1 \leq X < 3) = \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = 0.23 + 0.38 = 0.61.$$

Funkcija distribucije diskretne slučajne varijable

Definicija 2.1.5. *[12] Neka je dana diskretna slučajna varijabla*

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix},$$

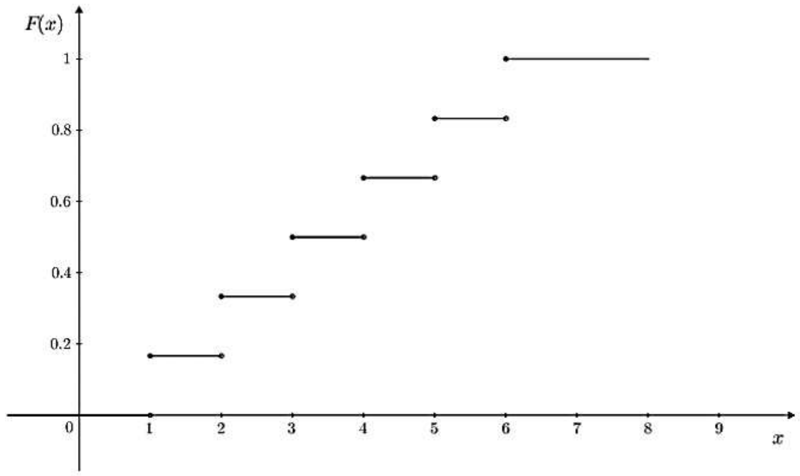
tj. $\mathcal{R}(X) = \{x_i, i \in I\}, I \subseteq \mathbb{N}$, *je dani skup vrijednosti slučajne varijable* X, *a* $(p_i, i \in I)$ *niz pripadnih vjerojatnosti za koji vrijedi*

$$p_i = \mathbb{P}(X = x_i) \geq 0, \forall i \in I, \sum_{i \in I} p_i = 1.$$

Za svaki $x \in \mathbb{R}$ definiramo ***funkciju distribucije*** *$F : \mathbb{R} \to \mathbb{R}$ formulom:*

$$F(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} p_i.$$

Takva funkcija distribucije je stepenasta funkcija tj. funkcija ima skokove u točkama x_i za visine p_i koje imaju istu vrijednost jednaku $F(x_i)$. U nastavku je prikazan graf funkcije distribucije za broj dobiven bacanjem simetrične kocke.



Slika 2.4: Funkcija distribucije za simetričnu kocku [11]

Očekivanje slučajne varijable

Slučajna varijabla predstavlja mjerenje s nepredvidivim ishodom, tj. jedno mjerenje prikazujemo kao funkciju koja može poprimiti više vrijednosti s različitim vjerojatnostima. U primjenama želimo pronaći pojedine parametre koji će nam sažeto opisati varijabilnost ishoda. Zato svakoj slučajnoj varijabli pridružujemo neke karakteristike koje su predvidljive i koje možemo izračunati, a koje nam pomažu da lakše radimo s njima.

Pogledajmo primjer bacanja kocke 120 puta. Zanima nas koliko puta možemo očekivati da će rezultat biti "šest". Da bismo odgovorili na pitanje, prvo moramo razmotriti sve moguće ishode bacanja kocke. Mogućnosti su 1, 2, 3, 4, 5 i 6, a svaka od njih je jednako vjerojatna. Stoga, očekivali bismo da će $\frac{1}{6}$ njih biti "šestica". Jedna šestina od 120 je 20, pa očekujemo da će od 120 bacanja kocke njih 20 dati "šesticu". Znači li to da ćemo dobiti 20 šestica ako kocku bacimo 120 puta? Odgovor je ne nužno. Pojam "očekivanje" u teoriji vjerojatnosti odnosi se na prosječni ishod u velikom broju ponovljenih pokusa. To ne znači da ćemo u svakom pojedinačnom eksperimentu dobiti točno taj broj. U stvarnosti, broj šestica može varirati zbog prirodne varijabilnosti u procesu slučajnog bacanja kocke. Ponekad ćemo možda dobiti više od 20 šestica, a ponekad manje.

Primjer 2.1.6. *[5] Mjesta A i B povezana su dvjema cestama. Prvom se može doći za 20 minuta ako nema gužve i za 40 minuta ako je gužva. Jednako je vjerojatno da će biti gužve kao i da neće. Drugom se cestom može doći za 25 minuta ako nema gužve, a ako je gužva, put će trajati 40 minuta. Vjerojatnost da nema gužve je 0.7. Koju cestu treba odabrati?*

Rješenje: Pri odabiru ceste, u obzir ćemo uzeti vrijeme u kojem moramo stići do mjesta B. Ako ondje moramo biti za 20 minuta, odabrat ćemo prvu cestu, ali što ako nam je prihvatljivo i ako stignemo za 25 minuta? U tom slučaju zanima nas povoljnija varijanta. Za prvu cestu možemo izračunati prosječno vrijeme kao aritmetičku sredinu danih vremena (vjerojatnost da će biti gužva, kao i da neće je jednaka) odnosno $\frac{20+40}{2} = \frac{60}{2} = 30$. Za drugu cestu aritmetička sredina nije smislena jer nije jednako vjerojatno da će biti gužve kao i da neće.

Pogledajmo račun u prvom slučaju i pokušajmo ga drugačije interpretirati:

$$\frac{1}{2} \cdot 20 + \frac{1}{2} \cdot 40 = 30.$$

Vjerojatnost da na putu neće biti gužve je $\frac{1}{2}$, a 20 je vrijeme putovanja u tom slučaju. Analogno za drugi pribrojnik, vjerojatnost da će biti gužva je $\frac{1}{2}$, a 40 je vrijeme putovanja u tom slučaju. Iz ovoga vidimo da smo prosječno vrijeme na prvoj cesti izračunali pomoću vjerojatnosti i vremena. Pokušajmo ponoviti postupak za drugu cestu.
Vjerojatnost da na putu neće biti gužve je 0.7, a 25 je vrijeme putovanja u tom slučaju. Vjerojatnost da će biti gužva je 0.3, a 40 je vrijeme putovanja u tom slučaju. Slijedi,

$$0.7 \cdot 25 + 0.3 \cdot 40 = 29.5.$$

Ako 100 puta idemo tim cestama, potrošit ćemo 3000 minuta na prvoj cesti, dok ćemo na drugoj cesti potrošiti 2950 minuta. Dakle, iz ovoga možemo zaključiti da je druga cesta bolja opcija. Brojeve koje smo na ovaj način izračunali nazvat ćemo matematičko očekivanje.

Definicija 2.1.7. *Neka je X diskretna slučajna varijabla s vrijednostima x_i i vjerojatnostnima p_i.* ***Matematičko očekivanje*** *slučajne varijable X je broj $\mathbb{E}(X) = \sum_i x_i p_i$.*

Primjer 2.1.8. *[5] U 4.a razredu od 28 učenika 10 je djevojaka. Na slučajan način biramo četveročlanu delegaciju. Odredimo očekivanje broja djevojaka u delegaciji.*

Rješenje: Slučajna varijabla X opisuje broj djevojaka u četveročlanoj delegaciji. Trebamo odrediti razdiobu varijeble X. Vrijednosti slučajne varijable X su 0, 1, 2, 3 i 4, a pripadne vjerojatnosti su:

$$\mathbb{P}(X=0)=\frac{\binom{18}{4}}{\binom{28}{4}}=\frac{68}{455}$$

$$\mathbb{P}(X=1)=\frac{\binom{10}{1}\binom{18}{3}}{\binom{28}{4}}=\frac{544}{1365}$$

$$\mathbb{P}(X=2)=\frac{\binom{10}{2}\binom{18}{2}}{\binom{28}{4}}=\frac{153}{455}$$

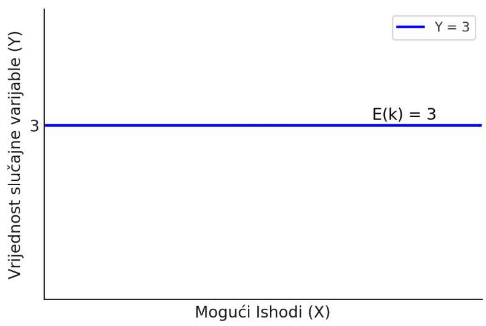
$$\mathbb{P}(X=3)=\frac{\binom{10}{3}\binom{18}{1}}{\binom{28}{4}}=\frac{48}{455}$$

$$\mathbb{P}(X=4)=\frac{\binom{10}{4}}{\binom{28}{4}}=\frac{2}{195}$$

Sada možemo izračunati očekivanje:

$$\mathbb{E}(X)=0\cdot\frac{68}{455}+1\cdot\frac{544}{1365}+2\cdot\frac{153}{455}+3\cdot\frac{48}{455}+4\cdot\frac{2}{195}=\frac{10}{7}\approx 1.43.$$

Napomena 2.1.9. *Ako kažemo da je slučajna varijabla X uvijek jednaka broju k, to znači da ne postoji nikakva nesigurnost u vezi s vrijednošću koju k može poprimiti. Drugim riječima, X je deterministička ili konstantna varijabla:* $\mathbb{P}(X=k)=1, \mathbb{P}(X\neq k)=0$. *Ovakva varijabla ima distribuciju s jednim jedinim stupcem na vrijednosti k, jer je to jedina vrijednost koju varijabla može poprimiti.*



Slika 2.5: Graf konstantne slučajne varijable k

Graf (slika 2.5) prikazuje konstantnu slučajnu varijablu X, koja uvijek poprima istu vrijednost 3. Tu slučajnu varijablu zvat ćemo konstantnom varijablom i ponekad označavati istom oznakom kao i jedinu vrijednost koju poprima. Očekivanje te varijable jednako je njenoj stalnoj vrijednosti, $\mathbb{E}(k) = k$.

Direktno iz definicije slijedi sljedeća propozicija:

Propozicija 2.1.10. *(Svojstva očekivanja) Neka su X i Y diskretne slučajne varijable i k realan broj. Tada je:*

(i) $\mathbb{E}(k) = k$

(ii) $\mathbb{E}(kX) = k\mathbb{E}(X)$

(iii) $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

U srednjoškolskoj nastavi, ova bi se svojstva mogla dokazati ako se želi učenike uvesti u dublje matematičke koncepte, posebno ako su zainteresirani za matematiku ili planiraju daljnje obrazovanje u tom području. Međutim, s obzirom na standardni kurikulum, nastavnik bi vjerojatno mogao ove dokaze spomenuti samo na intuitivnoj razini ili ih prikazati uz jednostavne primjere, a ne ulaziti u formalne dokaze. Tako bi učenici dobili osjećaj za ta svojstva bez potrebe za ulaskom u tehničke detalje.

Očekivanjem se možemo koristiti za procjenu dobitka u nekoj igri. Igra je fer ako je očekivanje dobitka 0. Igrač je na dobitku ako je očekivanje pozitivno, a na gubitku ako je negativno.

Primjer 2.1.11. *[10] U igri na sreću, igrač okreće kotačić označen brojevima 1, 2, 3 i 4. Igrač osvaja iznos novca prikazan u tablici pored, ovisno o broju koji izađe. Odredite:*

a. Očekivanu isplatu za jedno okretanje kotačića.

b. Očekivanu dobit igrača ako svaka igra košta 5 eura.

c. Biste li preporučili igranje ove igre?

Broj	1	2	3	4
Dobitak	1 eur	2 eur	5 eur	8 eur

Rješenje:

a. Neka Y označava isplatu nakon okretanja kotačića. Kako je svaki ishod jednako vjerojatan, vjerojatnost svakog ishoda je $\frac{1}{4}$. Tada je očekivana isplata jednaka:

$$\mathbb{E}(Y) = \frac{1}{4} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 5 + \frac{1}{4} \cdot 8 = 4$$

eura.

b. Neka X označava dobit igrača iz svake igre. Budući da svaka igra košta 5 eura, očekivana dobit jednaka je:

$$\mathbb{E}(X) = \mathbb{E}(Y) - 5 = 4 - 5 = -1$$

euro.

c. Budući da je $\mathbb{E}(X) \neq 0$, igra nije poštena. Konkretno, budući da je $\mathbb{E}(X) = -1$ euro, očekujemo da će igrač izgubiti 1 euro u prosjeku sa svakim okretanjem. Stoga, ne bismo preporučili igranje ove igre.

Varijanca slučajne varijable

Primjer 2.1.12. *Zadane su dvije razdiobe X i Y koje predstvalju ocjene u istoj gimnastičkoj disciplini. Slučajna varijabla X predstavlja ocjenu koju dobiva prva natjecateljica, a slučajna varijabla Y predstavlja ocjenu koju dobiva druga natjecateljica. Obje varijable, X i Y, su definirane na temelju istih kriterija za ocjenjivanje. Odredimo očekivanje.*

$$X \sim \begin{pmatrix} 11 & 13 & 15 \\ 0.2 & 0.2 & 0.6 \end{pmatrix} \qquad Y \sim \begin{pmatrix} 13 & 14 & 15 \\ 0.4 & 0.4 & 0.2 \end{pmatrix}.$$

Rješenje:

$$\mathbb{E}(X) = 11 \cdot 0.2 + 13 \cdot 0.2 + 15 \cdot 0.6 = 2.2 + 2.6 + 9 = 13.8$$

$$\mathbb{E}(Y) = 13 \cdot 0.4 + 14 \cdot 0.4 + 15 \cdot 0.2 = 5.2 + 5.6 + 3 = 13.8$$

Očekivanje ocjena za obje natjecateljice u disciplini je jednako 13.8 bodova. To znači da su prosječne ocjene koje bi natjecateljice mogle postići vrlo slične, što ukazuje na to da u prosjeku imaju sličan nivo izvedbi u ovoj disciplini. Ukoliko promotrimo razdiobe, možemo primjetiti da su vrijednosti koje može poprimiti varijabla Y bliže očekivanju nego što su to vrijednosti varijable X. Iz tog razloga, potreban nam je još jedan podatak koji će mjeriti koliko se vrijednosti razlikuju od očekivanja. Kako bismo izbjegli da se pozitivne i negativne razlike ponište, promatrat ćemo slučajnu varijablu $(X - \mathbb{E}(X))^2$ i računati njezino očekivanje.

Definicija 2.1.13. *Neka je X diskretna slučajna varijabla i* $\mu = \mathbb{E}(X)$. ***Varijanca*** *slučajne varijable X je broj* $\operatorname{Var} X = \mathbb{E}((X - \mu)^2)$.

Definicija 2.1.14. ***Standardna devijacija*** *od X, u oznaci* $\sigma(X)$ *je broj dan formulom*

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

U statistici se često koriste i varijanca i standardna devijacija jer svaka ima svoje prednosti i primjene, a njihovo zajedničko korištenje omogućuje bolje razumijevanje distribucije podataka. Standardna devijacija (σ) je kvadratni korijen varijance i koristi se jer ima iste dimenzije kao i izvorni podaci, pa ju je "lakše" interpretirati u kontekstu problema. To znači da, ako su podaci izraženi u metrima, i standardna devijacija je izražena u metrima. "Mali" $\sigma(X)$ znači da je većina vrijednosti od X grupirana oko očekivanja $\mathbb{E}(X)$, a "veliki" $\sigma(X)$ znači da su vrijednosti od X razvučene preko većeg raspona. Varijanca (σ^2) mjeri prosječnu kvadratnu udaljenost svakog podatka od srednje vrijednosti, ali je izražena u kvadratnim jedinicama (npr. kvadratnim metrima ako su podaci u metrima). Korisna je jer zadržava određena matematička svojstva koja standardna devijacija gubi. Na primjer, varijanca ima linearnost, što znači da je varijanca sume ili razlike nezavisnih slučajnih varijabli zbroj odnosno razlika njihovih varijanci:

$$\mathrm{Var}(X+Y) = \mathrm{Var}\,X + \mathrm{Var}\,Y.$$

Izračunajmo sada varijancu slučajne varijable X i slučajne varijable Y iz prethodnog primjera po definiciji. Trebamo odrediti očekivanje $\mathbb{E}(X)$ i $\mathbb{E}[(X-\mu)^2]$ te $\mathbb{E}(Y)$ i $\mathbb{E}[(Y-\mu)^2]$. Iz prethodnog znamo da je $\mu = \mathbb{E}(X) = \mathbb{E}(Y) = 13.8$. $\mathbb{E}[(X-\mu)^2]$ ćemo izračunati tako da ćemo odrediti $(x_i - \mu)^2$ za svaku vrijednost x_i:

$$(11 - 13.8)^2 = (-2.8)^2 = 7.84$$

$$(13 - 13.8)^2 = (-0.8)^2 = 0.64$$

$$(15 - 13.8)^2 = 1.2^2 = 1.44$$

Sada možemo odrediti očekivanje:

$$\mathbb{E}[(X-\mu)^2] = 7.84 \cdot 0.2 + 0.64 \cdot 0.2 + 1.44 \cdot 0.6 = 1.568 + 0.128 + 0.864 = 2.56.$$

Dakle, $\mathrm{Var}\,X = 2.56$. Postupak ponavljamo kako bi izračunali varijancu Y.

$$(13 - 13.8)^2 = (-0.8)^2 = 0.64$$

$$(14 - 13.8)^2 = 0.2^2 = 0.04$$

$$(15 - 13.8)^2 = 1.2^2 = 1.44$$

Sada možemo odrediti očekivanje:

$$\mathbb{E}[(Y-\mu)^2] = 0.64 \cdot 0.4 + 0.04 \cdot 0.4 + 1.44 \cdot 0.2 = 0.256 + 0.016 + 0.288 = 0.56.$$

Dakle, $\mathrm{Var}\,Y = 0.56$.

Varijanca za prvu natjecateljicu (razdioba X) je 2.56, dok je varijanca za drugu natjecateljicu (razdioba Y) 0.56. Navedeno ukazuje na to da su ocjene prve natjecateljice više raspršene, s većim rasponom ocjena, što znači da su njezini rezultati manje dosljedni. S druge strane, druga natjecateljica ima manju varijancu, što sugerira da su njezini rezultati stabilniji i bliži prosječnoj ocjeni. Na kraju možemo zaključiti da ako su prosječne ocjene za obje natjecateljice u istoj disciplini jednake, varijanca rezultata pokazuje različite nivoe stabilnosti u njihovim performansama.

Formulu za računanje varijance možemo svesti na jednostavniji oblik primijenjujući svojstva očekivanja:

$$\begin{aligned}
\operatorname{Var} X &= \mathbb{E}[(X-\mu)^2] = \mathbb{E}(X^2 - 2X\mu + \mu^2) \\
&= \mathbb{E}(X^2) - \mathbb{E}(2X\mu) + \mathbb{E}(\mu^2) \\
&= \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2 \\
&= \mathbb{E}(X^2) - 2\mu^2 + \mu^2 \\
&= \mathbb{E}(X^2) - \mu^2 \\
&= \mathbb{E}(X^2) - (\mathbb{E}(X))^2
\end{aligned}$$

Propozicija 2.1.15. *Neka je X diskretna slučajna varijabla s očekivanjem* $\mathbb{E}(X)$*. Za varijancu slučajne varijable X vrijedi* $\operatorname{Var} X = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

Izračunajmo varijancu iz prethodog primjera koristeći izvedenu formulu. Znamo da je $\mathbb{E}(X) = \mathbb{E}(Y) = 13.8$ odnosno $(\mathbb{E}(X))^2 = (\mathbb{E}(Y))^2 = 13.8^2 = 190.44$. Izračunajmo $\mathbb{E}(X^2)$ i $\mathbb{E}(Y^2)$.

$$\mathbb{E}(X^2) = 11^2 \cdot 0.2 + 13^2 \cdot 0.2 + 15^2 \cdot 0.6 = 24.2 + 33.8 + 135 = 198.$$

$$\mathbb{E}(Y^2) = 13^2 \cdot 0.4 + 143^2 \cdot 0.4 + 15^2 \cdot 0.2 = 67.6 + 78.4 + 45 = 191.$$

$$\operatorname{Var} X = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = 198 - 190.44 = 2.56.$$

$$\operatorname{Var} Y = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = 191 - 190.44 = 0.56.$$

U nastavku ćemo promotriti neke važne primjere standardnih diskretnih slučajnih varijabli.

Bernoullijeva razdioba

Dosad smo razmatrali svojstva općenitih diskretnih slučajnih varijabli. Sada ćemo istražiti posebnu vrstu diskretne slučajne varijable u kojoj su moguća samo dva ishoda: uspjeh i neuspjeh. Vjerojatnosnu distribuciju povezanu s tom varijablom nazivamo Bernoullijeva vjerojatnosna distribucija. Zamislimo da imamo jednu kovanicu i želimo ju baciti. Svaki put kada bacimo kovanicu, imamo samo dva moguća ishoda: "glava" (uspjeh) ili "pismo" (neuspjeh). Definirajmo Bernoullijevu slučajnu varijablu X kao: $X = 1$ ako padne "glava" i $X = 0$ ako padne "pismo". U ovom slučaju, vjerojatnost da kada dobijemo "glavu" označavamo s p. Tada je vjerojatnost da dobijemo "pismo" jednaka $1 - p$.

Definicija 2.1.16. *Diskretna slučajna varijabla X je Bernoullijeva s parametrom* $0 \leq p \leq 1$ *ako je njezina raspodjela oblika*

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

Bernoullijeva slučajna varijabla može se interpretirati kao ishod pokusa koji može rezultirati samo uspjehom ili neuspjehom.

Propozicija 2.1.17. *Za Bernoullijevu slučajnu varijablu X s vjerojatnošću uspjeha p vrijedi*

$$\mathbb{E}(X) = p, \ \mathrm{Var}\, X = p(1-p).$$

Binomna razdioba

Binomna razdioba jedna je od najvažnijih diskretnih razdioba u teoriji vjerojatnosti i statistici. Ova razdioba opisuje broj uspjeha u konačnom broju nezavisnih pokusa, gdje svaki pokus ima samo dva moguća ishoda – uspjeh ili neuspjeh. Razmatrajući pokuse poput bacanja novčića, odgovaranja na pitanja s dvije moguće opcije (točno/netočno) ili testiranja prisutnosti određene karakteristike kod uzoraka, binomna razdioba pruža matematički okvir za analizu i predviđanje ishoda. Ključna svojstva binomne razdiobe su broj pokusa n i vjerojatnost uspjeha u pojedinom pokusu p. Ova dva parametra omogućuju modeliranje širokog spektra stvarnih situacija, u kojima nas zanima koliko puta će se određeni ishod pojaviti tijekom niza pokušaja. Kako bismo bolje razumjeli binomnu razdiobu, istražit ćemo njezinu definiciju, osnovne karakteristike, te primjene u raznim područjima.

Primjer 2.1.18. *[10] Pretpostavimo da vrtuljak ima tri plava i jedno bijelo polje. Jasno je da ćemo pri svakom okretanju dobiti ili plavo ili bijelo polje. Šansa da se završi na plavom polju je* $\frac{3}{4}$, *a na bijelom je* $\frac{1}{4}$. *Ako plavi rezultat nazovemo 'uspjehom', a bijeli rezultat 'neuspjehom', tada imamo binomni eksperiment. Neka je p vjerojatnost dobivanja plavog polja, dakle* $p = \frac{3}{4}$. *Vjerojatnost dobivanja bijelog polja je* $1 - p = \frac{1}{4}$. *Izračunajte vjerojatnosti za različite brojeve uspjeha (plavih polja) u tri okretaja.*

Slika 2.6: Plavo-bijeli vrtuljak

Rješenje: Neka slučajna varijabla X predstavlja broj 'uspjeha' ili plavih rezultata, dakle X može biti 0, 1, 2 ili 3.

$$\mathbb{P}(X=0)=\mathbb{P}(\text{nijednom nije palo na plavu stranu})=\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}=\left(\frac{1}{4}\right)^3$$

$$\begin{aligned}\mathbb{P}(X=1)&=\mathbb{P}(\text{jednom je palo na plavu stranu i dva puta na bijelu})\\&=\mathbb{P}(\text{PBB, BPB, BBP})=\frac{3}{4}\cdot\left(\frac{1}{4}\right)^2\cdot 3\end{aligned}$$

$$\begin{aligned}\mathbb{P}(X=2)&=\mathbb{P}(\text{dva puta je palo na plavu stranu i jednom na bijelu})\\&=\mathbb{P}(\text{BPP, PBP, PPB})=\left(\frac{3}{4}\right)^2\cdot\frac{1}{4}\cdot 3\end{aligned}$$

$$\mathbb{P}(X=3)=\mathbb{P}(\text{tri puta je palo na plavu stranu})=\left(\frac{3}{4}\right)^3$$

Faktor 3 s kojim smo "dodatno" množili predstavlja broj načina za postizanje jednog uspjeha u tri pokušaja, što je $\binom{3}{1}$. Primijetimo:

$$\mathbb{P}(X=0)=\left(\frac{1}{4}\right)^3=\binom{3}{0}\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)^3\approx 0.0156$$

$$\mathbb{P}(X=1)=3\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^1=\binom{3}{1}\left(\frac{3}{4}\right)^1\left(\frac{1}{4}\right)^2\approx 0.1406$$

$$\mathbb{P}(X=2)=3\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^2=\binom{3}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^1\approx 0.4219$$

$$\mathbb{P}(X=3)=\left(\frac{3}{4}\right)^3=\binom{3}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^0\approx 0.4219$$

Ovo sugerira da je $\mathbb{P}(X=x)=\binom{3}{x}\left(\frac{3}{4}\right)^x\left(\frac{1}{4}\right)^{3-x}$ gdje je $x=0,1,2,3$. Ukoliko zbrojimo vjerojatnosti

$$\mathbb{P}(X=0)+\mathbb{P}(X=1)+\mathbb{P}(X=2)+\mathbb{P}(X=3)=\left(\frac{1}{4}\right)^3+3\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^1+3\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^2+\left(\frac{3}{4}\right)^3$$

dobivamo binomni izraz $\left(\frac{1}{4}+\frac{3}{4}\right)^3=1^3=1$.

Definicija 2.1.19. *Diskretna slučajna varijabla X je binomna s parametrima* $0\le p\le 1$ *i* $n\in\mathbb{N}$*, u oznaci* $X\sim B(n,p)$*, ako je njezina raspodjela oblika*

$$X\sim\begin{pmatrix}0 & 1 & \dots & n\\ p_0 & p_1 & \dots & p_n\end{pmatrix},$$

gdje je

$$p_k=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,1,...,n.$$

Binomna slučajna varijabla predstavlja broj uspjeha $k\in\{0,1,...,n\}$ u n nezavisnih ponavljanja pokusa kod kojeg kao ishod imamo uspjeh s vjerojatnošću p i neuspjeh s vjerojatnošću $(1-p)$. Vjerojatnost bilo kojeg fiksnog niza od k uspjeha i $n-k$ neuspjeha u n nezavisnih ponavljanja pokusa je $p^k(1-p)^{n-k}$, a broj takvih nizova je $\binom{n}{k}$.

Propozicija 2.1.20. *Za binomnu slučajnu varijablu X s n ponavljanja i vjerojatnošću uspjeha p vrijedi*

$$\mathbb{E}(X)=np, \operatorname{Var}X=np(1-p)=npq.$$

Napomena 2.1.21. *Binomna slučajna varijabla s parametrima n=1 i* $0\le p\le 1$*, u oznaci* $X\sim B(1,p)$ *zapravo je Bernoullijeva slučajna varijabla s parametrom p.*

2.2 Neprekidne slučajne varijable

U ovom poglavlju razmotrit ćemo osnovne koncepte vezane uz neprekidne slučajne varijable. Neprekidne slučajne varijable predstavljaju slučajne varijable čije vrijednosti mogu poprimiti bilo koji realan broj unutar određenog intervala, što ih čini izuzetno važnima za modeliranje i analizu pojava koje nisu ograničene na diskretne vrijednosti. Dok su diskretne slučajne varijable povezane s prebrojavanjem (npr. broj djece u obitelji), neprekidne slučajne varijable se obično povezuju s mjerenjem (npr. visina, težina, temperatura). Na kraju, u ovom poglavlju, analizirat ćemo normalnu distribuciju. Normalna distribucija je temeljni primjer razdiobe neprekidne slučajne varijable. Razumijevanje ovih funkcija i njihova primjena omogućuju precizno opisivanje i analizu složenih procesa te donošenje informiranih odluka u različitim područjima.

Jedan od osnovnih pojmova u analizi neprekidnih slučajnih varijabli je funkcija gustoće, koja opisuje koliko je vjerojatno da slučajna varijabla bude u blizini određenog broja. Ova funkcija omogućava nam da izračunamo vjerojatnost da će vrijednost varijable pasti unutar određenog intervala. Na primjer, u analizi distribucije visine ljudi, funkcija gustoće može pomoći u predviđanju vjerojatnosti da će slučajni uzorak osobe imati visinu u određenom rasponu. Uz funkciju gustoće, važan je i koncept funkcije distribucije, koja daje vjerojatnost da slučajna varijabla bude manja ili jednaka određenom broju. Funkcija distribucije je pojam koji definiramo za bilo koju slučajnu varijablu i pomoću kojeg definiramo kad je slučajna varijabla neprekidna. Da bismo shvatili kako vjerojatnosti variraju duž različitih vrijednosti slučajne varijable, trebamo nešto više od samih vrijednosti i njihovih vjerojatnosti. Kao što smo već rekli, za neprekidne slučajne varijable, vjerojatnost da varijabla poprimi točno određenu vrijednost jednaka je nula. Umjesto toga, važna je vjerojatnost da varijabla padne unutar određenog intervala, što je omogućeno funkcijom distribucije. Sada ćemo definirati funkciju distribucije i objasniti njene osnovne karakteristike koje čine ovaj alat ključnim za statističku analizu i modeliranje.

Definicija 2.2.1. *Neka je $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ vjerojatnostni prostor i neka je X slučajna varijabla. Funkciju $F : \mathbb{R} \to [0, 1]$ koja realnom broju x pridružuje vjerojatnost da dana slučajna varijabla bude manja ili jednaka tom broju, tj. funkciju*

$$F(x) = \mathbb{P}(\omega \in \Omega : X(\omega) \leq x) = \mathbb{P}(X \leq x),$$

*zovemo **funkcija distribucije** slučajne varijable X.*

Neprekidne slučajne varijable možemo definirati i kao funkcije kojima je funkcija distribucije neprekidna funkcija. Funkcija $F : \mathbb{R} \to [0, 1]$ ima sljedeća svojstva:

1. $F(x)$ je rastuća funkcija,
2. $F(x) = \mathbb{P}(X < x) = \mathbb{P}(-\infty < X < x) = \int_{-\infty}^{x} f(t)\, dt, x \in \mathbb{R}$,

3. $\lim_{x\to-\infty} F(x) = 0$ i $\lim_{x\to\infty} F(x) = 1$,
4. $\mathbb{P}(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f(x)\,dx = F(x_2) - F(x_1)$.

Specijalno, ako je slučajna varijabla diskretna, funkcija distribucije može se izraziti koristeći tablicu distribucije slučajne varijable, a ako je neprekidna, koristeći pripadnu funkciju gustoće.

Za neprekidne slučajne varijable X, vjerojatnost da X bude točno jednaka određenoj vrijednosti je nula. Dakle, $\mathbb{P}(X = x) = 0$ za svaki x. Na primjer, vjerojatnost da jaje teži točno 72.9 g je nula. Ako bismo vagali jaje na vagu koja mjeri do na točnost od 0.1 g, očitavanje od 72.9 g znači da težina leži negdje između 72.85 g i 72.95 g. Bez obzira na to koliko su precizne vage, možemo samo znati težinu jajeta unutar nekog raspona. Stoga, za neprekidnu varijablu možemo govoriti samo o vjerojatnosti da događaj leži u intervalu.

Primjer 2.2.2. *Pretpostavimo da želimo prošetati uz neku rijeku. Taj put neka je dug oko 2 km. Zbog različite brzine hoda i ostalih subjektivnih utjecaja na trajanje šetnje, jasno je da put koji osoba pritom prijeđe u prvih 5 minuta možemo modelirati kao slučajnu karakteristiku s neprebrojivim skupom mogućih ishoda [0, 2].*

Definicija 2.2.3. *Kažemo da je slučajna varijabla X neprekidna ili kontinuirana ako postoji nenegativna integrabilna funkcija f takva da je vjerojatnost da slučajna varijabla X poprimi vrijednosti između realnih brojeva a i b jednaka:*

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b) = \int_a^b f(x)\,dx.$$

Funkciju f nazivamo ***funkcijom gustoće*** *neprekidne slučajne varijable X.*

Dodatno, za funkciju $f : \mathbb{R} \to \mathbb{R}$ vrijedi svojstvo:

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

Napomena 2.2.4. *[12] Neka je X neprekidna slučajna varijabla s funkcijom gustoće f. Dakle,* $f(x) \geq 0$ *za sve* $x \in \mathbb{R}$, $\int_{-\infty}^{\infty} f(x)\,dx = 1$ *i*

$$\mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\,dt, x \in \mathbb{R}.$$

Iz svojstava integrala očigledno je da je, za razliku od diskretnog slučaja, funkcija distribucije neprekidne slučajne varijable neprekidna funkcija u svakoj točki $x \in \mathbb{R}$. Također, funkcija gustoće neprekidne slučajne varijable X može se dobiti kao derivacija pripadne funkcije distribucije u gotovo svakoj točki.

Napomena 2.2.5. *Funkcija distribucije neprekidne slučajne varijable X s funkcijom gustoće f derivabilna je realna funkcija F takva da je $F'(x) = f(x)$ za svaki x.*

Primjer 2.2.6. *Odredimo funkciju gustoće za neprekidnu slučajnu varijablu koja poprima vrijednosti u intervalu [0,5] i kojoj je vjerojatnost raspoređena proporcionalno udaljenosti od ishodišta. Izračunajmo vjerojatnost da X poprimi vrijednost između 3 i 4.*

Rješenje: Trebamo odredili funkciju gustoće vjerojatnosti za neprekidnu slučajnu varijablu X koja poprima vrijednosti u intervalu $[0, 5]$, a za koju je vjerojatnost raspoređena proporcionalno udaljenosti od ishodišta. Ako je nešto proporcionalno s x, to znači da funkcija koja opisuje tu proporcionalnost ima oblik $f(x) = C \cdot x$, gdje je C konstanta proporcionalnosti. U ovom slučaju, funkcija $f(x)$ je funkcija gustoće vjerojatnosti koja treba zadovoljiti uvjet proporcionalnosti, dok je C konstanta koja osigurava da je funkcija gustoće pravilno normalizirana, tj. da je ukupna vjerojatnost unutar intervala $[0, 5]$ jednaka 1. Konstantu C ćemo odrediti iz uvjeta normalizacije:

$$\int_0^5 f(x)\,dx = 1.$$

Zamjenjujući $f(x)$ s $C \cdot x$ slijedi:

$$\int_0^5 C \cdot x dx = 1.$$

Izračunajmo sada integral:

$$C \cdot \int_0^5 x dx = C \cdot \frac{x^2}{2}\Big|_0^5 = C \cdot \frac{25}{2}.$$

Da bismo odredili C prethodni izraz izjednačavamo s 1:

$$C \cdot \frac{25}{2} = 1,$$

iz čega slijedi da je $C = \frac{2}{25}$. Dakle, funkcija gustoće je $f(x) = \frac{2}{25}x, x \in [0, 5]$. Sada možemo izračunati vjerojatnost da X poprimi vrijednost između 3 i 4:

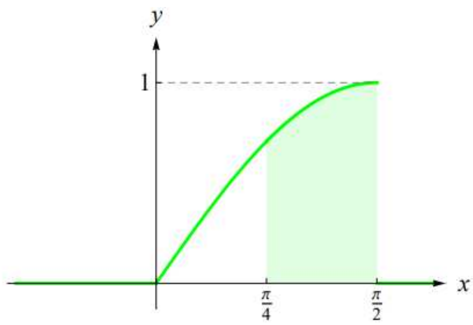
$$\mathbb{P}(3 \leq X \leq 4) = \int_3^4 f(x)\,dx = \int_3^4 \frac{2}{25}x\,dx = \frac{2}{25}\frac{x^2}{2}\Big|_3^4 = \frac{2}{25} \cdot \left(\frac{16}{2} - \frac{9}{2}\right) = \frac{2}{25} \cdot \frac{7}{2} = \frac{14}{50} = \frac{7}{25}.$$

Dakle, vjerojatnost da X poprimi vrijednost između 3 i 4 iznosi $\frac{7}{25}$ ili 0.28.

Primjer 2.2.7. *[12] Analizirajmo može li funkcija $f : \mathbb{R} \to \mathbb{R}$ definirana izrazom*

$$f(x) = \begin{cases} \sin x, & x \in \langle 0, \frac{\pi}{2}] \\ 0, & x \notin \langle 0, \frac{\pi}{2}] \end{cases}$$

poslužiti kao funkcija gustoće neke neprekidne slučajne varijable X. Ako može, odredimo $\mathbb{P}\left(\frac{\pi}{4} < X \leq \frac{\pi}{2}\right)$.



Slika 2.7: Funkcija gustoće slučajne varijable

Rješenje: Iz definicije funkcije f očito se radi o nenegativnoj funkciji:

- na intervalu $\langle 0, \frac{\pi}{2}]$, $f(x) = \sin x$, a funkcija sinus je pozitivna na tom intervalu,
- na $\langle -\infty, 0] \cup \langle \frac{\pi}{2}, \infty\rangle$, $f(x) = 0$.

Pokažimo da je funkcija f normirana:

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{\frac{\pi}{2}} f(x)\,dx + \int_{\frac{\pi}{2}}^{\infty} f(x)\,dx = \int_{0}^{\frac{\pi}{2}} \sin x\,dx = 1.$$

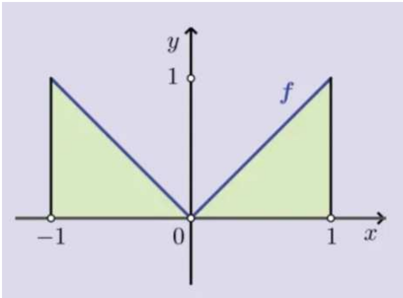
Budući da je funkcija f nenegativna i normirana, zaključujemo da može poslužiti kao funkcija gustoće neprekidne slučajne varijable. Traženu vjerojatnost računamo na sljedeći način:

$$\mathbb{P}\left(\frac{\pi}{4} < X \leq \frac{\pi}{2}\right) = \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \sin x\,dx = \frac{\sqrt{2}}{2}.$$

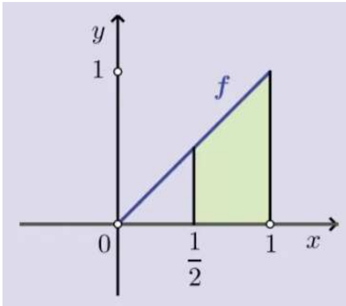
Primjer 2.2.8. *[14] Neka je funkcija gustoće neprekidne slučajne varijable X dana s*

$$f(x) = \begin{cases} |x|, & -1 \leq x \leq 1 \\ 0, & x < -1, x > 1 \end{cases}.$$

Odredimo $\mathbb{P}(-1 \leq X \leq 1), \mathbb{P}(-1 \leq X \leq 0)$ te $\mathbb{P}(0.5 \leq X \leq 1)$. Odredimo funkciju distribucije varijable X.



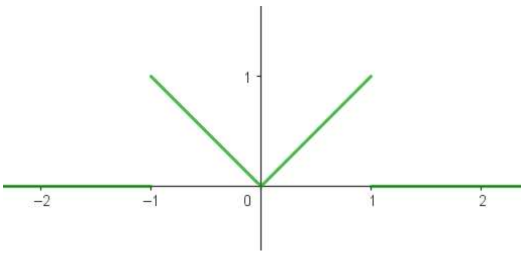
Slika 2.8: $\mathbb{P}(-1 \leq X \leq 1)$



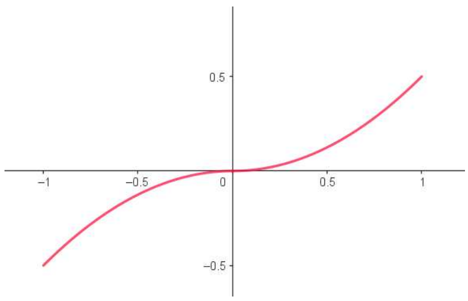
Slika 2.9: $\mathbb{P}(0.5 \leq X \leq 1)$

Rješenje: Vjerojatnost $\mathbb{P}(-1 \leq X \leq 1)$ jednaka je zbroju površina dvaju jednakokračnih pravokutnih trokuta kateta duljine 1, pa je $\mathbb{P}(-1 \leq X \leq 1) = 1$. Vjerojatnost $\mathbb{P}(-1 \leq X \leq 0)$ jednaka je površini jednog takvog trokuta, pa je $\mathbb{P}(-1 \leq X \leq 0) = \frac{1}{2}$. Kako bismo dobili vjerojatnost $\mathbb{P}(0.5 \leq X \leq 1)$, od površine jednokokračnog pravokutnog trokuta katete duljine 1 trebamo oduzeti površinu jednakokračnog pravokutnog trokuta katete duljine 0.5, pa slijedi: $\mathbb{P}(0.5 \leq X \leq 1) = \frac{1}{2} - \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{8}$.

Možemo uzeti da za funkciju distribucije $F(x)$ vrijedi $F(x) = 0$ za $x < -1$ i $x > 1$. Kako je $f(x) = -x$ za $x \in [-1, 0]$, možemo uzeti $F(x) = -\frac{x^2}{2}$ za $x \in [-1, 0]$ te slično $F(x) = \frac{x^2}{2}$ za $x \in [0, 1]$.



Slika 2.10: Funkcija gustoće za primjer 2.2.9.



Slika 2.11: Funkcija distribucije za primjer 2.2.9.

Očekivanje i varijanca neprekidne slučajne varijable

Slično kao i kod diskretnih slučajnih varijabli, uvodimo determinističke karakteristike neprekidnih slučajnih varijabli, u ovom slučaju pomoću funkcije gustoće. Također, sva prethodno spomenuta svojstva očekivanja i varijance za diskretne slučajne varijable vrijede i za neprekidne slučajne varijable.

Definicija 2.2.9. ***Matematičko očekivanje*** *neprekidne slučajne varijable X definiramo kao*

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} t f(t)\, dt.$$

Definicija 2.2.10. ***Varijanca*** *neprekidne slučajne varijable X je*

$$\operatorname{Var} X = \int_{-\infty}^{\infty} (t - \mathbb{E}(X))^2 f(t)\, dt.$$

Definicija 2.2.11. ***Standardna devijacija*** *neprekidne slučajne varijable X je*

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

Primjer 2.2.12. *[12] Neka je X slučajna varijabla čija je vrijednost vrijeme trajanja akumulatora i dana je funkcijom distribucije*

$$F(x) = \mathbb{P}(X \leq x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}x^2, & 0 \leq x < 2 \\ 1, & x > 2. \end{cases}$$

Odredimo funkciju gustoće od X, izračunajmo pripadno očekivanje, varijancu i standardnu devijaciju te odredimo vjerojatnost da se akumulator potrošio u razdoblju od 1.5 godine do 2 godine.

Rješenje: Uočimo, funkcija $F(x)$ nije derivabilna u $x = 2$. Odredimo funkciju gustoće:

$$f(x) = F'(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}x, & 0 \leq x < 2 \\ 0, & x > 2. \end{cases}$$

Sada imamo:

$$\mathcal{R}(X) = [0, 2]$$

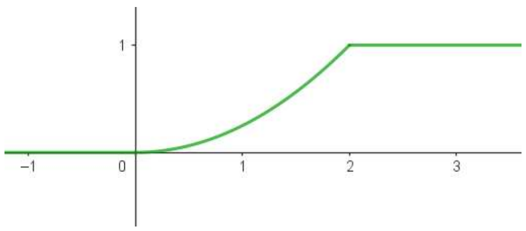
$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^2 \frac{x^2}{2} dx = \frac{4}{3}$$

$$\operatorname{Var} X = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mathbb{E}(X)^2 = \int_0^2 \frac{x^3}{2} dx - \frac{16}{9} = \frac{2}{9}$$

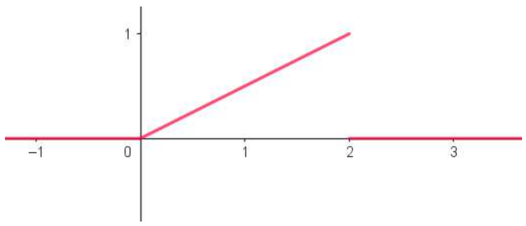
$$\sigma(X) = \sqrt{\operatorname{Var} X} = \frac{\sqrt{2}}{3}.$$

Tražena vjerojatnost je

$$\mathbb{P}(1.5 < X \leq 2) = F(2) - F(1.5) = 0.4375.$$



Slika 2.12: Funkcija distribucije za primjer 2.2.13.



Slika 2.13: Funkcija gustoće za primjer 2.2.13.

U nastavku pogledajmo primjer najpoznatije neprekidne slučajne varijable.

Normalna (Gaussova) slučajna varijabla

Normalna slučajna varijabla je temeljni koncept u statistici i teoriji vjerojatnosti koji omogućuje modeliranje i razumijevanje mnogih stvarnih fenomena. Kada govorimo o normalnoj slučajnoj varijabli, referiramo se na varijablu koja slijedi normalnu razdiobu, poznatu po svojoj zvonolikoj, simetričnoj krivulji. Ova krivulja, nazvana i Gaussova krivulja, opisuje način na koji se vrijednosti varijable raspoređuju oko srednje vrijednosti, pri čemu je najveća vjerojatnost pridružena vrijednostima bliskim sredini, dok ekstremne vrijednosti imaju manju vjerojatnost pojavljivanja. Normalne slučajne varijable su ključne u mnogim područjima, uključujući fiziku, biologiju, ekonomiju i psihologiju, jer mnogi prirodni i društveni procesi prirodno slijede normalnu razdiobu. Na primjer, visina ljudi, pogreške u mjerenju i rezultati testova često se modeliraju pomoću normalne slučajne varijable. U ovom potpoglavlju istražit ćemo karakteristike normalne slučajne varijable, kako se ona matematički definira i kako se koristi u analizi podataka.

Definicija 2.2.13. *Kažemo da neprekidna slučajna varijabla X ima* ***normalnu razdiobu*** *s parametrima μ i σ^2 i označavamo ju s $X \sim N(\mu, \sigma^2)$, ako joj je funkcija gustoće*

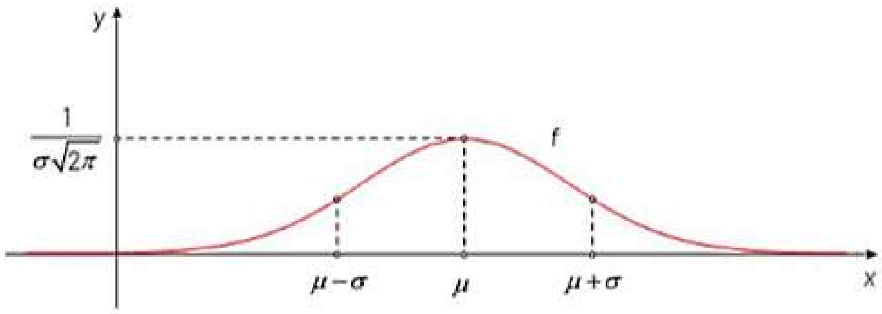
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Razlomak $\frac{1}{\sigma\sqrt{2\pi}}$ iz prethodne definicije ključan je za normalizaciju i skaliranje funkcije gustoće normalne distribucije. On osigurava da ukupna površina ispod krivulje bude 1 te omogućava da oblik distribucije pravilno ovisi o standardnoj devijaciji σ.

Za normalnu slučajnu varijablu X vrijedi

$$\mathbb{E}(X) = \mu, \operatorname{Var} X = \sigma^2.$$

Graf funkcije $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ zovemo normalna krivulja ili Gaussova krivulja.



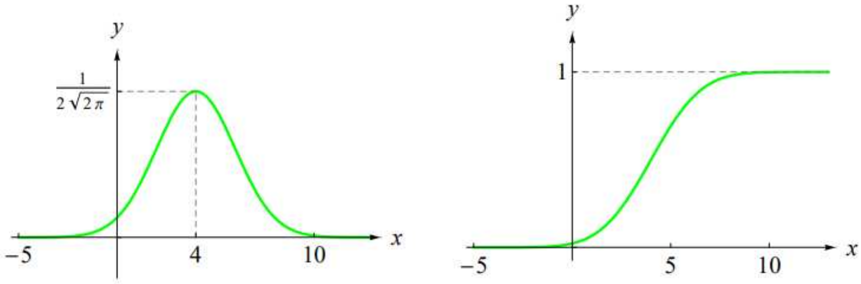
Slika 2.14: Normalna ili Gaussova krivulja [10]

Osnovna svojstva Gaussove krivulje [5]:

1. Funkcija f definirana je za sve realne brojeve.
2. Funkcija je strogo pozitivna, $f(x) > 0$ za svaki realni broj x. Površina ispod grafa funkcije jednaka je 1.
3. Graf funkcije asimptotski se približava osi x kada $x \to -\infty$ ili $x \to +\infty$.
4. Maksimum funkcije jest u točki $(\mu, \frac{1}{\sigma\sqrt{2\pi}})$.
5. Graf funkcije simetričan je s obzirom na pravac $x = \mu$.
6. Parametar σ utječe na oblik grafa funkcije. Ako se σ povećava, graf funkcije postaje spljošteniji, a ako se σ smanjuje, graf funkcije se isteže u visinu.
7. Parametar μ translatira graf funkcije duž osi x.

Funkcija distribucije normalne slučajne varijable definirana je izrazom

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



Slika 2.15: Graf funkcije gustoće i funkcije distribucije slučajne varijable $X \sim N(4, 4)$ [12]

Deriviranjem prethodne funkcije gustoće dobivamo:

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

- $f'(x) = 0$ za $x = \mu$ i to odgovara točki na grafu gdje funkcija $f(x)$ poprima maksimalnu vrijednost

Ponovnim deriviranjem dobivamo:

$$f''(x) = \frac{-1}{\sigma^2\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\left[\frac{1}{\sigma} - \frac{(x-\mu)^2}{\sigma^3}\right].$$

- $f''(x) = 0$ za $\frac{(x-\mu)^2}{\sigma^3} = \frac{1}{\sigma} \iff (x-\mu)^2 = \sigma^2 \iff x-\mu = \pm\sigma \iff x = \mu \pm \sigma.$

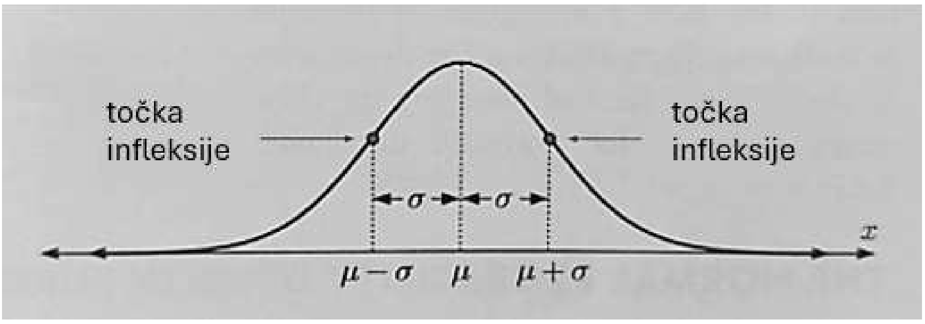
Dakle, točke infleksije su $x = \mu + \sigma$ i $x = \mu - \sigma$. Za normalnu krivulju, standardna devijacija je jedinstveno određena kao horizontalna udaljenost od linije simetrije $x = \mu$ do točke infleksije.

Za normalnu distribuciju sa srednjom vrijednosti μ i standardnom devijacijom σ, postotna raspodjela vjerojatnosti gdje slučajna varijabla može ležati prikazana je niže na slici. Primijetimo da vrijedi:

$\approx 68.26\%$ vrijednosti nalazi se između $\mu - \sigma$ i $x = \mu + \sigma$,

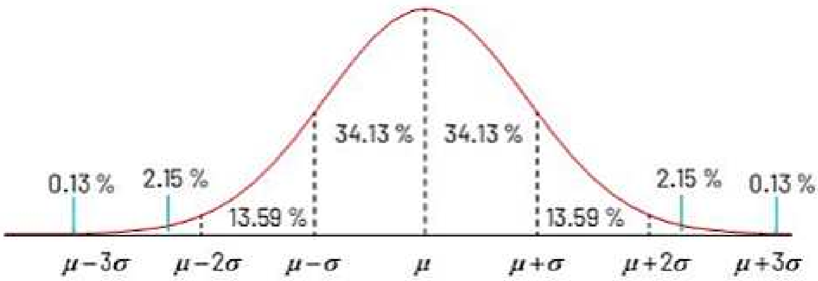
$\approx 95.44\%$ vrijednosti nalazi se između $\mu - 2\sigma$ i $x = \mu + 2\sigma$,

$\approx 99.74\%$ vrijednosti nalazi se između $\mu - 3\sigma$ i $x = \mu + 3\sigma$.



Slika 2.16: Točke infleksije na Gaussovoj krivulji [10]

Iz prikazanog vidimo da za $X \sim N(\mu, \sigma^2)$ vrijedi $\mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9974$ odnosno da vrijednosti slučajne varijable po normalnoj distribuciji padaju u interval koji se proteže za po tri standarne devijacije lijevo i desno od očekivanja. Ovu činjenicu nazivamo pravilom tri sigma. Slično, s 95-postotnom sigurnošću možemo tvrditi da se normalno distribuirana veličina nalazi za najviše dvije standardne devijacije lijevo i desno od očekivanja, odnosno da je u tom intervalu više od 95% površine ispod pripadne krivulje.



Slika 2.17: Pravilo tri sigma [10]

Standardna normalna distribucija

Među svim normalnim distribucijama posebnu ulogu ima **jedinična ili standardna normalna distribucija**. Svaka normalna X distribucija može se transformirati u standardnu normalnu distribuciju ili Z-distribuciju koristeći transformaciju $z = \frac{x-\mu}{\sigma}$. To je distribucija kojoj su parametri $\mu = 0$ i $\sigma^2 = 1$, odnosno očekivanje je 0 i varijanca je 1. Označavamo ju s $Z \sim N(0,1)$. Funkcija gustoće normalne distribucije $N(0,1)$ dana je izrazom

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, x \in \mathbb{R}.$$

Za $\mu = 0$ i $\sigma = 1$, vrijednost z iznosi $z = \frac{x-\mu}{\sigma} = x$. Dakle, distribucija $N(0,1)$ ostaje nepromijenjena pod transformacijom, a funkcija gustoće vjerojatnosti za Z-distribuciju je

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, z \in \mathbb{R}.$$

Da bismo transformirali $N(0,1)$ u $N(\mu,\sigma^2)$, trebamo:

- horizontalno rastezanje s faktorom skaliranja σ kako bismo premjestili točke infleksije na $x = \pm\sigma$,
- horizontalnu translaciju za μ jedinica udesno kako bismo srednju vrijednost premjestili na $x = \mu$.

Tada dobivamo:

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

Međutim, kada smo napravili horizontalno rastezanje, promijenili smo površinu ispod krivulje. Za funkciju gustoće znamo da površina mora biti jednaka 1. Stoga, dodatno trebamo vertikalno rastezanje s faktorom skaliranja $\frac{1}{\sigma}$ kako bismo osigurali da $\int_{-\infty}^{\infty} f(x)dx = 1$. Sada imamo funkciju gustoće

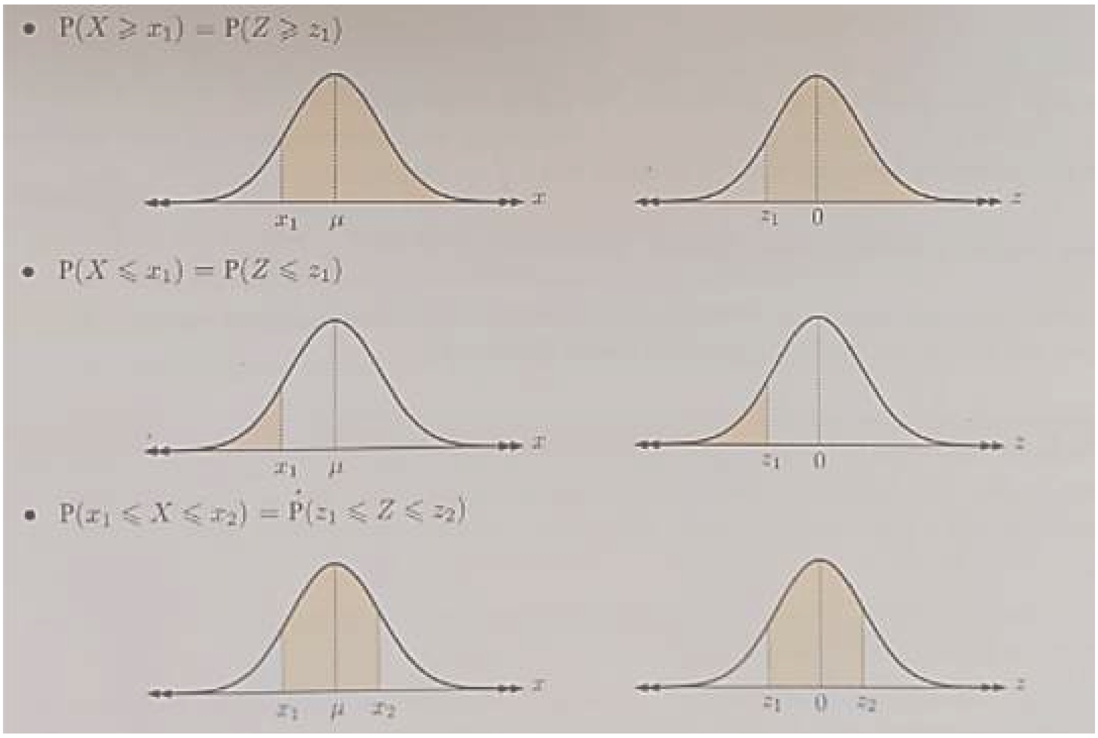
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Na primjer, ako je

- $z = 1.84$, tada je x za 1.84 standardne devijacije desno od srednje vrijednosti,
- $z = -0.273$, tada je x za 0.273 standardne devijacije lijevo od srednje vrijednosti.

Ovo znači da za $X \sim N(\mu,\sigma^2)$ i $Z \sim N(0,1^2)$, te za bilo koje $x_1, x_2 \in \mathbb{R}, x_1 < x_2$ sa pripadajućim z-vrijednostima $z_1 = \frac{x_1-\mu}{\sigma}$ i $z_2 = \frac{x_2-\mu}{\sigma}$ vrijedi:

- $\mathbb{P}(X \geq x_1) = \mathbb{P}(Z \geq z_1)$,
- $\mathbb{P}(X \leq x_1) = \mathbb{P}(Z \leq z_1)$,
- $\mathbb{P}(x_1 \leq X \leq x_2) = \mathbb{P}(z_1 \leq Z \leq z_2)$.



Slika 2.18: Prikaz vjerojatnosti za različite z vrijednosti [10]

Vrijednosti funkcije distribucije slučajne varijable moraju se računati korištenjem metoda numeričkog integriranja jer se integrali uglavnom ne daju riješiti eksplicitno. Iz tog razloga često su unutar knjiga priložene tablice vrijednosti funkcije distribucije standardne normalne slučajne varijable. Također, u današnje vrijeme, za računanje vjerojatnosti po normalnoj distribuciji najčešće se koriste naprednija džepna računala ili specijalizirani software na računalu.

Primjer 2.2.14. *[10] Nika je postigla 73% iz povijesti, gdje je prosjek razreda bio 68%, a standardna devijacija 10.2%. Iz matematike je postigla 66%, gdje je prosjek razreda bio 62%, a standardna devijacija 6.8%. U kojem je predmetu Nika bila bolja u usporedbi s ostatkom razreda? Pretpostavimo da su rezultati za oba predmeta normalno distribuirani.*

Rješenje: Nikina z-vrijednost za povijest iznosi $\frac{73-68}{10.2} \approx 0.490$. Nikina z-vrijednost za matematiku iznosi $\frac{66-62}{6.8} \approx 0.588$. Dakle, Nikin rezultat iz matematike bio je 0.588 standardnih devijacija iznad prosjeka, dok je njezin rezultat iz povijesti bio 0.490 standardnih devijacija iznad prosjeka što znači da je Nikin rezultat iz matematike bio je bolji u usporedbi s njezinim razredom, iako je njezin postotak riješenosti bio niži.

Povezivanje diskretnih i neprekidnih slučajnih varijabli na primjeru bacanja kockice

U ovom poglavlju, analizirat ćemo ponašanje slučajne varijable u kontekstu bacanja kockice. Povezat ćemo diskretne i neprekidne slučajne varijable koristeći primjere i distribucije koje se primjenjuju na slučajne varijable. Kroz primjer, primijenjivat ćemo binomnu distribuciju za diskretne slučajne varijable koju ćemo aproksimirati pomoću normalne distribucije za neprekidne slučajne varijable.

Primjer 2.2.15. *Razmotrimo slučaj bacanja simetrične kockice n puta. Izračunajte vjerojatnost da se broj 3 pojavi u n puta bacanja kockice.*

Rješenje: Pretpostavimo da imamo simetričnu kockicu i bacamo je n puta. Neka slučajna varijabla X predstavlja broj pojavljivanja broja 3. Početni korak je razumijevanje da X slijedi binomnu distribuciju s parametrima n (broj bacanja) i $p = \frac{1}{6}$ (vjerojatnost da se na jednom bacanju pojavi broj 3). Ako bacamo kockicu n puta, vjerojatnost da se broj 3 pojavi točno k puta je dana formulom binomne distribucije:

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

gdje je $p = \frac{1}{6}$, a $(1 - p) = \frac{5}{6}$.

Distrubucija bi tada izgledala kao:

$$X \sim \begin{pmatrix} 0 & 1 & \dots & k & \dots \\ \binom{n}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^n & \binom{n}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{n-1} & \dots & \binom{n}{k}\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{n-k} & \dots \end{pmatrix}.$$

Kad broj bacanja n postane velik, binomna distribucija se može aproksimirati normalnom distribucijom. To je zato što binomna distribucija ima tendenciju da se oblikuje u zvono kad su n i p takvi da su np i $n(1 - p)$ dovoljni za primjenu Centralnog graničnog teorema. Za $n = 10$, normalna aproksimacija nije savršena odnosno ima asimetričnu distribuciju, ali je korisna za uvid. Distribuciju za različite brojeve k (od 0 do 10) možemo prikazati tablično na sljedeći način:

$$X \sim \begin{pmatrix} 0 & 1 & 2 & 3 & \dots & 9 & 10 \\ \binom{10}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^{10} & \binom{10}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^9 & \binom{10}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^8 & \binom{10}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7 & \dots & \binom{10}{9}\left(\frac{1}{6}\right)^9\left(\frac{5}{6}\right)^1 & \binom{10}{10}\left(\frac{1}{6}\right)^{10}\left(\frac{5}{6}\right)^0 \end{pmatrix},$$

gdje brojevi od 0 do 10 predstavljaju "broj dobivenih 3" s pripadnim vjerojatnostima.

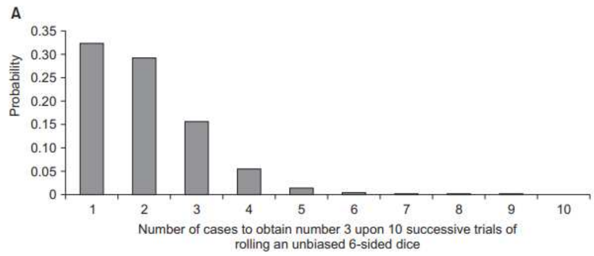
Srednja vrijednost (μ) binomne distribucije je $\mu = np = 10 \cdot \frac{1}{6} \approx 1.67$.

Standardna devijacija (σ) je $\sigma = \sqrt{np(1-p)} = \sqrt{10 \cdot \frac{1}{6} \cdot \frac{5}{6}} \approx 1.1785$.

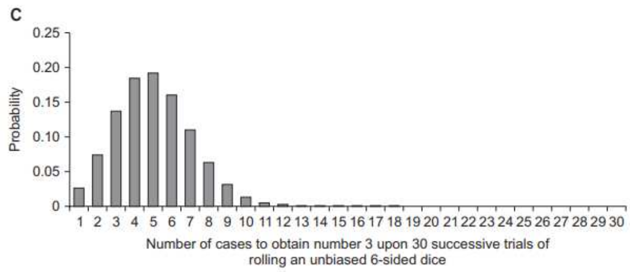
Ako koristimo normalnu distribuciju Y za aproksimaciju imamo

$$Y \sim N(\mu, \sigma^2) = N(1.67, 1.13^2).$$

Histogram distribucije broja pojavljivanja broja 3 u 10 bacanja kockice trebao bi izgledati kao diskretna raspodjela sa skokovima, dok bi normalna aproksimacija trebala izgledati kao zvono, sa sredinom oko 1.67 i rasponom pokriva sve vrijednosti k oko te sredine. Osi x prikazuju broj pojavljivanja broja 3 (0, 1, 2, ..., 10), dok osi y prikazuju vjerojatnosti ili frekvencije za svaki broj pojavljivanja (slike 2.19., 2.20.). Histogram za binomnu distribuciju oko 0 i 1 će imati veće frekvencije, a vjerojatnosti za k veće od sredine će brzo opadati. Histogram za normalnu aproksimaciju imat će oblik zvona s vrhom oko 1.67, s većom gustoćom oko sredine te opadajućom gustoćom prema krajevima.



Slika 2.19: Broj slučajeva da se dobije broj 3 u 10 uzastopnih pokušaja bacanja pravedne 6-strane kocke (primjer 2.2.15.) [15]



Slika 2.20: Broj slučajeva da se dobije broj 3 u 30 uzastopnih pokušaja bacanja pravedne 6-strane kocke (primjer 2.2.15.) [15]

Povezivanje diskretnih i neprekidnih slučajnih varijabli na primjeru Galtonove daske [13]

Galtonova daska, koja se koristi za prikazivanje distribucije kuglica koje padaju kroz niz čunjeva, omogućava vizualizaciju kako se diskretne vjerojatnosti mogu pretvoriti u kontinuiranu normalnu distribuciju kada se broj eksperimenata poveća. Dok kuglice padaju kroz čunjeve, one se raspoređuju u različite ”kolone” na dnu daske. Svaka od tih kolona predstavlja broj uspjeha (ili određeni broj objekata) koji se pojavljuje s određenom vjerojatnošću. S obzirom na to da broj kuglica raste, raspodjela u kolonama počinje sve više nalikovati normalnoj distribuciji, što je vidljivo u slučaju velikih uzoraka. Ova vizualizacija jasno prikazuje kako diskretne varijable, poput broja uspjeha u nizu nezavisnih Bernoullijevih pokušaja (koji slijedi binomnu distribuciju), mogu stvoriti kontinuiranu raspodjelu kada se razmatraju u velikim brojevima.

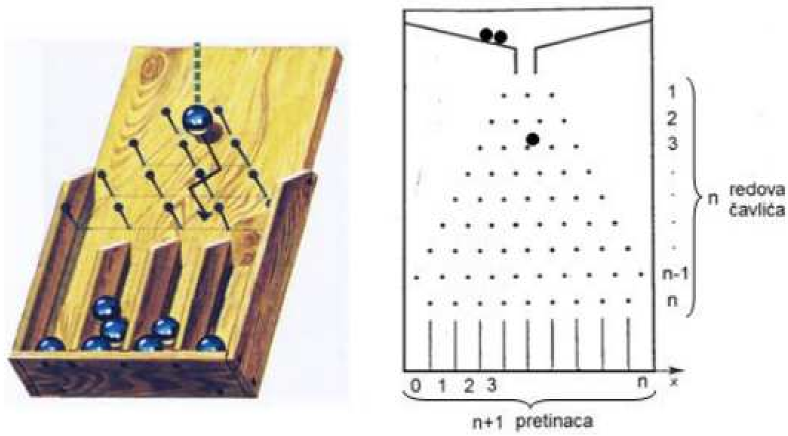
Kroz lijevak na vrhu Galtonove daske puštamo kuglice da se kotrljaju između čavlića. Kuglica pri svakom sudaru s čavlićem ima mogućnost da ga obiđe slijeva ili zdesna. Sve one na kraju završavaju svoje putanje u pretincima na dnu uređaja i tako podijeljene čine vizualni prikaz distribucije vjerojatnosti koji obično zovemo histogram.

Primjer 2.2.16. *Ako kroz lijevak pustimo N kuglica, i one prođu kroz n redova čavlića, koliko će ih dospijeti u pojedini pretinac?*

Rješenje: Označimo pretince brojevima od 0 do n počevši slijeva na desno (imamo n redova čavlića i $n+1$ pretinac). Možemo zaključiti da će pojedina kuglica nakon nalijetanja na n čavlića pasti u pretinac označen onim brojem koliko joj se desilo desnih obilazaka. Na primjer, da bi dospjela u $0-$ti pretinac, ne smije imati niti jedan desni obilazak i općenito, da bi dospjela u x-ti pretinac, mora joj se desiti x desnih obilazaka i naravno $x-n$ lijevih. Bavimo se s dvije mogućnosti koje se sastoje od toga da se neki događaj pojavi (desnii obilazak) ili ne pojavi (lijevi obilazak), pa možemo označiti vjerojatnost desnog s p, a lijevog s $q = (1-p)$. Kod precizno izrađene daske $p = \frac{1}{2}$ i $q = \frac{1}{2}$ tj. vjerojatnost lijevog i desnog obilaska su jednake, $p+q = 1$. To zapravo znači da je vjerojatnost da od n obilazaka njih x bude desnih dana funkcijom

$$f(x) = \binom{n}{x} \cdot \left(\frac{1}{2}\right)^x \cdot \left(\frac{1}{2}\right)^{n-x}.$$

$$f(x) = \binom{n}{x} \cdot \left(\frac{1}{2}\right)^n.$$



Slika 2.21: Galtonova daska [13]

2.3 Centralni granični teorem

U osnovnim i srednjim školama, učenici se susreću s osnovnim pojmovima vjerojatnosti i statistike koji su usmjereni na razumijevanje osnovnih statističkih prikaza i jednostavnih izračuna vjerojatnosti. Uvođenje kompleksnijih statističkih teorija, kao što je Centralni granički teorem (CGT), može nadmašiti okvir srednjoškolskog gradiva. Ovaj koncept, iako na prvi pogled može izgledati apstraktno, igra ključnu ulogu u naprednoj statistici i teoriji vjerojatnosti, te je temelj za mnoge metode i tehnike koje se koriste u znanstvenom istraživanju, analizi podataka i statističkoj primjeni.

Centralni granični teorem (CGT) objašnjava kako se raspodjela suma velikog broja neovisnih slučajnih varijabli približava normalnoj distribuciji, bez obzira na raspodjelu pojedinih varijabli. Ovo je izuzetno važno za razumijevanje kako se rezultati iz statističkih uzoraka mogu interpretirati i primijeniti u praksi, omogućujući statističarima da koriste normalnu raspodjelu kao aproksimaciju u mnogim situacijama.

Iako ovaj koncept nadmašuju osnovno i srednjoškolsko obrazovanje, u ovom poglavlju, usmjerit ćemo se na istraživanje ovog teorema koji nudi snažan okvir za razumijevanje statističkih principa i njihovih primjena.

Da bismo učenicima približili ovaj koncept, prvo je važno objasniti osnovne pojmove vezane uz vjerojatnost i statistiku. Nakon toga, uvođenje Centralnog graničnog teorema (CGT) može se olakšati korištenjem jednostavnog primjera. Na primjer, možemo koristiti eksperiment s bacanjem kockice više puta uz bilježenje prosjeka rezultata za svaki skup bacanja. Učenicima treba demonstrirati kako distribucija prosječnih rezultata postaje sve bliža normalnoj s povećanjem broja bacanja. Također, važno je koristiti vizualizacije kako bi se učenicima približio ovaj koncept. Različite simulacije na računalu ili stvarni podaci iz eksperimenta mogu biti korisni za prikazivanje normalne distribucije. Nakon što se objasni kako CGT funkcionira, treba naglasiti važnost normalne distribucije u statistici jer se mnoge statističke metode i testovi oslanjaju se na pretpostavku da su podaci normalno distribuirani.

Pogledajmo primjer bacanja kocke [16]. Zamislimo da 100 puta bacamo pravilnu šesterostranu kocku. Kocka ima šest mogućih ishoda: 1, 2, 3, 4, 5 i 6, a svaki od njih je jednako vjerojatan. Naš cilj je pronaći raspon vrijednosti unutar kojeg možemo biti 95% sigurni da će se nalaziti zbroj svih tih bacanja. Postupak možemo prikazati u nekoliko koraka. U prvom koraku trebamo definirati slučajnu varijablu. U ovom slučaju svaki ishod bacanja kocke možemo predstaviti kao slučajnu varijablu X_i, za $i = 1, ..., 100$. Dakle, $X_1, X_2, ..., X_{100}$ predstavljaju nezavisne i identično distribuirane varijable. U drugom koraku trebamo izračunati očekivanje, varijancu te standardnu devijaciju slučajne varijable X. Očekivana vrijednost pojedinačnog bacanja kocke (srednja vrijednost) je:

$$\mathbb{E}(X_i) = \frac{1+2+3+4+5+6}{6} = 3.5.$$

Za 100 bacanja, očekivana vrijednost zbroja svih ishoda je:

$$\mathbb{E}(S_{100}) = 100 \cdot \mathbb{E}(X_i) = 100 \cdot 3.5 = 350.$$

Varijanca pojedinačnog bacanja kocke izračunava se kao:

$$\operatorname{Var} X_i = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2,$$

gdje je

$$\mathbb{E}(X_i^2) = \frac{1^2 + 2^2 + \ldots + 6^2}{6} = 15.1667.$$

Dakle,

$$\operatorname{Var} X_i = 15.1667 - 3.5^2 = 15.1667 - 12.25 = 2.9167.$$

Za 100 bacanja, ukupna varijanca zbroja S_{100} je:

$$\operatorname{Var} S_{100} = 100 \cdot \operatorname{Var} X_i = 100 \cdot 2.9167 = 291.67.$$

Standardna devijacija zbroja S_{100} je:

$$\sigma_{S_{100}} = \sqrt{\operatorname{Var} S_{100}} = \sqrt{291.67} \approx 17.08.$$

Sada, prema centralnom graničnom teoremu, zbroj S_{100} slijedi približno normalnu razdiobu s očekivanjem 350 i standardnom devijacijom 17.08. Da bismo pronašli interval unutar kojeg će se zbroj S_{100} nalaziti s 95% sigurnošću, koristimo činjenicu da za standardnu normalnu razdiobu vrijedi:

$$\mu_S - 2\sigma_S = 350 - 2 \cdot 17.08 = 315.84 \approx 316$$

te

$$\mu_S + 2\sigma_S = 350 + 2 \cdot 17.08 = 384.16 \approx 384.$$

Stoga, prema centralnom graničnom teoremu, možemo biti 95% sigurni da će se zbroj ishoda nakon 100 bacanja šesterostrane kocke nalaziti unutar intervala $[316, 384]$.
Ovaj primjer ilustrira primjenu CGT-a odnosno iako pojedinačno bacanje kocke nije normalno distribuirano, zbroj velikog broja bacanja konvergira prema normalnoj distribuciji, omogućujući nam da koristimo normalnu distribuciju za procjenu intervala pouzdanosti.

Teorem 2.3.1. *(Centralni granični teorem [11]) Neka je $X_n, n \in \mathbb{N}$ niz nezavisnih jednako distribuiranih slučajnih varijabli sa zajedničkim očekivanjem μ i zajedničkom varijancom σ^2. Za $n \in \mathbb{N}$, definirajmo $S_n := X_1 + \ldots + X_n$. Tada za sve $x \in \mathbb{R}$ vrijedi*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n}} \leq x\right) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2\sigma^2}}\, dx.$$

Drugim riječima, za velike n, niz $\frac{S_n - n\mu}{\sqrt{n}}$ konvergira po distribuiciji k $N(0, \sigma^2)$.

Primjenjujući centralni granični teorem binomnu slučajnu varijablu možemo aproksimirati normalom.

Teorem 2.3.2. *[18] (de Moivre-Laplace CGT). Za svaki $n \geq 1$, neka je $X_n \sim B(n, p)$ pri čemu je $p \in (0, 1)$. Ako je $Z \sim N(0, 1)$, tada za svaki $x \in \mathbb{R}$ vrijedi*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_n - np}{\sqrt{npq}} \leq x\right) = \mathbb{P}(Z \leq x) = \Phi(x).$$

Uočite, $np = \mathbb{E}(X_n)$, a $\sqrt{npq} = \sqrt{\mathrm{Var}\, X_n}$.

Napomena 2.3.3. *[18] Za $X \sim B(n, p)$ i veliki n i $a \leq b$ također vrijedi:*

$$\mathbb{P}(X \geq a) \approx \mathbb{P}\left(Z \geq \frac{a - np}{\sqrt{npq}}\right) = 1 - \Phi\left(\frac{a - np}{\sqrt{npq}}\right)$$

i

$$\mathbb{P}(a \leq X \leq b) \approx \mathbb{P}\left(\frac{a - np}{\sqrt{npq}} \leq Z \leq \frac{b - np}{\sqrt{npq}}\right) = \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right).$$

Primjer 2.3.4. *[11] Neka je $X \sim B(50, 0.237)$. Izračunajmo $\mathbb{P}(7 \leq X \leq 9)$.*

Rješenje: Znamo da je $X = X_1 + X_2 + \ldots + X_{50}$, gdje su $X_1 + X_2 + \ldots + X_{50}$ nezavisne jednako distribuirane Bernoullijeve slučajne varijable

$$X_i \sim \begin{pmatrix} 0 & 1 \\ 0.763 & 0.237 \end{pmatrix}, \; i = 1, ..., 50.$$

Kako je $\mathbb{E}(X_i) = 0.237$, $\operatorname{Var} X_i = 0.18$ i $\sigma(X_i) = 0.42$ za $i = 1, ...50$, iz centralnog graničnog teorema imamo:

$$\begin{aligned}
\mathbb{P}(7 \leq X \leq 9) &= \mathbb{P}(7 \leq X_1 + ... + X_{50} \leq 9) \\
&= \mathbb{P}(-4.85 \leq X_1 + ... + X_{50} - 11.85 \leq -2.85) \\
&= \mathbb{P}(-1.613 \leq \frac{X_1 + ... + X_{50} - 11.85}{3} \leq -0.94) \\
&\approx \mathbb{P}(-1.613 \leq Z \leq -0.94) \\
&= \Phi(-0.94) - \Phi(-1.163) \\
&= \Phi(1.613) - \Phi(0.94) \\
&= 0.12,
\end{aligned}$$

gdje vrijednosti $\Phi(1.613)$ i $\Phi(0.94)$ iščitavamo iz tablica.

Poglavlje 3

Vjerojatnost i statistika u kurikulumu

U Hrvatskoj se vjerojatnost i statistika prvi put uvela u nastavni plan i program 2006. godine kroz teme *Prikazivanje i analiza podataka* i *Vjerojatnost slučajnog događaja* u 7. razredu osnovne škole. Uvođenje ovih dviju tema u kurikulum nije bilo dovoljno da bi dovelo do dubljeg razumijevanja ove grane matematike.

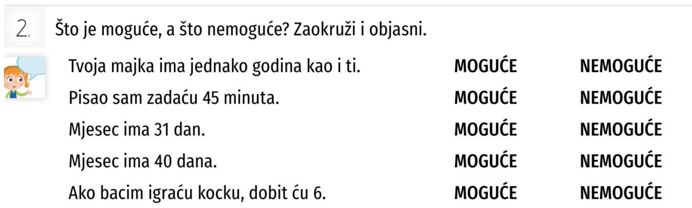
Proširenje uvođenja vjerojatnosti i statistike u nastavu matematike bio je u sklopu Nacionalnog okvirnog kurikuluma prema kojem su vjerojatnost i statistika bile uvrštene kao teme u svim ciklusima obrazovanja u sklopu matematičkog koncepta *Podatci*. U tom vremenu, većina udžbenika za razrednu nastavu matematike bila je pisana prema Hrvatskom nacionalnom obrazovnom standardu (HNOS-u) iz 2006. godine, a koji nisu uključivali nastavni sadržaj iz vjerojatnosti i statistike. Stoga, nastavnicima je ostalo da sami razmisle i osmisle kako i kada će uvesti navedene teme.

U sklopu Cjelovite kurikularne reforme izrađen je Nacionalni kurikulum nastavnog predmeta matematike. U njemu se kao jedna od domena nalazi domena *Podatci, statistika i vjerojatnost*. Trenutno je na snazi kurikulum koji je donesen 2019. godine. Prema njemu domena *Podatci, statistika i vjerojatnost* bavi se prikupljanjem, razvrstavanjem, obradom, analizom i prikazivanjem podataka u odgovarajućemu obliku. Od učenika se očekuje da podatke koji su prikazani grafički ili na neki drugi način pravilno očitaju, protumače i ispravno upotrijebe. Cilj je da to postignu kroz korištenje jezika statistike. Statistika uključuje korištenje matematičkih metoda za računanje mjera srednje vrijednosti, mjera raspršenja, mjera položaja i korelacije podataka. Promatrajući veze među podacima i frekvencije pojavljivanja, dolazi se do pojma vjerojatnosti. Na taj način se određuje broj povoljnih i svih mogućih ishoda, procjenjuje se i izračunava vjerojatnost, što omogućuje predviđanje događaja.

Od 2000. godine kvaliteta matematičkog obrazovanja u 43 zemalje OECD-a procjenjuje se prema rezultatima učenika u testiranjima PISA i TIMSS. PISA više od trećinu svojim zadataka iz područja matematičke pismenosti pridaje području podataka, statistike i vjerojatnosti. Istraživanja su pokazala da hrvatski učenici ostvaruju osrednje rezultate, dok najbolje rezultate ostvaruju učenici iz država koje sadržaje iz vjerojatnosti i statistike poučavaju od rane školske dobi.

3.1 Vjerojatnost i statistika u nastavi matematike u osnovnoj školi

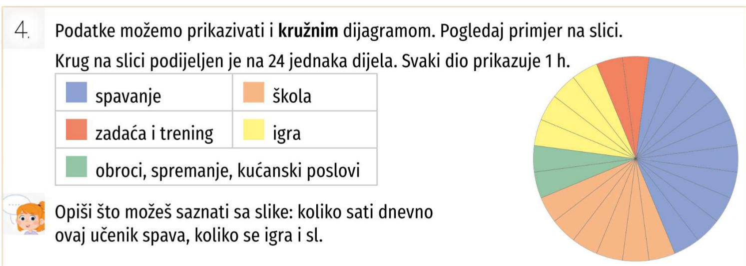
Prema novom hrvatskom kurikulumu, učenici se s osnovnim pojmovima vjerojatnosti prvi puta susreću u osnovnoj školi. Kurikulum predviđa učenje vjerojatnosti u drugom razredu osnovne škole. Ključni pojmovi s kojima se učenici upoznaju su moguć i nemoguć događaj. Od učenika se očekuje da u različitim situacijama predviđaju moguće i nemoguće događaje te da objasne zašto je neki događaj (ne)moguć.


2. Što je moguće, a što nemoguće? Zaokruži i objasni.

Tvoja majka ima jednako godina kao i ti.	MOGUĆE	NEMOGUĆE
Pisao sam zadaću 45 minuta.	MOGUĆE	NEMOGUĆE
Mjesec ima 31 dan.	MOGUĆE	NEMOGUĆE
Mjesec ima 40 dana.	MOGUĆE	NEMOGUĆE
Ako bacim igraću kocku, dobit ću 6.	MOGUĆE	NEMOGUĆE

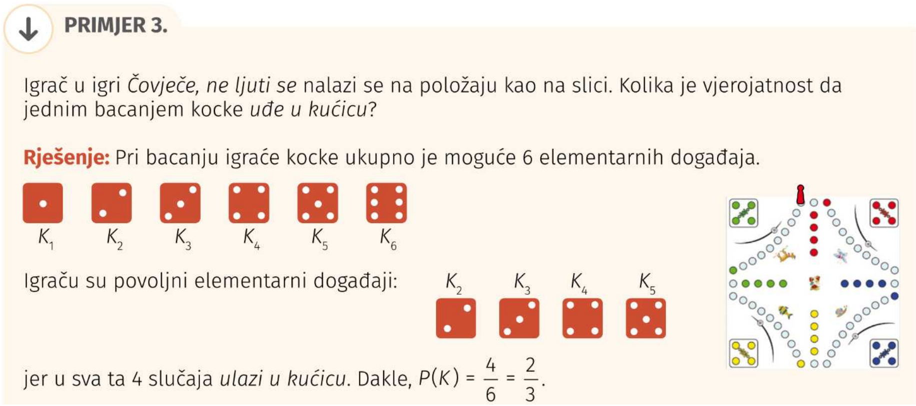
Slika 3.1: Primjer zadatka iz udžbenika za 2. razred osnovne škole [9]

U četvrtom razredu osnovne škole od učenika se očekuje da prikupljaju, razvrstavaju i prikazivaju podatke tablicama ili dijagramima. Nakon što učenici nauče samostalno kreirati tablice i dijagrame, važno je da nauče kako pravilno iščitati i interpretirati iste podatke. Također, za grafičko predočavanje često se upotrebljavaju piktogrami. Osim za statistiku, čitanje piktograma pogodno je i za uvježbavanje tablice množenja. Na kraju, od učenika se očekuje da mogu opisati vjerojatnost događaja.



Slika 3.2: Primjer zadatka iz udžbenika za 4. razred osnovne škole [8]

U osmom razredu osnovne škole od učenika se očekuje da računaju vjerojatnost i da na osnovi tih podataka donose odluke. Ključni pojmovi s kojima se učenici upoznaju su: slučajni događaj, relativna frekvencija te vjerojatnost slučajnog događaja. Od učenika se očekuje da mogu navesti elementarne događaje u promatranom pokusu ili promatranoj situaciji i odrediti koji su povoljni za zadani događaj te da znaju odrediti vjerojatnost slučajnog događaja. Ovim se ishodom ne provjerava tehnika računanja, nego učenikovo koceptualno razumijevanje vjerojatnosti i sposobnost analize problema.



PRIMJER 3.

Igrač u igri *Čovječe, ne ljuti se* nalazi se na položaju kao na slici. Kolika je vjerojatnost da jednim bacanjem kocke *uđe u kućicu*?

Rješenje: Pri bacanju igraće kocke ukupno je moguće 6 elementarnih događaja.

K_1 K_2 K_3 K_4 K_5 K_6

Igraču su povoljni elementarni događaji: K_2 K_3 K_4 K_5

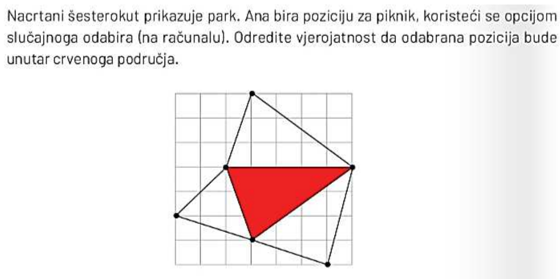
jer u sva ta 4 slučaja *ulazi u kućicu*. Dakle, $P(K) = \frac{4}{6} = \frac{2}{3}$.

Slika 3.3: Primjer zadatka iz udžbenika za 8. razred osnovne škole [7]

3.2 Vjerojatnost i statistika u nastavi matematike u srednjoj školi

Vjerojatnost se obrađuje i u drugim i četvrtim razredima srednjih škola (gimnazije, tehničke škole).

U drugom razredu srednje škole, učenici, nakon što se prisjete osnovnih pojmova, ususreću se s klasičnim definicijama vjerojatnosti, definiraju osnovna svojstva vjerojatnosti, proučavaju načine zadavanja vjerojatnosti. Kroz različite primjere i zadatke, stječu uvid u primjenu vjerojatnosti u stvarnim situacijama te razvijaju sposobnost donošenja zaključaka temeljenih na statističkim podacima. Tijekom ovog procesa, učenici se često susreću s određenim miskoncepcijama, kao što su pogrešno razumijevanje nasumičnosti ili pretjerano pojednostavljivanje složenih događaja. Učenike se potiče da kroz diskusiju i rješavanje problema razriješe ove miskoncepcije i steknu dublje razumijevanje koncepta vjerojatnosti.


Nacrtani šesterokut prikazuje park. Ana bira poziciju za piknik, koristeći se opcijom slučajnoga odabira (na računalu). Odredite vjerojatnost da odabrana pozicija bude unutar crvenoga područja.

Slika 3.4: Primjer zadatka iz udžbenika za 2. razred srednje škole [6]

U prvom dijelu četvrtog razreda srednje škole, učenici se susreću s pojmovima vjerojatnosnog prostora, nezavisnih i zavisnih događaja, uvjetne vjerojatnosti te Bayesove formule, što im omogućuje dublje razumijevanje i primjenu vjerojatnosti u složenijim problemima. Učenici razvijaju sposobnost analiziranja situacija koje uključuju nasumične događaje te primjene stečenih znanja u rješavanju praktičnih problema. Dodatno, učenici koji prema programu imaju 224 sata matematike godišnje, obrađuju i nastavnu cjelinu *Slučajne varijable*, u kojoj se upoznaju s diskretnim i neprekidnim slučajnim varijablama. Od njih se očekuje da razumiju osnovna svojstva ovih varijabli i da znaju primijeniti binomnu i normalnu razdiobu u konkretnim situacijama, čime se osposobljavaju za rješavanje zadataka koji zahtijevaju primjenu statističkih metoda i vjerojatnosnih modela.


Problem Montyja Halla poznati je problem oko kojega su se vodile mnoge polemike, a vezan je uz američku zabavnu emisiju *Let's Make a Deal* u kojoj sudionik može osvojiti automobil. Pred vama se nalaze troja vrata. Iza jednih je vrata automobil, a iza preostalih su dvojih vrata koze. Nakon što odaberete jedna vrata, Monty vam otvara jedna od preostalih vrata iza kojih je koza. Sada vam daje mogućnost da se držite svojega prvobitnog odabira ili se predomislite i odaberete druga vrata. Što je vjerojatnije? Osvojiti automobil ako se držite prvobitnoga odabira vrata ili osvojiti automobil ako promijenite svoje mišljenje i odaberete druga vrata?

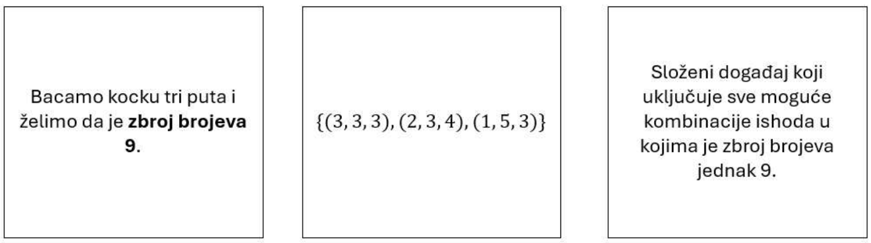
Slika 3.5: Primjer zadatka iz udžbenika za 4. razred srednje škole [17]

3.3 Aktivnosti

Aktivnosti koje se provode u nastavi matematike na temu vjerojatnosti i statistike osmišljene su kako bi učenicima omogućile praktično razumijevanje i primjenu teorijskih koncepata. Kroz interaktivne zadatke, simulacije, analizu stvarnih podataka i grupne projekte, učenici istražuju osnovne pojmove vjerojatnosti, koje smo spomenuli u prethodnim poglavljima, te razvijaju vještine interpretacije statističkih podataka. Aktivnosti su korisne i za učenike osnovnih, ali i srednjih škola jer pomažu u postepenom izgrađivanju temeljnog razumijevanja matematike, potiču logičko razmišljanje i analitičke sposobnosti, te pripremaju učenike za donošenje odluka temeljenih na podacima u stvarnim životnim situacijama.

Aktivnost 1: "Trimory" za uvježbavanje pojma složenog događaja

Cilj aktivnosti: Učenici će razumjeti i uvježbati pojam složenog događaja kroz igru u kojoj trebaju prepoznati trojke kartica prema određenim pravilima.
Ishodi učenja: Učenici će moći definirati složeni događaj.
Materijali: Kartice za igru "trimory", papir i olovka
Nastavni oblik: Rad u paru
Primjer kartica: Slika 3.6



Slika 3.6: Primjer kartice za igru *"Trimory"*

Tijek aktivnosti: Aktivnost započinje kratkim uvodom u pojam elementarnog događaja. Nastavnik podsjeća učenike na pojmove: elementarni događaji te složeni događaji, kroz jednostavne primjere s kojima su se učenici već susreli. Nakon uvoda, nastavnik učenicima predstavlja igru *"Trimory"* koja se temelji na otkrivanju trojki simbola s kartica. U igri je postavljeno barem 12 kartica okrenutih licem prema dolje, a učenici otvaraju po tri kartice u pokušaju pronalaska odgovarajućih trojki simbola (u tom slučaju zapravo traže 4 rješenja). Učenici igru igraju u paru. Nakon završetka igre, slijedi rasprava u kojoj učenici analiziraju svoje rezultate. Učitelj postavlja pitanja kako bi potaknuo diskusiju o pojmu elementarnog i složenog događaja. Učenici se potiču da definiraju svoje elementarne i složene događaje i razmisle o mogućim ishodima igre te o tome kako su neki događaji manje ili više vjerojatni. Kroz vođeni razgovor, učenici prepoznaju i zapisuju sve moguće ishode koje su susreli tijekom igre.

Aktivnost 2: Otkrivanje formule za očekivanje i varijancu binomne slučajne varijable

Cilj aktivnosti: Učenici će otkriti formulu za očekivanje i varijancu binomne slučajne varijable.
Ishodi učenja: Učenici će izračunati i interpretirati očekivanje i varijancu binomne slučajne varijable koristeći poznate formule za očekivanje i varijancu slučajnih varijabli.
Materijali: radni listić, papir i olovka
Nastavni oblik: rad u paru
Tijek aktivnosti: Aktivnost će se fokusirati na konkretne primjere s malim brojem pokušaja kako bi se otkrili obrasci koji se mogu generalizirati za opći slučaj. Na početku, nastavnik će s učenicima pponoviti osnovne pojmove vezane uz binomnu slučajnu varijablu. Treba istaknuti da binomna slučajna varijabla predstavlja broj uspjeha u n nezavisnih Bernoullijevih pokušaja, pri čemu svaki pokušaj ima istu vjerojatnost uspjeha p. Nakon toga, s učenicima analizirajte slučaja kada je broj pokušaja n jednak 1. S obzirom da učenici znaju formulu za očekivanje i varijancu slučajne varijable, za slučaj $n = 1$ najprije mogu prikazati razdiobu u obliku *Tablica 3.1*, a zatim koristeći poznate formule izračunati:

x_i	0	1
$p_i = \mathbb{P}(X_1 = x_i)$	q	p

Tablica 3.1: Razdioba slučajne varijable X_1

$$\mathbb{E}(X_1) = 0 \cdot q + 1 \cdot p = p.$$

$$\operatorname{Var} X_1 = \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2 = p - p^2 = p(1-p) = pq.$$

Analogno, za slučajeve $n = 2, n = 3$. Diskutirajte s učenicima kako rezultati iz ovog slučaja potvrđuju obrasce viđene u prethodnom primjeru. Nakon što su učenici razumjeli obrasce za $n = 2$ i $n = 3$, učenici otkrivaju formulu za opći slučaj kada je broj pokušaja n proizvoljan. Učenici prethodne primjere mogu generalizirati za bilo koji broj pokušaja. Na kraju aktivnosti, nastavnik organizira refleksiju i raspravu o tome što su učenici naučili. Učenici dijele svoja zapažanja, zaključke i poteškoće s kojima su se susreli u rješavanju zadataka.

Aktivnost 3: Motivacija za neprekidne slučajne varijable

Cilj aktivnosti: Učenici će se upoznati s konceptom neprekidnih slučajnih varijabli i razlikovati ih od diskretnih.
Ishodi učenja: Učenici će moći definirati neprekidne slučajne varijable. Učenici će moći razlikovati diskretne i neprekidne slučajne varijable.
Materijali: ploča i kreda/flomasteri
Nastavni oblik: rad u grupi (cijeli razred)
Nastavna metoda: metoda dijaloga
Tijek aktivnosti: Nastavnik započinje aktivnost predstavljanjem problema, primjer 3.3.1. Objašnjava da se podaci o težini lososa analiziraju kako bi se razumjeli osnovni pojmovi vezani uz statistiku i vjerojatnost. Učenici razmatraju zadani primjer i razmišljaju o načinu na koji se podaci mogu analizirati i prikazati. Pokušavaju razumjeti zašto je važno imati podatke raspoređene u intervalima i kako to pomaže u analizi podataka. Također, postavljaju pitanja i raspravljaju o tome kako bi mogli koristiti podatke za određivanje određenih karakteristika raspodjele. Navedeni primjer može biti motivacija za uvođenje neprekidne slučajne varijable i normalne distribucije. S obzirom da učenici nisu još upoznati s normalnom distribucijom, profesori na prethodna pitanja mogu očekivati jednostavnije odgovore koji se oslanjaju na intuiciju i osnovne matematičke operacije. Od učenika bi mogli očekivati niže napisani odgovori.

Primjer 3.3.1. *[10] Uzgajivač lososa ulovi stotine odraslih riba. Bilježi njihove težine u tablici u intervalima* $3 \leq w \leq 3.1$ *kg,* $3.1 \leq w \leq 3.2$ *kg,* $3.2 \leq w \leq 3.3$ *kg... Prosječna težina ribe je* 4.73 *kg, a standardna devijacija je* 0.53 *kg. Histogram podataka ima oblik zvona i simetričan je oko srednje vrijednosti. Razmislimo:*

a. *Možemo li koristiti samo srednju vrijednost i standardnu devijaciju kako bismo pronašli udio lososa čija je težina:*
   i. *veća od 6 kg,*
   ii. *između 4 kg i 6 kg?*

b. *Kako možemo pronaći težinu:*
   i. *ispod koje teži 90% lososa,*
   ii. *iznad koje teži 25% lososa?*

Rješenje:

a. i. ”Ako uzmemo u obzir da je prosječna težina 4.73 kg i da se težine ne razlikuju puno od te vrijednosti, moglo bi biti jako malo lososa čija je težina veća od 6 kg.”

ii. "Prosječna težina lososa iznosi približno 4.73 kg, što sugerira da se mnogi lososi nalaze u rasponu od 4 kg do 6 kg, to je razuman raspon oko prosječne težine."

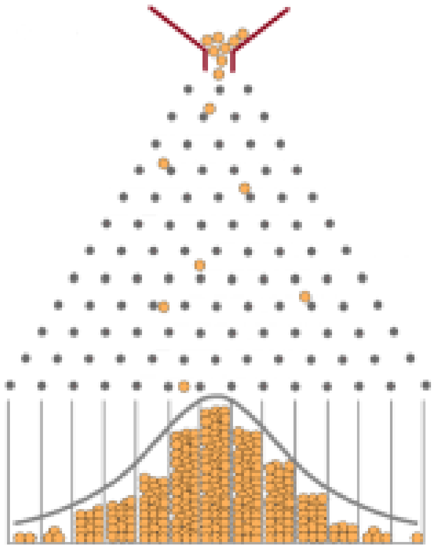
b. i. "Većina lososa ima težinu oko 4.73 kg, pa bi težina ispod koje teži 90% lososa mogla biti nešto malo veća od 4.73 kg, jer će manji broj lososa imati puno veću težinu."

ii. "Budući da je prosječna težina 4.73 kg, težina iznad koje teži 25% lososa mogla bi biti nešto veća od prosjeka, možda malo iznad 5 kg."

Nakon što se uvedu svi potrebni pojmovi, profesor se s učenicima može vratiti na motivacijski primjer kako bi učenici vidjeli je li njihovo razmišljanje bilo ispravno.

Aktivnost 4: Galtonova daska

Cilj aktivnosti: Učenici će se upoznati s pojmom normalne raspodjele koristeći Galtonovu dasku kao praktičan primjer.
Ishodi učenja: Učenici koriste Galtonovu dasku za ilustraciju normalne raspodjele i analiziraju rezultate.
Materijali: Galtonova daska (stvarna ili simulirana), papir i olovka
Nastavni oblik: rad u grupi (cijeli razred)
Nastavna metoda: metoda dijaloga, metoda eksperimenta
Tijek aktivnosti: Nastavnik započinje aktivnost predstavljanjem Galtonove daske. Objašnjava kako će se koristiti za vizualizaciju i istraživanje raspodjele podataka. Ako je moguće, nastavnik koristi stvarnu Galtonovu dasku ili računalnu simulaciju. Učenici promatraju kako kuglice prolaze kroz dasku i bilježe rezultate. Svaki učenik bilježi broj kuglica koje su završile u svakom od dostupnih odjeljaka na dnu daske. Nastavnik vodi raspravu o obrascima koje učenici uoče na dnu daske. Učenici primjećuju da kuglice formiraju određeni obrazac prilikom raspoređivanja. Raspravlja se o tome što učenici primjećuju, poput simetrije i oblika raspodjele. Nakon što učenici uoče da raspodjela kuglica ima oblik sličan zvonu, nastavnici uvode pojam normalne raspodjele. Objašnjavaju kako normalna raspodjela opisuje raspodjelu podataka koja se oblikuje u obliku zvona kada se podaci prikupe u velikim količinama. Nastavnik objašnjava osnovne karakteristike normalne raspodjele, uključujući simetriju oko srednje vrijednosti i kako standardna devijacija utječe na oblik raspodjele. Na kraju, nastavnik povezuje rezultate pokusa s matematičkim pojmovima. Objašnjava kako Galtonova daska ilustrira karakteristike normalne raspodjele i kako se srednja vrijednost i standardna devijacija koriste za opisivanje raspodjele. Pruža dodatne materijale, uključujući grafove normalne raspodjele i Z-tablice, kako bi učenici bolje razumjeli ove pojmove. Na kraju aktivnosti, nastavnik organizira refleksiju i raspravu o tome što su učenici naučili.



Slika 3.7: Galtonova daska [13]

Aktivnost 5: Računanje vjerojatnosti s kalkulatorom

Cilj aktivnosti: Učenici će korištenjem kalkulatora računati vjerojatnosti koje su normalno distribuirane.
Ishodi učenja: Učenici će pomoću kalkulatora izračunati vjerojatnost da varijabla padne unutar zadanog intervala, te pravilno interpretirati dobivene rezultate.
Materijali: kalkulator, tablice za normalnu distribuciju za svakog učenika, papir i olovka, radni listići
Nastavni oblik: individualan rad
Nastavna metoda: metoda eksperimenta
Tijek aktivnosti: Nastavnik započinje aktivnost kratkim uvodom u kojem s učenicima ponavlja osnovne pojmovove vezane uz normalnu distribuciju, uključujući srednju vrijednost i standardnu devijaciju. Nakon toga, nastavnik učenicima objašnjava kako su se prije korištenja kalkulatora upotrebljavale tablice za normalnu distribuciju. U ovom dijelu, nastavnik može na ploči ili projektoru prikazati jedan primjer kako koristiti tablice za normalnu distribuciju. U središnjem dijelu nastavnog sata, nastavnik demonstrira kako koristiti kalkulator za izračunavanje vjerojatnosti normalno distribuiranih varijabli. Nastavnik učenicima daje konkretan primjer problema i korak po korak ih vodi kroz rješavanje tog problema koristeći kalkulator. Tijekom ove faze, učenici će imati priliku postavljati pitanja i dobiti potrebnu pomoć dok prolaze kroz primjer. Nakon vođenog rada, učenici su sposobni raditi samostalno. Dobit će radne listove s različitim zadacima koji uključuju

izračunavanje vjerojatnosti za normalno distribuirane varijable. Učenici će koristiti kalkulatore i tablice za normalnu distribuciju za rješavanje zadatka na radnim listovima. Samostalni rad omogućuje učenicima da primijene naučeno. Na kraju aktivnosti, potrrebno je organizirati kratku diskusiju kako bi se razgovaralo o rješavanju zadataka, uključujući moguće izazove i rješenja koja su pronađena. Nastavnik će na taj način provjeriti točnost rješenja i pružiti povratne informacije.

Primjer 3.3.2. *[10] X je slučajna varijabla koja je normalno distribuirana s srednjom vrijednošću 70 i standardnom devijacijom 4. Pronađite:* $\mathbb{P}(70 \leq X \leq 74)$.

Rješenje: *(za CASIO fx-991EX)* Koraci: Menu - 7: Distribution - 2: Normal CD

Lower: 70 (stisnuti znak " = " nakon svakog retka)
Upper: 74
σ: 4
μ: 70

Rješenje: $P = 0.3413447461 \approx 0.341$.

U nastavku još neki primjeri zadataka za učenike:

Primjer 3.3.3. *[10] X je slučajna varijabla koja je normalno distribuirana s srednjom vrijednošću 60 i standardnom devijacijom 5. Pronađite:*

a. $\mathbb{P}(60 \leq X \leq 65)$

b. $\mathbb{P}(62 \leq X \leq 67)$

c. $\mathbb{P}(X \geq 64)$

d. $\mathbb{P}(X \leq 68)$

e. $\mathbb{P}(X \leq 61)$

f. $\mathbb{P}(57.5 \leq X \leq 62.5)$

Bibliografija

[1] L. Kralj, *Nastava vjerojatnosti u osnovnoj školi*, Miš, 2006.

[2] I. Fundurulić, *Diplomski rad: Problem uvođenja vjerojatnosti i vjerojatnosnog prostora u nastavi matematike prirodoslovno-matematičkih gimnazija*, PMF-MO, 2009.

[3] N. Sarapa, *Teorija vjerojatnosti*, Školska knjiga, 2002.

[4] S. Loparić, *Vjerojatnost i statistika - zašto, kada, kako?*, Poučak, 2019.

[5] Z. Šikić, S. Antoliš, A. Copić, R. Kalazić, S. Lukač, E. Špalj, *Matematika 4, udžbenik za četvrti razred gimnazije i srednje strukovne škole, 2. svezak*, ProfilKlett, 2021., Pristupljeno: 02.07.2024.

[6] Z. Šikić, A. Copić, R. Kalazić, S. Lukač, K.J. Penzar, *Matematika 2, udžbenik za drugi razred gimnazije i srednje strukovne škole, 2. svezak*, ProfilKlett, 2021., Pristupljeno: 03.07.2024.

[7] G. Pajić, Ž. Bošnjak, B. Čulina, N. Grgić, *Matematički izazovi 8, udžbenik sa zadatcima za vježbanje za osmi razred osnovne škole, prvi dio*, Alfa, 2023., Pristupljeno: 26.08.2024.

[8] D. Glasnović Gracin, G. Žokalj, T. Soucie, *Otkrivamo matematiku 4, radni udžbenik iz matematike za četvrti razred osnovne škole, prvi dio*, Alfa, 2023., Pristupljeno: 26.08.2024.

[9] D. Glasnović Gracin, G. Žokalj, T. Soucie, *Otkrivamo matematiku 2, radni udžbenik iz matematike za drugi razred osnovne škole, prvi dio*, Alfa, 2023., Pristupljeno: 26.08.2024.

[10] D. Martin, R. Haese, S. Haese, M. Haese, M. Humphries, *Matematics for the international student – Matematics HL (Core), third edition*, Haese Mathematics, 2012.

[11] N. Adžaga, A. Martinčić Špoljarić, N. Sandrić, *Vjerojatnost i statistika*, Građevinski fakultet, Sveučilište u Zagrebu, 2017., Pristupljeno: 02.07.2024.

[12] M. Benšić, N. Šuvak, *Uvod u vjerojatnost i statistiku*, Odjel za matematiku, Sveučilište u Osijeku, 2014., Pristupljeno: 26.08.2024.

[13] H. Mesić, *Galtonova daska*, Pristupljeno: 02.09.2024.

[14] I. Matić, Lj. Jukić Matić, M. Zelčić, M. Šujansky, T. Srnec, T. Vukas, Ž. Dijanić *Matematika 4, udžbenik matematike u četvrtom razredu srednje škole sa zadacima za rješavanje, 2. dio*, Školska knjiga, 2024., Pristupljeno: 26.08.2024.

[15] S. Gyu Kwak, J. Hae Kim, *Central limit theorem: the cornerstone of modern statistics*, Korean Journal of Anesthesiology, 2017., Pristupljeno: 02.09.2024.

[16] G. Sanderson, *But what is the Central Limit Theorem?*, 3blue1brown, 2023., Pristupljeno: 03.09.2024.

[17] Z. Šikić, S. Antoliš, A. Copić, R. Kalazić, S. Lukač, E. Špalj, *Matematika 4, udžbenik za četvrti razred gimnazije i srednje strukovne škole, 1. svezak*, ProfilKlett, 2021., Pristupljeno: 02.07.2024.

[18] Materijali iz kolegija *Vjerojatnost i statistika*, PMF-MO, 2019/2020.

Sažetak

Ovaj diplomski rad istražuje temeljne aspekte vjerojatnosti i statistike kroz analizu osnovnih pojmova, teorijskih koncepata i njihovih primjena u obrazovanju. U prvom poglavlju razmatramo definicije vjerojatnosti, uključujući skupove elementarnih događaja te osnovna svojstva vjerojatnosnog prostora. Posebna pažnja posvećena je razumijevanju i formaliziranju vjerojatnosti kao matematičkog koncepta, što stvara čvrst temelj za daljnju analizu. Drugo poglavlje fokusira se na slučajne varijable, razlikujući diskretne i neprekidne slučajne varijable te naglašavajući njihove karakteristike, distribucije i primjene. Također, razmatra se povezivanje ovih varijabli na primjeru bacanja kockice, čime se ilustriraju njihove zajedničke i specifične aspekte. Centralni granični teorem, kao ključni koncept u statistici, omogućuje razumijevanje ponašanja sumiranih slučajnih varijabli i aproksimaciju normalnih distribucija. Na kraju, istražuje se primjena vjerojatnosti i statistike u hrvatskom obrazovnom kurikulumu te predlažemo aktivnosti za unapređenje nastave matematike. Ovaj rad pruža sveobuhvatan pregled ključnih koncepata vjerojatnosti i statistike, s naglaskom na njihovu primjenu u obrazovanju, čime doprinosi boljem razumijevanju i primjeni statističkih metoda u različitim kontekstima.

Summary

This thesis explores the fundamental aspects of probability and statistics through the analysis of basic concepts, theoretical principles, and their applications in education. The first chapter examines definitions of probability, including the sets of elementary events and the basic properties of probability spaces. Special attention is given to understanding and formalizing probability as a mathematical concept, providing a solid foundation for further analysis. The second chapter focuses on random variables, distinguishing between discrete and continuous random variables while emphasizing their characteristics, distributions, and applications. Additionally, it explores the connection between these variables using the example of rolling a die, illustrating their common and specific aspects. The Central Limit Theorem, as a key concept in statistics, facilitates the understanding of the behavior of summed random variables and the approximation of normal distributions. Finally, the application of probability and statistics in the Croatian educational curriculum is examined, and activities are proposed to enhance mathematics teaching. This thesis offers a comprehensive overview of key probability and statistics concepts, with an emphasis on their application in education, contributing to a better understanding and application of statistical methods in various contexts.

Životopis

Rođena sam 28. prosinca 1996. godine u Splitu. Osnovnoškolsko obrazovanje završavam u Osnovnoj školi Ivana Mažuranića u Obrovcu Sinjskom te potom upisujem opći gimnazijski smjer u Općoj gimnaziji Dinka Šimunovića u Sinju. Godine 2017. upisujem Preddiplomski sveučilišni studij Matematika na Prirodoslovno-matematičkom fakultetu u Zagrebu; smjer nastavnički. Godine 2021. stječem naziv sveučilišne prvostupnice edukacije matematike te iste godine upisujem Diplomski sveučilišni studij Matematika; smjer nastavnički. Tijekom preddiplomskog studija zapošljavam se u Zagrebačkoj banci u odjelu *Proizvodi financiranja*, a potom prelazim u odjel *Aplikativni razvoj kreditne aplikacije*. Također, tijekom diplomskog studija sudjelovala sam na projektu *„Financijska pismenost u suvremenom matematičkom obrazovanju“* te sam u sklopu studija odradila metodičku praksu nastave matematike u Osnovnoj školi Dobriše Cesarića i XV. gimnaziji u Zagrebu (redovni i IB program).