SVEUČILIŠTE U RIJECI

TEHNIČKI FAKULTET

Preddiplomski sveučilišni studij elektrotehnike

Završni rad

SIMULACIJE SLUČAJNIH VARIJABLI

Rijeka, rujan 2022.

Ivan Frković
0069080734

SVEUČILIŠTE U RIJECI

TEHNIČKI FAKULTET

Preddiplomski sveučilišni studij elektrotehnike

Završni rad

SIMULACIJE SLUČAJNIH VARIJABLI

Mentor: doc. dr. sc. Ivan Dražić

Komentor: prof. dr. sc. Viktor Sučić

Rijeka, rujan 2022.

Ivan Frković
0069080734

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
POVJERENSTVO ZA ZAVRŠNE ISPITE

Rijeka, 5. ožujka 2021.

Zavod: **Zavod za matematiku fiziku, strane jezike i kineziologiju**
Predmet: **Inženjerska matematika ET**
Grana: **1.01.08 teorija vjerojatnosti i statistika**

ZADATAK ZA ZAVRŠNI RAD

Pristupnik: **Ivan Frković (0069080734)**
Studij: Preddiplomski sveučilišni studij elektrotehnike

Zadatak: **Simulacije slučajnih varijabli // Simulations of random variables**

Opis zadatka:

U radu je potrebno precizno definirati slučajne varijable i opisati njihova temeljna svojstva. Potrebno se osvrnuti na klasifikaciju slučajnih varijabli prema skupu vrijednosti i pripadne stohastičke modele povezane s inženjerskom strukom, s posebnim naglaskom na primjene u elektrotehnici. Zatim je potrebno opisati temeljne ideje algoritama za simuliranje slučajnih varijabli, uz njihovu implementaciju u nekom programskom jeziku. Sve opisane algoritme potrebno je analizirati s obzirom na robusnost, preciznost i brzinu izvođenja.

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

Zadatak uručen pristupniku: 15. ožujka 2021.

Mentor:

Doc. dr. sc. Ivan Dražić

Prof. dr. sc. Viktor Sučić (komentor)

Predsjednik povjerenstva za završni ispit:

Prof. dr. sc. Viktor Sučić

IZJAVA

Sukladno članku 8. Pravilnika o završnom radu, završnom ispitu i završetku preddiplomskih sveučilišnih studija Tehničkog fakulteta Sveučilišta u Rijeci od 1. veljače 2020., izjavljujem da sam samostalno izradio završni rad prema zadatku preuzetom dana 15. ožujka 2021.

Rijeka, 19. rujna 2022.

Ivan Frković

Veliku zahvalu želim uputiti mome mentoru doc. dr. sc. Ivan Dražiću. Želim mu se posebno zahvaliti na njegovoj predanosti i angažiranosti u radu sa nama studentima. Svoj završni rad prvenstveno posvećujem svojoj obitelji bez koje bi sve do sada postignuto bilo teško ili pak nemoguće i jednoj meni veoma voljenoj osobi koja nažalost nije više među nama.

Sadržaj

1. Uvod

Vjerojatnost kao takva jedna je od najprimjenjenijih matematičkih disciplina suvremenog svijeta. Ona utječe na "velike", a i na "male" životne odluke. Bilo da se radi o operaciji gdje mi kao ljudi dajemo određenu suglasnost ili dopuštenje za izvođenje zahvata gdje se liječnik ograđuje ako nešto što je lako "vjerojatno" pođe po zlu ili da se radi o planiranju vikenda sa obitelji gdje je vrlo "moguće" da padne koja kapljica kiše ili će ipak zasjati sunce. Vrlo bitan udio ima i u gospodarstvu gdje zbog određenih zbivanja na svjetskoj političkoj sceni može doći do pada ili rasta vrijednosti određenog proizvoda ili robe. Ne navodeći dalje primjere jer ih ima bezbroj vidimo da skoro svaki dio naših života je isprepleten određenom vjerojatnošću zbivanja događaja. U prošlosti se određena vjerojatnost događaja morala utvrđivati ponavljanjem pokusa da bi se došlo do zaključka, ali danas uz pomoć suvremenih tehnologija taj segment je uvelike olakšan jer je moguće virtualno simulirati određene situacije te se puno lakše prognoziraju određeni ishodi

U ovome radu opisati ćemo temeljne pojmove koje vežemo uz teoriju vjerojatnosti, proučavati diskretnu slučajnu varijablu te sistematizaciju razdioba diskretne slučajne varijable. Odabrane primjere iz područja teorije vjerojatnosti riješiti ćemo na dva načina: standardno odnosno analitički te uz pomoć računalnog programa Phyton. Svaki od navedenih načina će biti detaljno opisani.

Rad se sastoji od četiri poglavlja.

U drugom poglavlju dan je uvod sa osnovnim pojmovima kojima se problematika teorije vjerojatnosti bavi te kratak povjesni pregled od nastanka do danas.

U trećem poglavlju definirali smo slučajnu varijablu. S obzirom na njezinu podjelu na slučajnu varijablu kontinuiranog tipa te varijablu diskretnog tipa objašnjene su temeljne razlike između ove dvije vrste te naposljetku i definirana diskretna slučajna varijabla.

U četvrtom poglavlju analizirali smo vrste razdioba diskretne slučajne varijable.

U petom poglavlju opisali smo algoritme za simuliranje nekih slučajnih varijabli te njihovu implementaciju unuta programskog jezika Phyton.

2. Uvod u teoriju vjerojatnosti

Cardano i Galilei u 16 st. pročavaju igre na sreću i tako postavljaju temelj teoriji vjerojatnosti kojoj malo kasnije Pascal i Fermat postavljaju matematičku osnovu. Od toga trenutka vjerojatnost kao teorija se razvija i tretira kao samostalna matematička grana. Prve njezine primjene bile su u igrama na sreću, dok se u industriji počinje primjenjivati od kraja 19. stoljeća, a početkom 20. stoljeća doživljava primjenu u svim znanstvenim i proizvodnim granama [1].

Osnovni pojam teorije vjerojatnosti je događaj, a događaji za koje je neminovno da će se dogoditi nazivaju se sigurni događaji. Postoje i događaji koji se mogu, ali i ne moraju dogoditi. Pri bacanju igraće kocke ona će sigurno pasti, ali hoće li pasti paran ili neparan broj ne možemo znati. Ova vrsta događaja naziva se slučajnim događajima. Dođagaji koji se nikada ne mogu dogoditi nazivaju se nemogućim događajima. Tako je primjerice događaj da će na igraćoj kocki pasti broj 9 nemoguć događaj.

Upravo se na osnovu eksperimenata te slučajevima koji proizlaze ponavljanjem eksperimenata razmatra vjerojatnost nekog događaja. Možemo reći da je vjerojatnost kvantifikacija šanse da će se neki događaj dogoditi.

Kako bi mogli matematički korektno definirati pojam vjerojatnosti prvo moramo uvesti pojam jedne matematičke strukture koja se zove σ-algebra. Tu strukturu koristimo kako bismo modelirali događaje, drugim riječima događaji će biti skupovi koji će imati strukturu σ-algebre, a oni će biti podskupovi nekog skupa ishoda.

Definicija 2.1. *Neka je U skup objekata koje u teoriji vjerojatnosti zovemo elementarnim događajima. Skup S neka je neprazni skup podskupova U te neka S ima sljedeća svojstva:*

1. $U \in S$,
2. *Ako je* $A \in S \rightarrow \bar{A} \in S$, $\left(A \cup \bar{A} = U \right)$,
3. *Ako je* $A_i \in S \rightarrow \bigcup_i A_i \in S$ *za* $i = 1, 2, ...$

Takav skup S (neprazna familija podskupova skupa U), koji zadovoljava gore navedena svojstva zove se ***σ-algebra skupova*** *ili* ***Borelovo polje****.*

Sad možemo definirati samu vjerojatnost.

Definicija 2.2. *Neka je P funkcija definirana na σ-algebri S podskupova skupa U. Tu ćemo funkciju zvati vjerojatnost ako zadovoljava sljedeće uvjete:*

1. $P\left(U\right) = 1$,

2. $P(A) \geq 0$ *za svaki* $A \in S$,

3. *Ako su* $A_i \in S$ *za* $i = 1, 2, 3, ...$, *događaji koji se međusobno isključuju, tj. za koje je* $A_i \cap A_j = \emptyset$ *ako* $i \neq j$, *za njih vrijedi*

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i). \tag{2.1}$$

Uvjeti navedeni u prethodnoj definiciji zovu se aksiomi vjerojatnosti ili aksiomi Kolomogorova. Tako vrijednost funkcije P na skupu $A \in S$ zovmo vjerojatnost događaja A.

Često se koristi i sljedeća definicija.

Definicija 2.3. *Uređenu trojku* (U, S, P) *koja se sastoji od skupa ishoda* U, σ*-algebre događaja* S *i funkcije vjerojatnosti* P *definirane na* S *zovemo* ***prostor vjerojatnosti****.*

Prikažimo sada na jednom primjeru definiciju vjerojatnosti koja se vrlo često koristi, a naziva se klasična definicija vjerojatnosti.

Primjer 2.1. *Neka su* $A_1, A_2, ..., A_n$ *jednako vjerojatni događaji, koji se međusobno isključuju i čine potpuni sustav događaja, tako da vrijedi*

$$\bigcup_{i=1}^{n} A_i = U. \tag{2.2}$$

Ovime smo definirali jedan skup ishoda U*. Ovaj skup se još nazivam i skupom elementarnih događaja. Jasno je da vrijedi*

$$P(A_i) = \frac{1}{n}, \tag{2.3}$$

za $\forall = 1, ..., n.$

Skup događaja S*, odnosno* σ*-algebra događaja* S *sastojati će se od događaja* A *koji se mogu prikazati kao unija od* m *elementarnih događaja iz skupa* S*.*

Prema zadnjem aksiomu vjerojatnosti, vjerojatnost unije od m *jednakovjerojatnih događaja vjerojatnosti* $\frac{1}{n}$ *je:*

$$P(A) = \frac{m}{n}, \tag{2.4}$$

čime je definirana ***klasična definicija vjerojatnosti****.*

Prema terminologiji klasične teorije vjerojatnosti, elementarni događaji koji definiraju događaj A nazivaju se "povoljni ishodi" za događaj A, pa prema toj terminologiji klasičnu definiciju vjerojatnosti obrazlažemo na sljedeći način:

Vjerojatnost ostvarenja nekog događaja A *jednaka je kvocijentu broja povoljnih ishoda* m *i svih mogućih ishoda* n*.*

Primijetimo da je za klasičnu definiciju vjerojatnosti bilo nužno imati konačan skup jednakovjerojatnih ishoda što ponekad može biti značajno ograničenje.

Sada je jasno da će događaj opisan kao unija većeg broja elementarnih događaja imati veću vjerojatnost. Ako niti jedan elementarni događaj ne opisuje događaj A, tj. ako je $m(A) = 0$, biti će i $P(A) = 0$. To znači da događaju koji je nemoguć pripada i vjerojatnost jednaka 0.

Ako je $m(A) = n$ tj. ako svaki ishod realizira događaj A, biti će $P(A) = 1$. Tako da sigurnom događaju odgovara vjerojatnost jednaka 1. Za svaki drugi događaj je $0 < P(A) < 1$.

Budući se opisana vjerojatnost može izračunati bez izvođenja bilo kakvog pokusa na temelju nekog modela, kažemo da je dobivena teoretska ili matematička vjerojatnost događaja A ili vjerojatnost "*a priori*".

Često se definiranju vjerojatnosti pristupa na drugi način - eksperimentalno. U tom se slučaju bilježi broj realizacija događaja A koji se označava sa m, a broj izvršenih pokusa s n. Tada je omjer $\frac{m}{n}$ zapravo relativna frekvencija događaja A, poznata kao vjerojatnost "*a posteriori*" ili statistička vjerojatnost.

Kada broj promatranja raste, vjerojatnost "*a priori*" i "*a posteriori*" se sve manje razlikuju. J. Bernoulli formulirao je tu činjenicu u obliku svog zakona "Velikih brojeva". On kaže da: "Kad broj pokusa raste, apsolutna razlika između relativne frekvencije i vjerojatnosti opada."

Navedene definicije i teze prema [1]

Navedimo sada nekoliko primjera korištenja klasične definicije vjerojatnosti.

Primjer 2.2. *Promotrimo problem bacanja igraće kocke i odredimo kolika je vjerojatnost pada parnog broj te broja djeljivog s 3.*

Jasno je da će skup ishoda u tom slučaju biti

$$U = \{1, 2, 3, 4, 5, 6\}. \tag{2.5}$$

Događaj A "pao je paran broj" može se opisati sljedećim skupom ishoda:

$$A = \{2, 4, 6\}, \tag{2.6}$$

dok se događaj B "pao je broj djelji s 3 opisuje skupom

$$B = \{3, 6\}. \tag{2.7}$$

Sada je jasno da vrijedi

$$P(A) = \frac{3}{6} = \frac{1}{2}, \tag{2.8}$$

budući je bilo ukupno 6 ishoda, a povoljna su 3 ishoda. Analogno dobivamo

$$P(B) = \frac{2}{6} = \frac{1}{3}. \tag{2.9}$$

Primjer 2.3. *Iz hrpe od* 52 *karte izvlači se na sreću po jedna karta. Skup ishoda u tom slučaju će biti sve karte, tj. unija četiri skupa po* 13 *karata, odnosno skupova pika, kare, herca i trefa; što nam daje sve ukupno* 52 *karte.*

a) *Odredimo vjerojatnost događaja* A *koji znači da smo izvukli asa. On je definiran skupom koji čine* 4 *povoljna događaja. Imamo* 4 *skupine simbola u špilu (tref, herc, karo, pik) i svaki skup simbola sadrži po jednog asa te sukladno tome možemo reći da imamo* 4 *povoljna ishoda, a* 52 *moguća ishoda.*

   Sada je jasno da vrijedi

$$P(A) = \frac{4}{52} = \frac{1}{13}. \tag{2.10}$$

b) *Odredimo vjerojatnost događaja* B *koji znači da smo izvukli pika. On je definiran skupom od* 13 *karata u špilu u simbolu pika. Tako imamo 13 povoljnih ishoda, a* 52 *moguća ishoda.*

   Sada je jasno da vrijedi

$$P(A) = \frac{13}{52} = \frac{1}{4}. \tag{2.11}$$

c) *Odredimo vjerojatnost događaja* C *koji znači da smo izvukli pika asa. Definiran je skupom koji čini jedan ishod, odnosno u špilu postoji samo jedan pik as. Znači imamo jedan povoljan ishod, a* 52 *moguća ishoda. Sada je jasno da vrijedi*

$$P(A) = \frac{1}{52}. \tag{2.12}$$

Pokažimo sada na jednom primjeru korištenje statističke definicije vjerojatnosti.

Primjer 2.4. *U ožujku 1953. u Monte Carlu izvršena su* 4894 *pokusa u kojima je kuglica pala na crno* 2396 *puta, na crveno* 2360 *puta, a na bijelo* 138 *puta. Iz tih podataka zaključujemo da prema statističkoj definiciji vjerojatnosti vrijedi*

$$P_{crno} = \frac{2396}{4894} = 0.48958, \tag{2.13}$$

$$P_{crveno} = \frac{2360}{4894} = 0.4822, \tag{2.14}$$

$$P_{bijelo} = \frac{139}{4894} = 0.282. \tag{2.15}$$

Odredimo sada ove vjerojatnosti korištenjem klasične definicije vjerojatnosti. Polje ruleta podijeljeno je na 18 *crvenih,* 18 *crnih i* 1 *bijelo polje. Vjerojatnost a priori da će kuglica pasti na crveno, odnosno crno polje jednaka je*

$$\frac{18}{37} = 0.486, \tag{2.16}$$

a za bijelo

$$\frac{1}{37} = 0,027. \tag{2.17}$$

Ovime smo i demonstrirali zakon velikih brojeva pokazavši da se vjerojatnost dobivena a priori i a posteriori vrlo malo razlikuje ako je broj ponavljanja eksperimenta veći.

Pokažimo sada primjenu vjerojatnosti na jednom primjeru iz inženjerske struke.

Primjer 2.5. *Pri montaži se upotrebljavaju vijci nabavljeni od dva proizvođača. Prema deklaracijama prvi je izradio* 3 *posto nestandardnih, a drugi* 5 *posto nestandardnih elemenata. Od prvog dobavljača je nabavljeno* 1000, *a od drugog* 8000 *komada. Odredimo vjerojatnost da će slučajno upotrijebljeni vijak biti nestandardan.*

Primjetimo najprije da je ukupan broj vijaka, odnosno ishoda jednak $n = 10000 + 8000 = 18000$.

S druge strane, broj nestandardnih vijaka, odnoso povoljnih ishoda je $m = 0.3 \cdot 10000 + 0.05 \cdot 8000 = 700$.

Sada možemo izračunati vjerojatnost da je slučajno odabrani vijak nestandardan:

$$P(A) = \frac{700}{18000} \approx 0.39 = 39\%. \tag{2.18}$$

Analitički primjeri riješeni u radu napravljeni su prema [1].

3. Slučajne varijable

Svaki događaj A iz S je ishod promatranja ili eksperimenta i može biti okarakteriziran realnim brojem x. Na taj način moguće je preći iz algebre događaja u algebru podskupova od R, ali pritom se prenosi i sama vjerojatnost. Takvu vrstu preslikavanja nazivamo **slučajnom varijablom**.

Kako bi slučajna varijabla bila određena potrebno je odrediti skupove vrijednosti koje ona može poprimiti te potrebno je odrediti i vjerojatnosti povezane sa vrijednostima koje poprima.

Broj vrijednosti koje ona može primiti može biti prebrojiv ili neprebrojiv. Razmatrajući prebrojiv broj vrijednosti slučajne varijable govorimo o **diskretnoj slučajnoj varijabli** ili **varijabli diskretnog tipa**. Neki od primjera iz svakodnevnog života su npr. broj padova šestice pri ponovljenom bacanju kocke, broj putnika u autobusu itd.

Slučajna varijabla je **kontinuiranog tipa** ako s pozitivnom vrijednošću prima vrijednosti određenog intervala realnih brojeva. Npr. potrošnja goriva na 100 km prijeđenog puta, trajanje radne opreacije itd.

3.1. Diskretna slučajna varijabla

Neka je $(\Omega, \mathcal{F}, P)$ **diskretan** prostor vjerojatnosti, što znači da je prostor ishoda

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_k, \ldots\} \tag{3.1}$$

diskretan (konačan ili prebrojiv). S ω_k označili smo elementarne događaje, odnosno ishode promatranog slučajnog pokusa.

Diskretna slučajna varijabla je preslikavanje $X : \Omega \to \mathbb{R}$ koje svakom elementarnom događaju $\omega \in \Omega$ pridruži realan broj x_k, tj. vrijedi

$$X(\omega) = x_k. \tag{3.2}$$

Događaje $\{\omega \in \Omega : X(\omega) = x_k\}$ poistovjetit ćemo s izrazom $(X = x_k)$, dok ćemo pripadajuće vjerojatnosti označavati s

$$p_k = P(X = x_k). \tag{3.3}$$

Skup vrijednosti $\mathcal{R}(X) = \{x_1, x_2, \ldots, x_k, \ldots\} \subseteq \mathbb{R}$ diskretne slučajne varijable X je diskretan (konačan ili prebrojiv). Za proizvoljan podskup $R_x \subseteq \mathcal{R}(X)$ uvedimo oznaku:

$$(X \in R_x) = \{\omega \in \Omega : X(\omega) \in R_x\}. \tag{3.4}$$

Jasno je da je $(X \in R_x)$ podskup prostora događaja Ω za koje vrijednosti slučajne varijable pripadaju skupu R_x, te je $(X \in R_x)$ događaj koji pripada algebri događaja $\mathcal{F}$.

Zakon razdiobe diskretne slučajne varijable X uobičajeno pišemo u sljedećem obliku:

$$X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_k & \cdots \\ p_1 & p_2 & \cdots & p_k & \cdots \end{pmatrix}, \tag{3.5}$$

pri čemu u prvom retku navodimo vrijednosti koje slučajna varijabla može poprimiti, a u drugom retku pripadajuće vjerojatnosti. Pritom mora vrijediti:

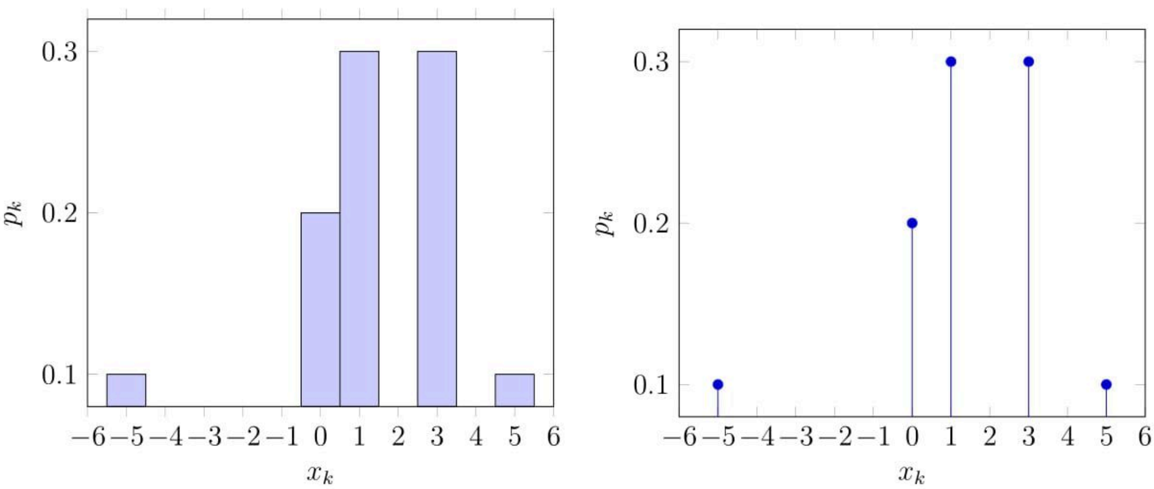
$$\sum_{k=1}^{\infty} p_k = 1.$$

Za slučajnu varijablu zadanu razdiobom (3.5), funkciju f za koju vrijedi

$$f(x_k) = P(X = x_k) = p_k$$

nazivamo **funkcija gustoće slučajne varijable** ili **funkcija vjerojatnosti**.

Funkcija gustoće kod diskretnih slučajnih varijabli najčešće se prikazuje histogramom ili točkastim dijagramom kako je prikazano na sljedećoj slici.



(a) Histogramski prikaz

(b) Prikaz točkastim dijagramom

Slika 3.1. Funkcija gustoće slučajne varijable X.

Primjer 3.1. *Računalo generira tri binarne znamenke. Neka je X slučajna varijabla koja opisuje broj generiranih jedinica, a Y slučajna varijabla koja opisuje poziciju prve generirane jedinice.*

Odredite prostor ishoda te za svaki ishod odredite vrijednosti slučajnih varijabli X i Y i zapišite njihove zakone razdioba.

Prostor ishoda je $\Omega = \{000, 001, 010, 011, 100, 101, 110, 111\}$. U tablici su dane vrijednosti koje poprimaju slučajne varijable X i Y za svaki od ishoda.

ishod	$X(\omega_k)$	$Y(\omega_k)$
$\omega_1 = 000$	*0*	*0*
$\omega_2 = 001$	*1*	*3*
$\omega_3 = 010$	*1*	*2*
$\omega_4 = 011$	*2*	*2*
$\omega_5 = 100$	*1*	*1*
$\omega_6 = 101$	*2*	*1*
$\omega_7 = 110$	*2*	*1*
$\omega_8 = 111$	*3*	*1*

Iz tablice vidimo da slučajna varijabla X poprima vrijednosti iz skupa $\mathcal{R}(X) = \{0, 1, 2, 3\}$, pri čemu je $(X = 0) = \{\omega_1\}$, $(X = 1) = \{\omega_2, \omega_3, \omega_5\}$, $(X = 2) = \{\omega_4, \omega_6, \omega_7\}$ te $(X = 3) = \{\omega_8\}$. S obzirom da su svi ishodi jednako vjerojatni, vidimo da vrijedi $P(X = 0) = \frac{1}{8}$, $P(X = 1) = \frac{3}{8}$, $P(X = 2) = \frac{3}{8}$ i $P(X = 3) = \frac{1}{8}$. Stoga je zakon razdiobe slučajne varijable X

$$X \sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{pmatrix}. \tag{3.6}$$

Slično dobivamo

$$Y \sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{8} & \frac{4}{8} & \frac{2}{8} & \frac{1}{8} \end{pmatrix}. \tag{3.7}$$

3.2. Neprekidna slučajna varijabla

Promatrajmo pokus u kojem točku slučajno biramo s intervala duljine L. Udaljenost točke do ruba intervala je slučajna veličina. S obzirom da ta udaljenost može biti jednaka bilo kojoj vrijednosti s intervala $\left[0, \frac{L}{2}\right]$, očito je da prostor događaja Ω nije prebrojiv. Slijedi da diskretna slučajna varijabla uvedena u prethodnom poglavlju, koja se odnosila samo na diskretan prostor vjerojatnosti, u ovom slučaju ne može biti primjenjiva. Stoga u nastavku uvodimo pojam **slučajne varijable** za proizvoljan prostor vjerojatnosti.

Neka je $(\Omega, \mathcal{F}, P)$ prostor vjerojatnosti. Preslikavanje $X : \Omega \to \mathbb{R}$ je **slučajna varijabla** ako za svaki $x \in \mathbb{R}$ vrijedi da je podskup prostora događaja dan s

$$(X < x) = \{\omega \in \Omega : X(\omega) < x\} \tag{3.8}$$

događaj, tj. element σ-algebre događaja $\mathcal{F}$.

Funkcija distribucije ili **funkcija razdiobe** slučajne varijable X je funkcija $F : \mathbb{R} \to [0, 1]$ definirana s

$$F(x) = P(X < x). \tag{3.9}$$

Intuitivno, za odabrani realan broj x, vrijednost funkcije distribucije $F(x)$ predstavlja vjerojatnost da je vrijednost slučajne varijable X manja od x. $F(x)$ nazivamo još i **kumulativna funkcija distribucije** (*eng. cumulative distribution function, CDF*).

Neka je F funkcija distribucije slučajne varijable X. X je **neprekidna slučajna varijabla** ako postoji nenegativna funkcija $f : \mathbb{R} \to \mathbb{R}$ takva da vrijedi

$$F(x) = \int_{-\infty}^{x} f(t)dt. \tag{3.10}$$

Funkcija $f(x)$ nije nužno neprekinuta, no u točkama neprekinutosti vrijedi

$$f(x) = F'(x). \tag{3.11}$$

Funkciju $f(x)$ nazivamo **funkcija gustoće** ili **funkcija vjerojatnosti**.

Funkcija distribucije je neprekinuta pa vrijedi $P(X = x) = F(x + 0) - F(x) = 0$ za svaki $x \in \mathbb{R}$. Stoga su događaji $(x_1 < X < x_2)$, $(x_1 \leq X < x_2)$, $(x_1 < X \leq x_2)$ i $(x_1 \leq X \leq x_2)$ svi jednako vjerojatni.

Primjer 3.2. *Na autocesti se tehnička i druga pomoć može zatražiti putem SOS telefona uz autocestu. SOS telefoni međusobno su udaljeni 2 kilometra duž autoceste. Pod pretpostavkom da se prilikom vožnje autocestom auto pokvari na slučajnom mjestu, odredimo funkciju koja će opisati vjerojatnost da je udaljenost do najbližeg SOS telefona manja od odabrane vrijednosti, a zatim izračunajmo vjerojatnost da je ta udaljenost manja od* 200 *metara.*

Uočimo najprije da je analizu dovoljno ograničiti na dionicu autoceste na kojoj se pokvario auto, a koja se nalazi između dvaju najbližih SOS telefona. Prema postavkama zadatka duljina te dionice je 2 kilometra, pa je možemo poistovjetiti s intervalom $[0, 2]$. *Pritom točka* $x = 0$ *odgovara poziciji SOS telefona kojeg je auto prošao prije kvara, a* $x = 2$ *poziciji sljedećeg SOS telefona do kojeg nije uspio stići. Udaljenost do najbližeg SOS telefona slučajna je veličina koju ćemo označiti s* X. *Slučajna varijabla* X *zapravo predstavlja udaljenost slučajne pozicije na kojoj se je dogodio kvar do (bližeg) ruba intervala* $[0, 2]$. *Vrijednosti koje slučajna varijabla* X *može poprimiti leže na intervalu* $[0, 1]$, *a za proizvoljan* $x \in [0, 1]$, *za događaj*

$$(X < x) = \text{"udaljenost do najbližeg SOS telefona manja je od } x\text{''},$$

vrijedi da je njegova vjerojatnost proporcionalna duljini unije podintervala $G_x = [0, x\rangle \cup \langle 2 - x, 2]$. *Skup* G_x *obuhvaća točke udaljene od ruba za manje od* x. *U skladu s definicijom funkcije distribucije slijedi*

$$F(x) = P(X < x) = \frac{2x}{2} = x, \quad x \in [0, 1]. \tag{3.12}$$

Za $x < 0$ *vrijedit će* $P(X < x) = 0$, *a za* $x > 1$, $P(X < x) = 1$ *(udaljenost od najbližeg SOS telefona je sigurno manja od* x *za* $x > 1$*), pa je funkcija distribucije slučajne varijable* X *jednaka*

$$F(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}. \tag{3.13}$$

Vrijednost funkcije distribucije $F(x)$ predstavlja vjerojatnost da je udaljenost do najbližeg SOS telefona manja od x (kilometara). Vjerojatnost da je udaljenost manja od 200 metara jednaka je $F(0.2) = 0.2$.

Napomenimo da ćemo se u ovom radu koncentrirati na diskretne slučajne varijable te da je poglavlje uglavnom obrađeno prema izvorima [1], [2] i [3].

4. Diskretne razdiobe slučajne varijable

U ovom se poglavlju detaljnije bavimo diskretnim slučajnim varijablama, odnosno nekim teorijskim modelima povezanim sa diskretnim slučajnim varijablama. Ovo je poglavlje uglavnom obrađeno prema izvoru [6].

4.1. Uniformna razdioba

Kod uniformne je razdiobe svaki od ishoda jednako vjerojatan, pa je uniformna razdioba s n vrijednosti definirana sljedećim zakonom:

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}. \tag{4.1}$$

Primjer 4.1. *Bacanjem kockice slučajni događaj bi bio pad na jedan od brojeva od 1 do 6. Svaki taj broj možemo shvatiti kao vrijednost koju varijabla može poprimiti, a toj vrijednosti pripada i vjerojatnost* $p = \frac{1}{6}$. *Tako određena varijabla je diskretna i očito ima uniformnu razdiobu sa sljedećim zakonom:*

$$X \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}. \tag{4.2}$$

Analitički izražen zakon vjerojatnosti prikazujemo ovako:

$$p(x) = \frac{1}{6}, \tag{4.3}$$

$x = 1, 2, 3, 4, 5, 6.$

4.2. Bernoullijeva razdioba

Neka je A odabrani događaj i neka je vjerojatnost njegove realizacije $p = P(A)$. Bernoullijev pokus je pokus s dva moguća ishoda:

"događaj A se je realizirao" i "događaj A se nije realizirao".

Pridružimo li prvom ishodu vrijednost 1, a drugom ishodu vrijednost 0, definirat ćemo **Bernoullijevu (indikatorsku)** slučajnu varijablu. Njezina je razdioba jednaka

$$X \sim \begin{pmatrix} 1 & 0 \\ p & 1-p \end{pmatrix}. \tag{4.4}$$

Uobičajeno pišemo $X \sim Ber(p)$.

Primjer 4.2. *Novčić se pri bacanju može okrenuti ili na pismo ili na glavu. Ako je novčić koji bacamo pravilan , ti su ishodi jednako vjerojatni. Ako ishod "okrenula se glava" proglasimo uspjehom i označimo ga s 1, a ishod "okrenulo se pismo" proglasimo s neuspjehom i označimo ga s 0, primjećujemo da ishod njegova bacanja opisujemo slučajnom varijablom X s Bernoullijevom razdiobom s parametrom $p = \dfrac{1}{2}$ i tablicom razdiobe*

$$X \sim \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{4.5}$$

4.3. Binomna razdioba

Neka je A odabrani događaj i neka je vjerojatnost njegove realizacije $p = P(A)$. Pretpostavimo da je promatrani slučajni pokus ponovljiv te da su ishodi pokusa u svakom ponavljanju međusobno nezavisni. Pretpostavimo da pokus ponavljamo n puta. Slučajna varijabla X neka predstavlja **broj realizacija događaja A u n ponavljanja pokusa**. Tako definirana slučajna varijabla poprima vrijednosti iz skupa $\{0, 1, \dots, n\}$. Pritom $(X = k)$ predstavlja događaj da se je promatrani događaj A realizirao k puta u n ponavljanja pokusa. Pripadajuće vjerojatnosti jednake su:

$$P(X = k) = \binom{n}{k} p^k q^{n-k},\ k = 0, 1, \dots, n, \tag{4.6}$$

pri čemu je $q = 1 - p$. Slučajna varijabla X ravna se po **binomnoj razdiobi**. Uobičajeno pišemo $X \sim B(n, p)$.

Vrijednosti $n \in \mathbb{N}$ i $0 < p < 1$ su parametri binomne razdiobe.

Ako su X_i, $i = 1, \dots, n$ Bernoullijeve slučajne varijable koje prate realizaciju događaja A u i-tom ponavljanju pokusa, tada je njihov zbroj

$$X = X_1 + X_2 + \dots + X_n \tag{4.7}$$

binomna slučajna varijabla s razdiobom $B(n, p)$.

Binomna razdioba je jedna od najvažnijih diskretnih razdioba te ju često susrećemo u primjenama kao što je primjerice kontrola pomoću uzoraka.

Primjer 4.3. *Vjerojatnost da tijekom jednog sata zakaže element nekog sklopa je 0.01. Sklop se sastoji od 100 elemenata. Odredimo:*

a) vjerojatnost da će tijekom jednog sata otkazati točno jedan element,

b) vjerojatnost da će tijekom jednog sata otkazati najviše jedan element,

c) vjerojatnost da će tijekom jednog sata otkazati barem jedan element.

Ako je X broj elemenata koji otkažu, vrijedi $X \sim B(100, 0.01)$. Za proizvoljan $k = 0, \ldots, 100$, događaj $(X = k)$ odgovara ishodu u kojem je točno k elemenata sklopa otkazalo. Koristeći svojstva binomne razdiobe imamo:

a)

$$P(X = 1) = 100 \cdot 0.01 \cdot 0.99^{99} = 0.36973, \tag{4.8}$$

b)

$$\begin{aligned} P(X \leq 1) = P(X = 0) + P(X = 1) &= 0.99^{100} + 100 \cdot 0.01 \cdot 0.99^{99} \\ &= 0.36603 + 0.36973 = 0.73576, \end{aligned} \tag{4.9}$$

c)

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - 0.36603 = 0.63397. \tag{4.10}$$

Primjer 4.4. *Vjerojatnost ispravnog rada nekog uređaja je svega 0.54. Kako bi se osigurala protočnost proizvodnje tvornica uvijek ima na raspolaganju četiri takva uređaja. Pritom za ispravno funkcioniranje proizvodnje tvornica mora imati barem jedan ispravan uređaj. Odredimo:*

a) S kojom je vjerojatnošću trenutno zadovoljena protočnost proizvodnje?

b) Koliko bi uređaja tvornica trebala imati na raspolaganju ako protočnost proizvodnje mora biti zadovoljena s vjerojatnošću od barem 99.5%?

Pretpostavimo li da je X broj ispravnih uređaja, X je slučajna varijabla koja se ravna po binomnoj razdiobi.

a) Ako tvornica raspolaže s četiri uređaja, slijedi $X \sim B(4, 0.54)$. Budući da za protočnost proizvodnje barem jedan uređaj mora biti ispravan, tražimo vjerojatnost događaja $(X \geq 1)$:

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - 0.46^4 = 0.9552. \tag{4.11}$$

b) S obzirom da tražimo broj uređaja koji bi tvornica trebala imati da bi se s dovoljno velikom vjerojatnošću osigurala protočnost proizvodnje, sada vrijedi $X \sim B(n, 0.54)$. Tražimo najmanji n za koji vrijedi

$$P(X \geq 1) \geq 0.995. \tag{4.12}$$

Dobivamo:

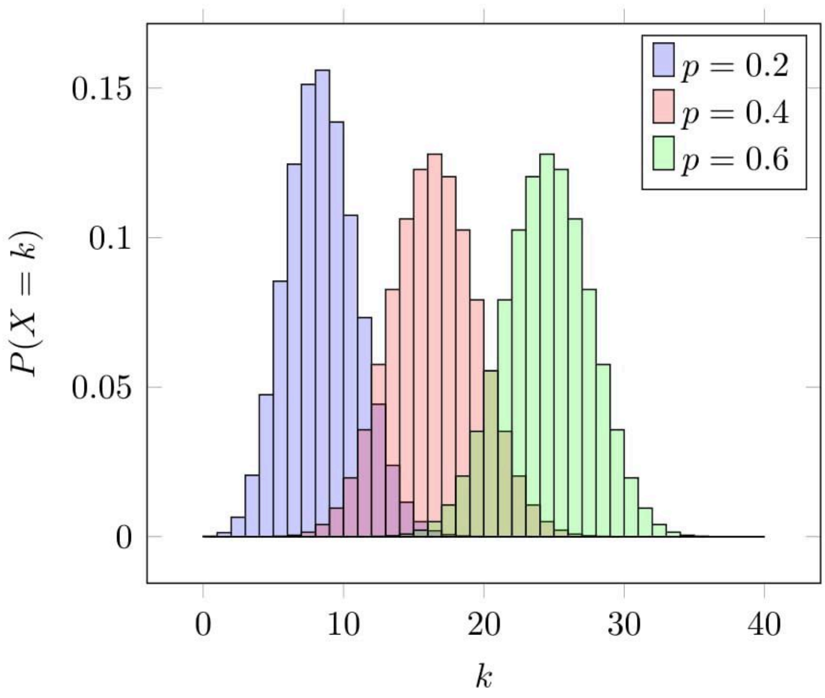
$$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.46^n \geq 0.995 \tag{4.13}$$

$$0.46^n \leq 0.005 \tag{4.14}$$

$$n \geq \frac{\ln 0.005}{\ln 0.46} = 6.82. \tag{4.15}$$

Zaključujemo da je za osiguranje tražene protočnosti potrebno imati barem 7 uređaja.

Funkcija gustoće binome razdiobe prikazana je na sljedećoh slici.



Slika 4.1. Funkcija gustoće binome razdiobe $B(40, p)$ za različite vrijednosti parametra p

4.4. Poissonova razdioba

Neka je $\lambda > 0$ **intenzitet** pojavljivanja nekog događaja u određenom vremenskom intervalu ili na određenoj prostornoj domeni. Slučajna varijabla X neka predstavlja **broj realizacija tog događaja u promatranoj domeni** (vremenskoj ili prostornoj). X poprima vrijednosti iz skupa $\{0, 1, 2, \ldots\}$, te su pripadajuće vjerojatnosti jednake:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},\ k = 0, 1, 2, \ldots. \tag{4.16}$$

Slučajna varijabla X ravna se po **Poissonovoj razdiobi**. Uobičajeno pišemo $X \sim \mathcal{P}(\lambda)$, gdje je $\lambda > 0$ parametar Poissonove razdiobe.

Primjeri situacija koje se mogu modelirati Poissonovom razdiobom su sljedeći: broj pristiglih telefonskih poziva, broj nesreća, broj ljudi koji čekaju u redu za određene usluge, broj pukotina u materijalu i sl.

Poissonova razdioba zapravo je granični slučaj binomne kada teži u beskonačnost, ali mora bit zadovoljen uvjet, a to je da je umnožak np konstantan. To je dakle slučaj kada je vjerojatnost događaja vrlo mala, a veličina skupa "beskonačna".

Zakon vjerojatnosti pri binomnoj razdiobi je :

$$P(x) = \binom{n}{x} p^x q^{n-x} = \frac{n(n-1)(n-2)\ldots(n-x+1)}{x!} p^x q^{n-x}. \tag{4.17}$$

Uvedemo li $np = m$ tj. $p = \frac{n}{m}$ slijedi:

$$P(x) = \frac{1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{x}{n}+\frac{1}{n}\right)}{x!} m^x \left(1-\frac{m}{n}\right)^{m-x}, \tag{4.18}$$

a za $n \to \infty$

$$p(x) = \lim_{n\to\infty} P(x) = m^x \cdot \frac{1}{x!} \lim_{n\to\infty} \left(1 - \frac{m}{n}\right)^n = \frac{m^x}{x!} e^{-m} \tag{4.19}$$

stoga slijedi:

$$P(x) = \frac{m^x}{x!} e^{-m}, \tag{4.20}$$

što smo već definirali uzevši da je $m = \lambda$.

Razdiobu je otkrio Poisson 1837. godine, ali i Bortkewitsch ponovno 1898. godine daje zakonitosti ponašanja rijetkih događaja pod nazivom "zakon malih brojeva".

Primjer 4.5. *Kod organizacije rada šalterske službe bilježi se broj klijenata koji dolaze u banku. Utvrđeno je da u prosjeku tijekom deset minuta dolazi troje klijenata. Odredimo*

a) Kolika je vjerojatnost da će u prvih deset minuta nakon otvaranja u banku doći točno dva klijenta?

b) Kolika je vjerojatnost da će u prvih deset minuta nakon otvaranja u banku doći najviše dva klijenta?

c) Pod pretpostavkom da radi samo jedan šalter te da službenica na šalteru za svakog klijenta utroši pet minuta, kolika je vjerojatnost da će klijent trebati pričekati na pružanje usluge.

Označimo li s X broj klijenata koji dođu u banku u periodu od deset minuta, možemo pretpostaviti da se taj broj ravna po Poissonovoj razdiobi $\mathcal{P}(\lambda)$. Budući da u promatranom periodu u prosjeku dolazi troje klijenata, vrijedi $E(X) = \lambda = 3$, odnosno $X \sim \mathcal{P}(3)$.

a)

$$P(X = 2) = \frac{3^2}{2!} e^{-3} = 0.22404. \tag{4.21}$$

b)

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= e^{-3} \left(1 + 3 + \frac{3^2}{2!}\right) = 0.42319. \end{aligned} \tag{4.22}$$

c) Broj klijenata koji uđu u banku tijekom pet minuta također je slučajna varijabla koja se ravna po Poissonovoj razdiobi. Označimo li je s Y vrijedi $Y \sim \mathcal{P}(1.5)$. Vjerojatnost da će klijent trebati pričekati na pružanje usluge odgovara vjerojatnosti da tijekom pet minuta u banku uđu barem dvije osobe, pa je tražena vjerojatnost jednaka

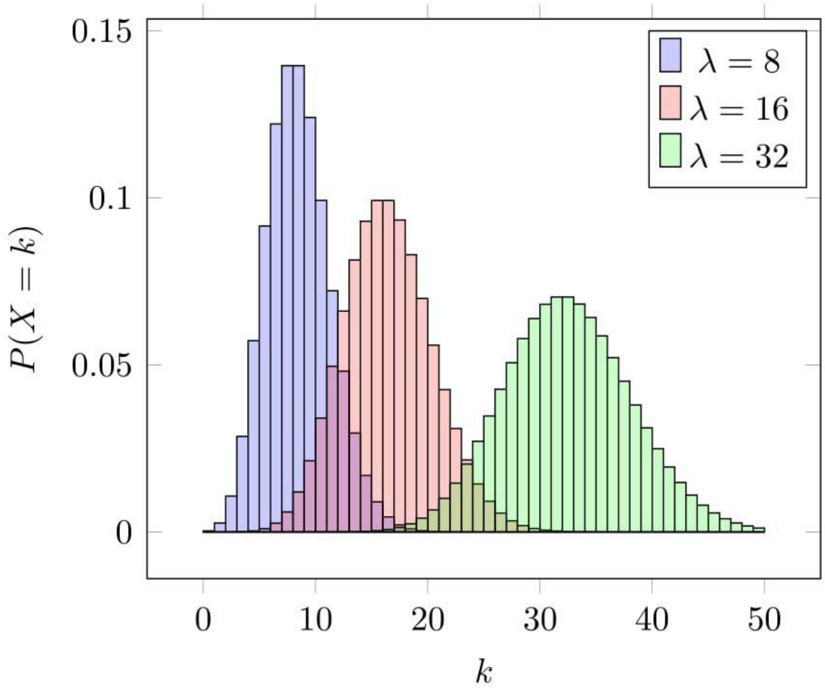
$$P(Y \geq 2) = 1 - P(Y = 0) - P(Y = 1) = 1 - e^{-1.5}(1 + 1.5) = 0.44217. \tag{4.23}$$

Primjer 4.6. *Onečišćenje česticama prašine predstavlja problem kod proizvodnje memorijskih ploča. Broj čestica koje se pojave na takvoj ploči je slučajan i ravna se po Poissonovoj razdiobi, s prosječnim brojem 0.02 čestica po kvadratnom centimetru ploče. Površina ploče je 100* cm^2. *Odredimo kolika je vjerojatnost da će se na ploči pojaviti više od 3 čestice?*

Za slučajnu varijablu X *koja predstavlja broj čestica na promatranoj ploči vrijedi* $X \sim \mathcal{P}(100 \cdot 0.02) = \mathcal{P}(2)$. *Tražena vjerojatnost jednaka je*

$$P(X > 3) = 1 - P(X \leq 3) = 1 - e^{-2}\left(1 + 2 + \frac{2^2}{2!} + \frac{2^3}{3!}\right) = 0.14288. \tag{4.24}$$

Funkcija gustoće Poissonove razdiobe prikazana je na sljedećoj slici.



Slika 4.2. Funkcija gustoće Poissonove razdiobe $\mathcal{P}(\lambda)$ *za različite vrijednosti parametra* λ

4.5. Geometrijska razdioba

Neka je A odabrani događaj i neka je vjerojatnost njegove realizacije $p = P(A)$. Pokus ponavljamo do prve realizacije događaja A, pretpostavljajući pritom da su ishodi pojedinih ponavljanja pokusa međusobno nezavisni. Slučajna varijabla X neka predstavlja **broj ponavljanja pokusa do prve realizacije događaja** A. Tako definirana slučajna varijabla poprima vrijednosti iz skupa $\{1, 2, \ldots\}$. Pritom $(X = k)$ predstavlja događaj da se je promatrani događaj A **prvi put** realizirao u k-tom ponavljanju pokusa. Pripadajuće vjerojatnosti jednake su:

$$P(X = k) = pq^{k-1}, \ k = 1, 2, \ldots, \tag{4.25}$$

pri čemu je $q = 1 - p$. Slučajna varijabla X ravna se po **geometrijskoj razdiobi**. Uobičajeno pišemo $X \sim G(p)$. Vrijednost $0 < p < 1$ je parametar geometrijske razdiobe.

Primjer 4.7. *Tipkovnica na prijenosnom računalu nekog proizvođača sastoji se od 86 tipki. Pretpostavimo da zatvorenih očiju trebamo stisnuti tipku za slovo A. Budući da je vjerojatnost pogotka bilo koje tipke jednaka* $\frac{1}{86}$ *(pretpostavimo da pri svakom pokušaju pogodimo neku tipku, tj. da nikada ne promašimo tipkovnicu), zaključujemo da slučajna varijabla koja opisuje broj stisnutih tipki do uspješnog stiskanja tipke za slovo A ima geometrijsku razdiobu s parametrom* $p = \frac{1}{86}$. *Tako definirana slučajna varijabla prima vrijednosti iz skupa* $R(X) = \{1, 2, \ldots\}$ *s pripadnim vjerojatnostima:*

$$P_k = P\{X = k\} = \frac{1}{86}\left(1 - \frac{1}{86}\right)^{k-1}. \tag{4.26}$$

a) Vjerojatnost da je tipka za slovo A stisnuta u petnaestom pokušaju je:

$$P\{X = 15\} = \frac{1}{86}\left(1 - \frac{1}{86}\right)^{14} = 0.00987. \tag{4.27}$$

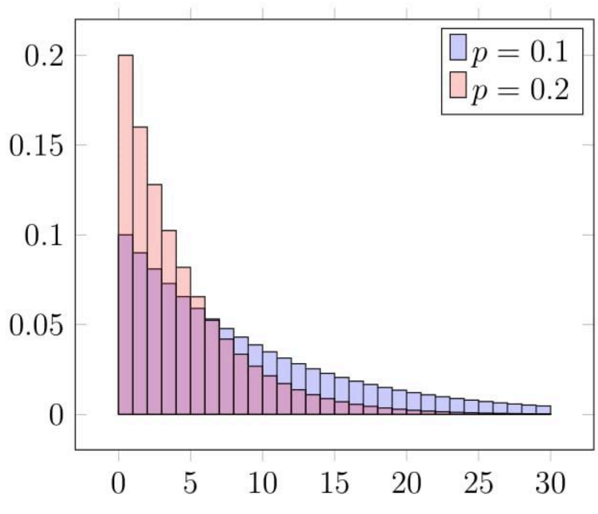
b) Vjerojatnost da je tipka za slovo A stisnuta u manje od 5 pokušaja je :

$$P[X < 5] = P\{X \leq 4\} = \sum_{k=1}^{4} \frac{1}{86}\left(1 - \frac{1}{86}\right)^{k-1} = 0.0157. \tag{4.28}$$

c) Vjerojatnost da nizi nakon 20 pokušaja još nismo uspijeli stisnuti tipku za slovo A:

$$P\{X > 20\} = 1 - P\{X \leq 10\} = 1 - \sum_{k=1}^{10} \frac{1}{86}\left(1 - \frac{1}{86}\right)^{k-1} = 0.7914. \tag{4.29}$$

Funkcija gustoće geometrijske razdiobe prikazana je na sljedećoj slici.



Slika 4.3. Funkcija gustoće geometrijske razdiobe $\mathcal{G}(p)$ *za različite vrijednosti parametra* p

4.6. Hipergeometrijska razdioba

Za "neiscrpive" ili nonekshaustivne skupove vrijedi zakon binomne razdiobe. Međutim, u praksi susrećemo često "iscrpive" ili ekshaustivne skupove kod kojih se kad uzmemo jedan element iz promatranog skupa mijenja se vjerojatnost preostalih elemenata u skupu.

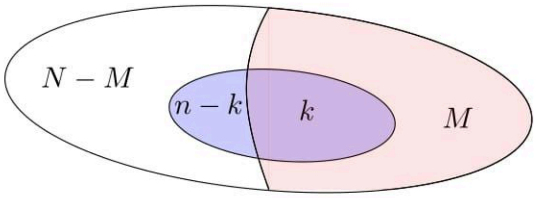
Zakon hipergeometrijske razdiobe vrijedi za takve ekshaustivne ili "iscrpive" skupove.

Pretpostavimo da imamo skup od N elemenata od kojih M elemenata ima neko promatrano obilježje, dok preostalih $N-M$ elemenata nema promatrano obilježje. Iz promatranog skupa biramo uzorak od n elemenata.

Slučajna varijabla X neka predstavlja **broj izabranih elemenata s promatranim obilježjem**. Vrijednosti koje slučajna varijabla X poprima su iz skupa $\{0, 1, \ldots, \min\{n, M\}\}$, te su pripadajuće vjerojatnosti jednake:

$$P(X=k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}. \tag{4.30}$$

Slučajna varijabla X ravna se po **hipergeometrijskoj razdiobi**. Uobičajeno pišemo $X \sim H(n, M, N)$. Ovdje su M, N i n parametri promatrane razdiobe.



Slika 4.4. Skup elemenata kod hipergeometrijske razdiobe

Primjer 4.8. *U skladištu se nalazi* 20 *ležajeva* ($N=20$) *i to 8 loših* ($M=8$)*, a 12 ispravnih* ($N-M=12$)*. Ležajevi se uzimaju jedan za drugim, dok ih se ne uzme 5 komada* ($n=5$)*. Odredimo kolika je vjerojatnost da od uzetih 5 komada bude* $0, 1, 2, ..., 5$ *loših?*

$$P(0) = \frac{12! \cdot 15!}{20!\,(20-8-5)!} = \frac{33}{646} = 0.051, \tag{4.31}$$

$$P(1) = 0.051 \cdot \frac{8 \cdot 5}{8} = 0.255, \tag{4.32}$$

$$P(2) = 0.255 \cdot \frac{7 \cdot 4}{2 \cdot 9} = 0.255 \cdot 0.555 = 0.3977, \tag{4.33}$$

$$P(3) = 0.3977 \cdot \frac{6 \cdot 3}{3 \cdot 10} = 0.3977 \cdot 0.6 = 0.23862, \tag{4.34}$$

$$P(4) = 0.23862 \cdot \frac{5 \cdot 2}{4 \cdot 11} = 0.23862 \cdot 0.254 = 0.0542, \tag{4.35}$$

$$P(5) = 0.0542 \cdot \frac{4 \cdot 1}{5 \cdot 12} = 0.0542 \cdot 0.06 = 0.0036. \tag{4.36}$$

5. Algoritimi za simuliranje diskretnih slučajnih varijabli

U ovom se poglavlju bavimo algoritmima za simuliranje diskretnih slučajnih varijabli. Naime, generatori slučajnuh brojeva u računalu pretpostavljaju da su brojevi uniformno distribuirani, što je vrlo rijetko slučaj. Upravo stoga potrebno je odrediti algoritme kojima možemo simulirati i brojeve koji imaju neku drugu razdiobu. Ovo poglavlje obrađeno je prema izvorima [10], [11] i [12].

5.1. Inverzna metoda

U ovom poglavlju bavimo se jednom od najjednostavnijih metoda za simuliranje slučajnih varijabli koja se temelji na inverzu funkcije razdiobe. Stoga najprije moramo i definirati funkciju razdiobe diskrente slučajne varijable.

Definicija 5.1. *Neka je slučajna varijabla X zadana zakonom razdiobe*

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}. \tag{5.1}$$

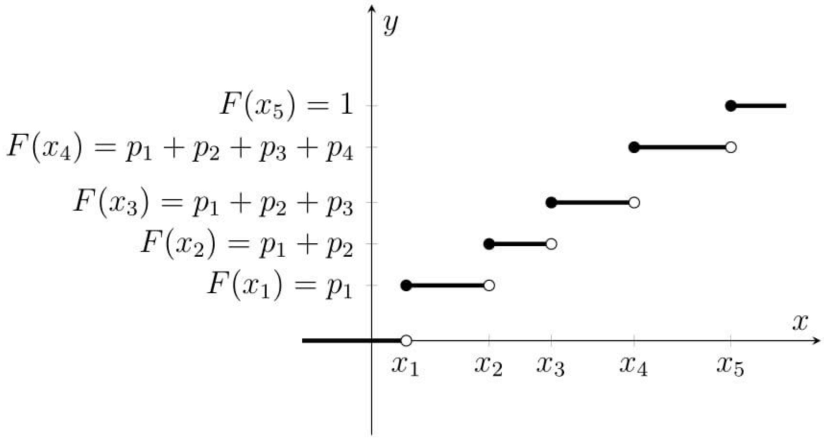
Funkcija razdiobe slučajne varijable X je tada funkcija $F : \mathbb{R} \to [0,1]$ zadana sa

$$F(x) = P(X \leq x) = \begin{cases} 0, \ x < x_1 \\ p_1, \ x_1 \leq x < x_2 \\ p_1 + p_2, \ x_2 \leq x < x_3 \\ \vdots \\ p_1 + p_2 + \cdots p_{n-1}, \ x_{n-1} \leq x < x_n \\ 1, \ x \geq x_n \end{cases}. \tag{5.2}$$

Jedan primjer funkcije F za $n = 5$ vidimo na sljedećoj slici.

Sada je jasno da se može definirati funkcija $F^* : [0,1] \to \{x_1, x_2, \dots, x_n\}$ na način

$$F^*(u) = \begin{cases} x_1, \ u \leq F(x_1) \\ x_2, \ F(x_1) < u \leq F(x_2) \\ x_3, \ F(x_2) < u \leq F(x_3) \\ \vdots \\ x_{n-1}, \ F(x_{n-2}) < x \leq F(x_{n-1}) \\ x_n, \ F(x_{n-1}) < x \leq 1 \end{cases}. \tag{5.3}$$



Slika 5.1. Grafički prikaz funkcije razdiobe

Dakle, ako smo u mogućnosti dobiti realizaciju uniformno distribuirane slučajne varijable na intervalu $[0, 1]$ (koju ćemo u nastavku zvati slučajnim brojem i označavati sa u), funkcijom F^* tu ćemo realizaciju preslikati u realizaciju promatrane slučajne varijable X. U praksi realizacija broja u ne predstavlja neki veliki problem budući je generator slučajnih brojeva dio gotovo svakog programskog jezika.

Pokažimo sada kako se opisani algoritam koristi unutar programskog jezika Phyton

U programskom jeziku Python za dobivanje slučajnog broja najprije moramo uvesti paket za znanstveno računanje, što radimo naredbom

```
import numpy as np
```

pri čemu smo samom paketu dodijelili alias `np`.

Slučajni broj sada dobivamo pomoću naredbe

```
np.random.rand()
```

Sada možemo formirati algoritam za simuliranje slučajne varijable X.

Pretpostavimo da je zakon razdiobe slučajne varijable X zadan dvodimenzionalnim poljem X. Najprije ćemo formirati polje S u koje ćemo spremati kumulativne vjerojatnosti slučajne varijable X, tj. u polju S nalazit će se vrijednosti $F(x_i)$. To u sustavu Python postižemo naredbom

```
np.cumsum(x[1,:])
```

Prema navedenom funkcija `GenerirajX(A)` koja bi davala jednu realizaciju slučajne varijable X u sustavu Python može izgledati ovako:

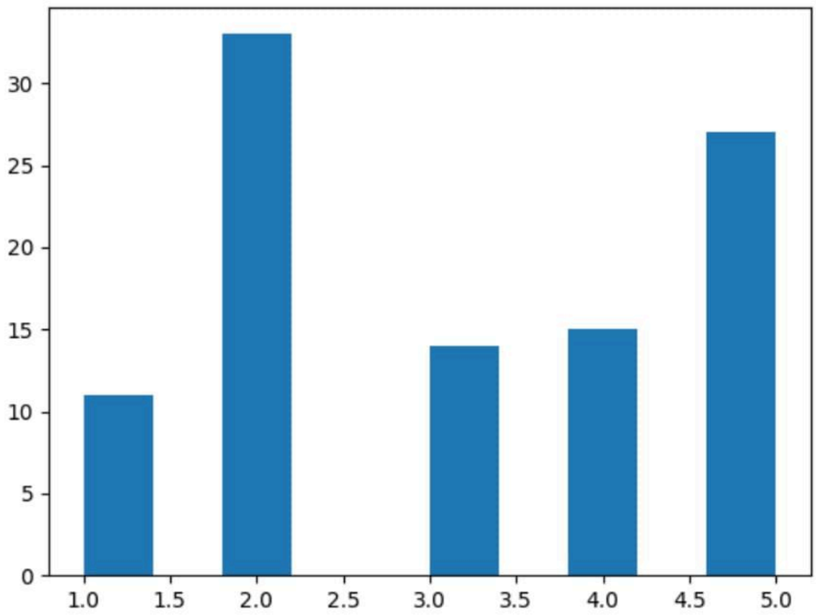
```
def GenerirajX(X):
        S=np.cumsum(X[1])
        i=0
        ulaz_u_petlju=1
        u=np.random.rand()
```

```
while ulaz_u_petlju:
        if u<=S[i]:
                broj=X[0][i]
                ulaz_u_petlju=0
        else:
                i=i+1
return broj
```

Kako bi provjerili točnost algoritma možemo generirati veći broj podataka i prikazati ih histogramom. To možemo učiniti na sljedeći način:

```
X=[[1,2,3,4,5],[0.1,0.3,0.2,0.15,0.25]]
podaci=np.zeros(100)
for i in range(100):
        podaci[i]=GenerirajX(X)
plt.hist(podaci)
plt.show()
```

U navedenom kodu generirali smo 100 brojeva i histogram iz tog slučaja prikazan je na sljedećoj slici.



Slika 5.2. Simulacija 100 slučajnih brojeva

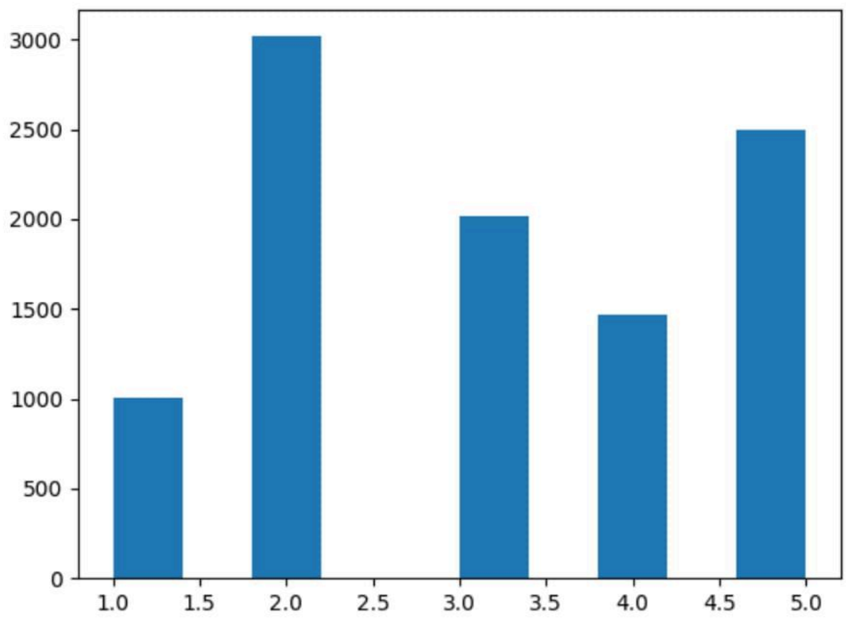
Možemo vidjeti da je razdioba vrlo slična zadanoj razdobi

$$X \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0.1 & 0.3 & 0.2 & 0.15 & 0.25 \end{pmatrix}. \tag{5.4}$$

Ako želimo precizniji rezultat, potrebno je generirati više slučajnih brojeva. Kod za 10000 slučajnih brojeva izgledao bi ovako:

```
X=[[1,2,3,4,5],[0.1,0.3,0.2,0.15,0.25]]
podaci=np.zeros(10000)
for i in range(10000):
        podaci[i]=GenerirajX(X)
plt.hist(podaci)
plt.show()
```

dok je pripadni histogram prikazan na sljedećoj slici.



Slika 5.3. Simulacija 10000 slučajnih brojeva

Napomenimo samo da je za korištenje naredbe `hist` nužno uvesti paket za crtanje što radimo naredbom

```
import matplotlib.pyplot as plt
```

pri čemu smo samom paketu dodijelili alias `plt`.

Ako je skup vrijednosti slučajne varijable relativno malen. Njenu realizaciju možemo napraviti i jednostavnije, bez korištenja petlje `while`. Pogledajmo kako bi u tom slučaju izgledao kod. Generiramo realizaciju slučajne varijable

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 \\ p_1 & p_2 & p_3 \end{pmatrix}. \tag{5.5}$$

Pripadna funkcija `GenerirajX1(x1,x2,x3,p1,p2,p3)` izgledati će ovako

```python
def GenerirajX1(x1,x2,x3,p1,p2,p3):
        u=np.random.rand()
        if u<=p1:
                broj=x1
        elif u<=p1+p2:
                broj=x2
        else:
                broj=x3
        return broj
```

Pokažimo sada na nekoliko primjera implementaciju opisanih algoritama.

Primjer 5.1. *Na nekoj nagradnoj igri može se dobiti 100 kn s vjerojatnošću od* 20%*, 200 kn s vjerojatnošću od* 10%*, a ne dobiti ništa s vjerojatnošću od* 70%*. Igru igra 500 igrača. Simulirajmo 5 realizacija igre tako da odredimo zaradu organizatora ako svaki igrač u igru ulaže 50 kn. Zaradu po realizacijama prikažimo grafički.*

Uočimo najprije da je temelj ovog zadatka simulacija slučajne varijable

$$X \sim \begin{pmatrix} 100 & 200 & 0 \\ 0.2 & 0.1 & 0.7 \end{pmatrix}. \tag{5.6}$$

Ovu varijablu možemo simulirati tako što ćemo koristiti već definiranu funkciju

`GenerirajX(X)`

Samu varijablu zadati ćemo matricom

`X=[[100,200,0],[0.2,0.1,0.7]]`

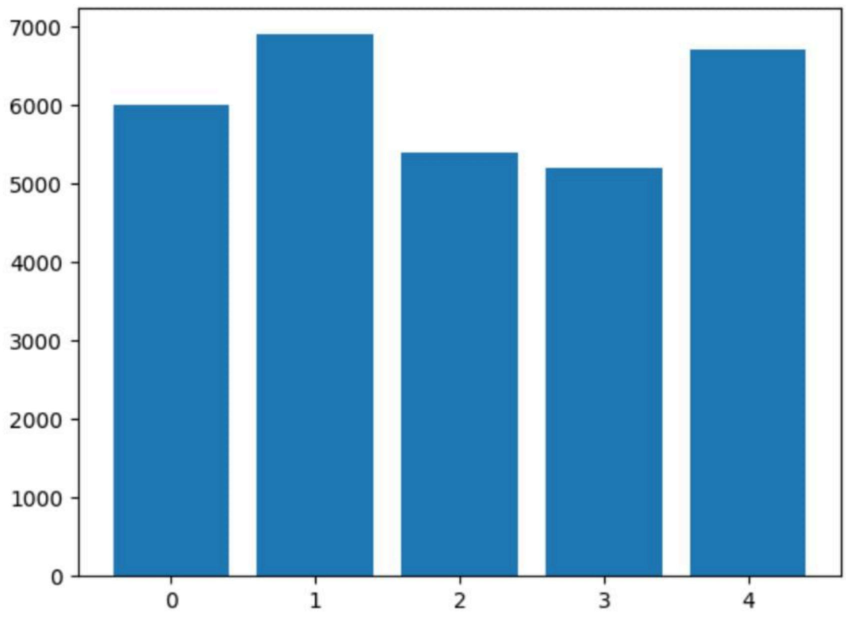
Ovdje treba uočiti da nas zanima nova slučajna varijbla koja predstavlja zaradu i da ona predstavlja sumu zarada u svakoj realizaciji. Također je bitno napomenuti da se pod zaradom smatra zarada organizatora, a ne igrača. Drugim riječima od ukupne zarade u svakoj iteraciji oduzimamo dobitak igrača.

Kao rezultat simulacije najpogodniji je grafički prikaz u vidu stupčastog dijagrama.

Grafički prikaz podataka radimo korištenjem naredbe `bar`*. Kod rješenja dan je u nastavku:*

```python
X=[[100,200,0],[0.2,0.1,0.7]]
zarada=np.zeros(5);
for i in range(5):
        for j in range(500):
                zarada[i]=zarada[i]+50-GenerirajX(X)
plt.bar(range(5),zarada)
plt.show()
```

Rezultat simulacije prikazan je na sljedećoj slici.



Slika 5.4. Rezultat simulacije

Primjer 5.2. *Kupac u trgovini kupuje pisač. Na raspolaganju ima 5 modela, označimo ih redom* 1, 2, 3, 4 *i* 5. *Vjerotnosti prodaje svakog od njih su respektivno* 0.2, 0.1, 0.15, 0.25 *i* 0.3.

Model 1 i model 2 prodaju se po cijeni od 480 kn, model 3 po cijeni od 620 kn, model 4 po cijeni od 950 kn, a model 5 po cijeni od 450 kn.

Broj kupaca koji kupe printer u tijeku jednog dana kreće se između 5 i 12, tako da je svaki broj kupaca u tom rasponu jednakovjerojatan.

Simulirajmo prodaju printera i prikažimo grafički prosječnu dnevnu zaradu za svaki mjesec (uz pretpostavku da mjesec u prosjeku ima 30 dana) tijekom perioda od 12 mjeseci.

U ovom slučaju promatramo dvije slučajne varijable. Varijablu P *koja predstavlja pisač za koji se mušterija odlučila, odnosno njegovu cjenu te varijablu* B *koja predstavlja broj mušterija koji su ušli u trgovinu tijekom dana.*

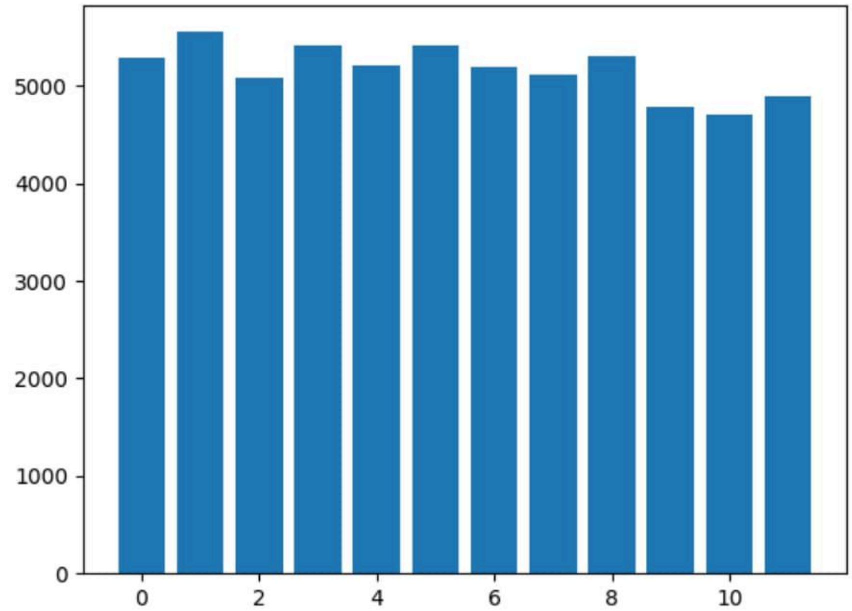
Najprije ćemo generirati broj mušterija i za dobiveni broj generirati odgovarajući broj realizacija cijena.

Za izračunavanje aritmetičke sredine koristimo naredbu `mean`.

Nakon toga izračunavamo ukupnu zaradu tijekom dana. Kod rješenja dan je u nastavku:

```
P=[[480,480,620,950,450],[0.2,0.1,0.15,0.25,0.3]]
B=[[5,6,7,8,9,10,11,12],
[0.125,0.125,0.125,0.125,0.125,0.125,0.125,0.125]]
```

```
zarada=np.zeros((12,30));
prosjeci=np.zeros(12);
for i in range(12):
        for j in range(30):
                n=GenerirajX(B)
                for k in range(n):
                        zarada[i][j]=zarada[i][j]+GenerirajX(P);
        prosjeci[i]=np.mean(zarada[i]);
plt.bar(range(12),prosjeci);
plt.show()
```



Slika 5.5. Rezultat simulacije

Objasnimo ukratko ovaj programski kod. Najprije smo poljima P i B zadali parametre spomenutih slučajnih varijabli. Zatim definiramo polje zarada i polje prosjeci. Polje zarada je dvodimenzionalno polje u koje će se spremati dnevna zarada za svaki od 30 dana kroz 12 mjeseci u godini. Polje prosjeci ima 12 unosa u koje će se spremati mjesečni dnevni prosjeci zarade.

Petlja s iteratorom i odnosi se na mjesece, dok se petlja s iteratorom j odnosi na dane unutar pojedinog mjeseca. Zatim generiramo broj mušterija koji je dan slučajnoj varijablom B i taj broj spremamo u varijablu n. Tada unutar petlje s iteratorom k za svaku mušteriju generiramo iznos koji će potrošiti, odnosno pisač koji će kupiti. Svi se ti iznosi zbrajaju po danima. Kada se završi generiranje svih dana u mjesecu izračuna se prosjek dnevne zarade koji se sprema u odgovarajuće mjesto polja prosjeci. Nakon toga kreće se na simuliranje idućeg mjeseca. Na kraju se stupčastim *grafikonom prikazuje rezultat polja prosjeci što je u biti i bio zadatak ove simulacije. Taj je grafički prikaz dan na slici.*

Napomenimo, da bi program radio, nužno je prije njega imati uključene odgovarajuće biblioteke i definiraju funkciju GenerirajX.

Primjer 5.3. *Zadana je slučajna varijabla*

$$X \sim \begin{pmatrix} -20 & 0 & 10 & 50 & 100 \\ 0.3 & 0.2 & 0.1 & 0.2 & 0.2 \end{pmatrix}.$$

Generirat ćemo uzorke slučajnih brojeva (koji imaju zadanu razdiobu) veličine 10, 100, 1000 i 10000. Za svaki uzorak prikazat ćemo grafički apsolutno odstupanje aritmetičke sredine od matematičkog očekivanja slučajne varijable X.

Objasnimo najprije pojam matematičkog očekivanja. ***Matematičko očekivanje*** *diskretne slučajne varijable računa se po formuli*

$$E(X) = \sum_{k=1}^{\infty} x_k p_k, \tag{5.7}$$

ako navedena suma postoji. Često se koristi oznaka: $\mu = E(X)$. Matematičko očekivanje mjera je centralne tendencije slučajne varijable.

Ako je slučajna varijabla X u programskom jeziku reprezentirana dvodimenzionalnim poljem, tj. matricom X kod koje su u prvom retku vrijednosti, a u drugom vjerojatnosti tada se matematičko očekivanje može izračunati naredbom

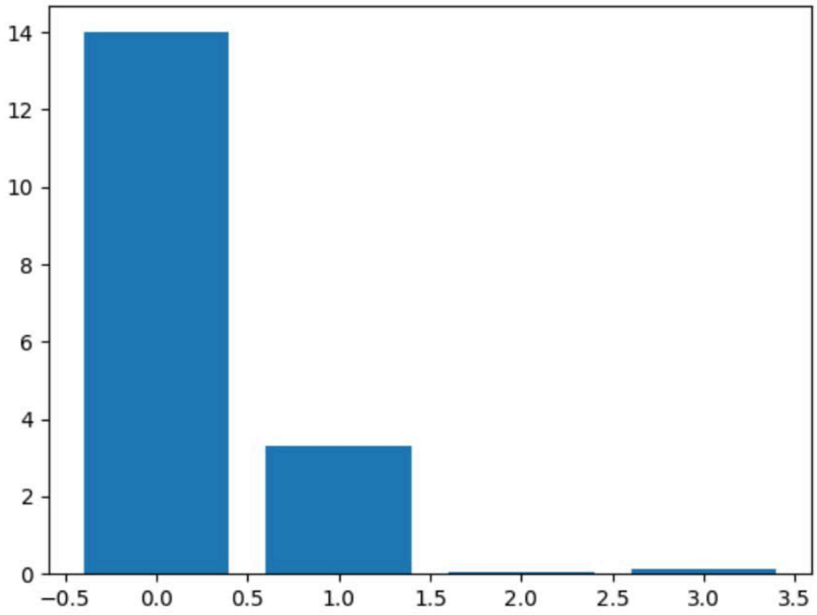
```
np.dot(x,y)
```

budući se matematičko očekivanje može shvatiti kao skalarni produkt prvog i drugog retka matrice.

Kod rješenja dan je u nastavku:

```
X=[[-20,0,10,50,100],[0.3,0.2,0.1,0.2,0.2]]
ocekivanje=np.dot(X[0],X[1])
n=[10,100,1000,10000]
greske=np.zeros(4)
for k in range(4):
        podaci=np.zeros(n[k])
        for j in range(n[k]):
                podaci[j]=GenerirajX(X)
        greske[k]=np.abs(ocekivanje-np.mean(podaci))
plt.bar(range(4),greske);
plt.show()
```

Rezultat simulacije prikazan je na sljedećoj slici.



Slika 5.6. Rezultat simulacije

Sa slike možemo vidjeti da se greška smanjuje povećavanjem uzorka, drugim riječima što je uzorak veći aritmetička sredina biti će bliža svojoj teorijskog vrijednosti, odnosno matematičkom očekivanju.

5.2. Simulacija binomne slučajne varijable

Kao što smo već rekli, za slučajnu varijablu sa zakonom razdiobe

$$X \sim \begin{pmatrix} 1 & 0 \\ p & 1-p \end{pmatrix} \tag{5.8}$$

kažemo da ima Bernoullijevu razdiobu i označavamo je s $X \sim Ber(p)$. Na osnovu opisanog algoritma za simuliranje općenite diskretne slučajne varijable lako dolazimo da funkcije kojom ćemo dobiti jednu realizaciju slučajne varijable sa Bernoullijevom razdiobom sa parametrom p:

```
def GenerirajBernoulliX(p):
    u=np.random.rand()
    if u<=p:
        broj=1
    else:
        broj=0
    return broj
```

Također smo već rekli da za slučajnu varijablu kojoj je skup vrijednosti iz skupa $X \in \{0, 1, \ldots, n\}$, $n \in \mathbb{N}$ i za koju vrijedi

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{5.9}$$

kažemo da ima binomnu $B(n, p)$ razdiobu. Algoritam za simulaciju binomno distribuirane slučajne varijable temelji se na sljedećem svojstvu binomne razdiobe.

Teorem 5.1. *Neka su X_i, $i = 1, ..., n$ slučajne varijable sa Bernoullijevom $Ber(p)$ razdiobom. Tada slučajna varijabla*

$$Z = \sum_{i=1}^{n} X_i$$

ima binomnu $B(n, p)$ razdiobu.

Prema tome, funkcija kojom ćemo dobiti jednu realizaciju binomno distribuirane slučajne varijable može izgledati ovako:

```
def GenerirajBinomnuX(n,p):
    broj=0
    for i in range(n):
        broj=broj+GenerirajBernoulliX(p)
    return broj
```

Primjer 5.4. *U nekom razredu ima 20 učenika. Vjerojatnost da učenik dobije odličnu ocjenu je 15%. Simulirajmo broj odličnih ocjena u 10 razreda. Rezultate simulacije prikažimo grafički.*

Uočimo najprije da će slučajna varijabla "broj odličnih" biti binomno distribuirana s parametrima $n = 20$ i $p = 0.15$. Slučajne brojeve spremati ćemo u polje veličine 10, gdje svaki unos predstavlja jedan razred. Pripadni kod izgledao bi ovako:

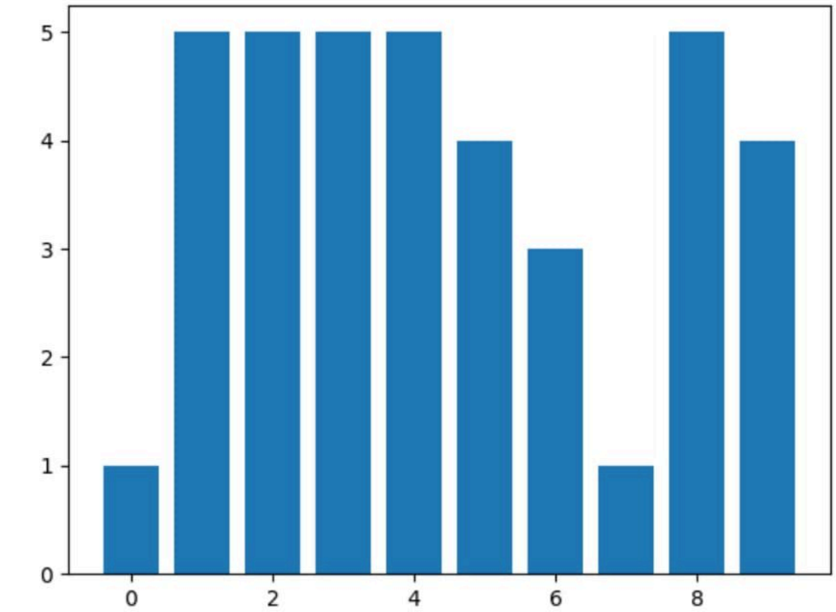
```
ocjene=np.zeros(10);
for i in range(10):
    ocjene[i]=GenerirajBinomnuX(20,0.15)
plt.bar(range(10),ocjene);
plt.show()
```

Napomenimo da funkcije koje smo prethodno objasnili također moraju biti definirane prije navođenja ovog koda. Rezultat simulacije možemo vidjeti na sljedećoj slici.

Spomenuti algoritam možemo testirati na sljedeći način. Matematičko očekivanje binomno distribuirane slučajne varijable s parametrima n i p dano je izrazom

$$E(X) = np, \tag{5.10}$$

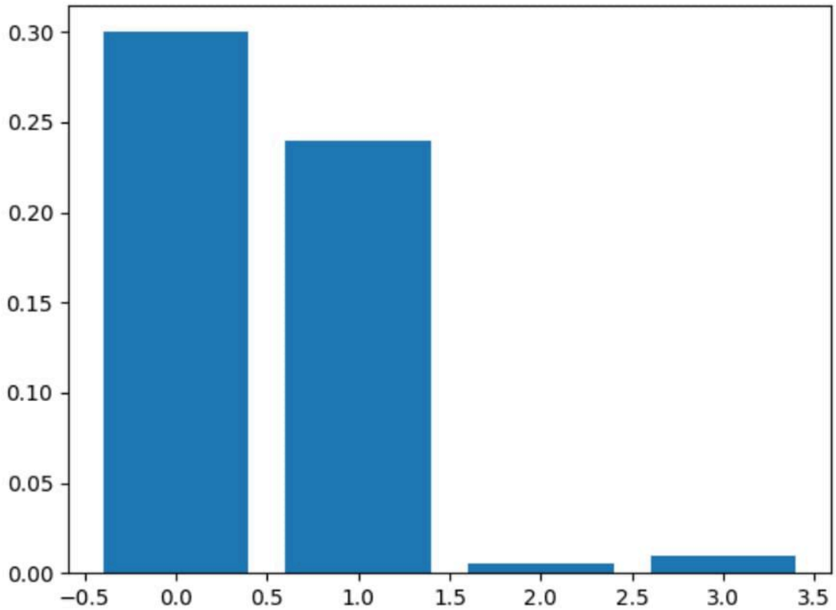
a u ovom slučaju je $E(X) = 20 \cdot 0.15 = 3$ što znači da bi se u većini simulacija dobiveni broj trebao kretati oko 3, što na ovoj slici zbog malog broja simulacija i ne možemo precizno vidjeti.



Slika 5.7. Rezultat simulacije

Kako bi bili sigurni u točnost simulacije, provesti ćemo simulaciju iz zadnjeg primjera u prethodnom poglavlju, odnosno pogledati ćemo koliko se aritmetička sredina generiranog uzorka razlikuje od testiranog matematičkog očekivanja.

Rezultat te simulacije prikazan je na sljedećoj slici.



Slika 5.8. Rezultat simulacije

Sa slike možemo vidjeti da se kod većeg uzorka aritmetička sredina približava matematičkom očekivanju. Pripadni kod dan je u nastavku.

```
n=[10,100,1000,10000]
greske=np.zeros(4)
for k in range(4):
        podaci=np.zeros(n[k])
        for j in range(n[k]):
                podaci[j]=GenerirajBinomnuX(20,0.15)
        greske[k]=np.abs(3-np.mean(podaci))
plt.bar(range(4),greske);
plt.show()
```

Kod algoritma za generiranje binomne razdiobe na značajno velikom uzorku može se uočiti i određena sporost algoritma, što se i navodi u literaturi. Problem s binomom razdiobom je bio u tome što nije postojala eksplicitna funkcija gustoće iz koje bi se mogao naći inverz, pa tako nije bilo moguće koristiti općenitu inverzu metodu. Upravo nam to govori o velikom problemu sa ovim tipom algoritama, odnosno većina razdioba zahtjevati će zasebne algoritme, budući opći algoritmi neće biti primjenjivi ili će pak postojati problem brzine izvršavanja.

6. Zaključak

U ovom radu uveden je pojam vjerojatnosti s posebnim naglaskom na pojam slučajne varijable. Pokazali smo da je slučajna varijabla jedan od važnijih koncepata kako teorije vjerojatnosti tako i inženjerstva.

Kada govorimo o slučajnom broju obično mislimo na broj koji dolazi iz uniformne razdiobe, no ovim radom pokazali smo da slučajni broj može imati i neku drugu razdiobu. Primjerice, ako želimo simulirati broj intervencija nekog servisa unutar jednog sata koristit ćemo Poissovovu razdiobu, a ako pak želimo simulirati broj ispravnih brojeva u nekom skladištu prirodno je koristiti bionmnu razdiobu. Drugim rječima, simuliranje različitih poslovnih i inženjerskih procesa nužno je znati simulirati slučajne brojeve koji imaju određenu razdiobu.

Ovim radom demonstrirali smo neke algoritame za generiranje slučajnih brojeva s definiranim razdiobama. Pokazalo se da su obrađeni algoritmi teorijski relativno jednostavni. Međutim, prilikom opsežnijih simulacija bio je očit problem vremena izvršavanja algoritma što otvara potrebu za razvojem specifičnih algoritama vezanih uz svaku pojedinu razdiobu.

Za simulacije unutar ovog rada korišten je programski jezik Phyton koji se pokazao optimalnim prije svega jer se radi o besplatnom alatu čija je sintaksa prilično jednostavna. Drugim riječima, za simuliranje osnovnih inženjerskih problema povezanih sa slučajnim varijablama Phyton može biti dobar izbor.

Literatura

[1] Tomašić, L.: "Matematika IV", Sveučilište u Rijeci, Rijeka, 1993.

[2] Elezović, N.: "Vjerojatnost i statistika", Element, Zagreb, 2007.

[3] Sarapa, N.: "Teorija vjerojatnosti", Školska knjiga, Zagreb, 1992.

[4] Jaynes, E.T.: "Probability Theory: The Logic of Science", Washington University, St. Louis, 1995.

[5] Shiryaev, A.N.: "Veroiatnost", Nauka, Moskva, 1980.

[6] Simčić, L., Dražić, I., Bilješke s predavanja iz kolegija Inženjerska statistika, Tehnički fakultet u Rijeci

[7] "OSNOVE TEORIJE VJEROJATNOSTI - PMF", s Interneta, https://www.pmf.unizg.hr/download/repository/PREDAVANJE6.pdf, 15.04.2022

[8] "3.2. Diskretna slučajna varijabla - Stedy", s Interneta, https://stedy.hr/slucajna-varijabla/diskretna-slucajna-varijabla, 07.05.2022

[9] Krajnović, M.: "Transformacije diskretnih i neprekidnih slučajnih varijabli", završni rad, Sveučilište u Osijeku, Osijek, 2017.

[10] Ross, S.M.: Introduction to probability models, 10. izdanje, Elsevier, 2010.

[11] "Python Tutorial - Python Full Course for Beginners", s Interneta, https://www.youtube.com/watch?v=uQrJ0TkZlcabchannel=ProgrammingwithMosh, 19.06.2022

[12] "Simulating probability events in Python | by Elliott Saslow", s Interneta, https://medium.com/future-vision/simulating-probability-events-in-python-5dd29e34e381, 05.07.2022

Sažetak i ključne riječi

U ovome radu ukratko je opisana teorija vjerojatnosti. Definirali smo slučajnu varijablu i njeno zadavanje putem zakona razdiobe. Temeljitije smo se bavili diskretnom slučajnom varijablom i sistematizacijom njezinih razdioba (unifromna, Benoullijeva, binomna, Poissonova, geometrijska, hipergeometrijska). Uz svaku vrstu razdiobe navedeni su primjeri od kojih su neki povezani inženjerskom strukom. U zadnjem dijelu rada bavili smo se algoritmima za simuliranje slučajnih varijabli zadanih zakonima razdiobe. Objasnili smo inverznu metodu kao i neke specifične metode povezane sa konkretnim razdiobama. Opisani algoritmi realizirani su u programskom jeziku Phyton pri čemu je napravljen veći broj simulacija.

Ključne riječi: vjerojatnost, slučajna varijabla, simulacija, algoritam simulacije, Phyton

Summary and key words

In this paper, we briefly describe the theory of probability. We have defined a random variable and its assignment by the distributive law. We have dealt in more detail with the discrete random variable and the systematization of its distributions (uniform, Benoulli, binomial, Poisson, geometric, hypergeometric). Examples are given for each type of distribution, some of which are related to the engineering profession. In the last part of the paper we dealt with algorithms for simulation of random variables given by distribution laws. We explained the inverse method as well as some specific methods related to particular distributions. The described algorithms were implemented in the Python programming language, and a large number of simulations were performed.

Keywords: probability, random variable, simulation, simulation algorithm, Python