SVEUČILIŠTE U ZAGREBU

FAKULTET KEMIJSKOG INŽENJERSTVA I TEHNOLOGIJE

SVEUČILIŠNI PREDDIPLOMSKI STUDIJ

Alena Đogić

ZAVRŠNI RAD

Zagreb, rujan 2024.

SVEUČILIŠTE U ZAGREBU

FAKULTET KEMIJSKOG INŽENJERSTVA I TEHNOLOGIJE

POVJERENSTVO ZA ZAVRŠNE ISPITE

Kandidatkinja Alena Đogić

Predala je izrađen završni rad dana: 18. rujna 2024.

Povjerenstvo u sastavu:

doc. dr. sc. Miroslav Jerković, Sveučilište u Zagrebu Fakultet kemijskog inženjerstva i tehnologije

dr. sc. Monika Šabić Runjavec, Sveučilište u Zagrebu Fakultet kemijskog inženjerstva i tehnologije

izv. prof. dr. sc. Vanja Kosar, Sveučilište u Zagrebu Fakultet kemijskog inženjerstva i tehnologije

doc. dr. sc. Iva Movre Šapić, Sveučilište u Zagrebu Fakultet kemijskog inženjerstva i tehnologije (zamjena)

povoljno je ocijenilo završni rad i odobrilo obranu završnog rada pred povjerenstvom u istom sastavu.

Završni ispit održat će se dana: 23. rujna 2024.

SVEUČILIŠTE U ZAGREBU

FAKULTET KEMIJSKOG INŽENJERSTVA I TEHNOLOGIJE

SVEUČILIŠNI PREDDIPLOMSKI STUDIJ

Alena Đogić

DISKRETNE SLUČAJNE VARIJABLE I PRIMJENE

ZAVRŠNI RAD

Mentorica: izv.prof.dr.sc. Erna Begović Kovač, Sveučilište u Zagrebu Fakultet kemijskog inženjerstva i tehnologije

Članovi povjerenstva:

1. doc. dr. sc. Miroslav Jerković, Sveučilište u Zagrebu Fakultet kemijskog inženjerstva i tehnologije
2. dr. sc. Monika Šabić Runjevac, Sveučilište u Zagrebu Fakultet kemijskog inženjerstva i tehnologije
3. izv. prof. dr. sc. Vanja Kosar, Sveučilište u Zagrebu Fakultet kemijskog inženjerstva i tehnologije

Zagreb, rujan 2024.

SAŽETAK

Ovaj završni rad bavi se tematikom diskretnih slučajnih varijabli te njihovom primjenom u statistici i inženjerstvu. U prvom dijelu završnog rada definiramo diskretne slučajne varijable i njihove osnovne karakteristike kao što su funkcija gustoće vjerojatnosti, funkcija distribucije vjerojatnosti, očekivana vrijednost, varijanca i standardna devijacija. U drugom dijelu rada analiziramo nekoliko važnih distribucija, uključujući Bernoullijevu, binomnu, Poissonovu, negativnu binomnu, geometrijsku te hipergeometrijsku distribuciju. Svaka distribucija prikazana je različitim primjerima i grafičkim prikazima.

Rad omogućuje bolje shvaćanje i razumijevanje diskretnih slučajnih varijabli, te njihovo korištenje.

SUMMARY

This final thesis deals with the topic of discrete random variables and their application in statistics and engineering. In the first part of the thesis, we define discrete random variables and their fundamental characteristics such as probability density function, probability distribution function, expected value, variance and standard deviation. In the second part of the thesis, we analyze several important distributions, including Bernoulli, binomial, Poisson, negative binomial, geometric and hypergeometric distribution. Each distribution is illustrated with various examples and graphical representations.

The thesis aims to provide a better understanding of discrete random variables, and their use.

SADRŽAJ

1. UVOD

U suvremenoj statistici i teoriji vjerojatnosti, razumijevanje slučajnih varijabli ključno je za analizu i modeliranje različitih fenomena u prirodnim i društvenim znanostima. Među različitim vrstama slučajnih varijabli, diskretne slučajne varijable zauzimaju poseban značaj zbog svoje primjenjivosti u mnogim praktičnim situacijama, gdje su događaji ograničeni na određeni skup vrijednosti. Diskretne slučajne varijable su one varijable koje mogu poprimiti samo konačan ili brojčano beskonačan skup vrijednosti, gdje svaka vrijednost ima određenu pridruženu vrijednost.

Ovaj rad istražuje teorijske osnove diskretnih slučajnih varijabli, analizirajući njihove funkcije gustoće vjerojatnosti, funkcije distribucije vjerojatnosti, očekivanu vrijednost, varijancu te standardnu devijaciju. U ovom radu analizirali smo i distribucije karakteristične za diskretne slučajne varijable kao što su Bernoullijeva, binomna, Poissonova, negativna binomna, geometrijska te hipergeometrijska distriucija.

Bernoullijeva distribucija daje nam jednostavan, ali moćan alat za razumijevanje binarnih ishoda. Binomna distribucija govori o broju uspjeha koji se pojavljuju u n pokusa. Može se koristiti za istraživanje tržišta, kvalitete u proizvodnji, genetike, medicinskih studija te političkih anketa. Poissonova distribucija je statistička distribucija koja se koristi za modeliranje broja događaja koji se događaju u fiksnom intervalu prostora ili vremena, kada su ti događaji rijetki i nezavisni. Možemo ju koristiti u epidemiologiji, astronomiji, geologiji, za procjenu kvarova na uređajima te za procjenu defekata u proizvodnoj liniji. Negativna binomna distribucija koristi se za modeliranje broja neuspjeha (ili pokušaja) koji su potrebni da postignemo određeni broj uspjeh u seriji. Pogodna je za istraživanje kvalitete u proizvodnji, određenih kliničkih ispitivanja te u marketinškim kompanijama. Geometrijska distribucija korisna je u situacijama gdje želimo saznati koliko pokušaja i neuspjeha je potrebno do prvog uspjeha. Hipergeometrijska distribucija korisna je za modeliranje u situacijama u kojima analiziramo uzorak iz ograničenih populacija, bez vraćanja uzorka, odnosno zamjene.

Cilj ovog rada je pružiti sveobuhvatan pregled diskretnih slučajnih varijabli i njihovih distribucija, uključujući njihove matematičke osobine i primjene u praksi. Kroz analizu konkretnih primjera koji su analizirani te riješeni korištenjem računalnog programa Excel, rad nam omogućuje bolje razumijevanje diskretnih slučajnih varijabli te na koji način one mogu biti korištene u donošenju odluka te rješavanju složenih problema.

2. DISKRETNE SLUČAJNE VARIJABLE

2.1. DEFINICIJA

Slučajna varijabla je statistička varijabla čije su vrijednosti slučajne, odnosno ne mogu se predvidjeti sa sigurnošću, već samo s određenom vjerojatnošću. Izraz slučajna varijabla odnosi se na matematičku funkciju koja svakom mogućem ishodu nekog eksperimenta pridružuje određeni realan broj. Koncept slučajne varijable omogućuje nam prijelaz sa samih eksperimentalnih ishoda koji često nisu numerički na numeričke ishode. Slučajne varijable označavaju se velikim slovima npr. *X,Y, Z.*

Postoje dvije različite vrste slučajnih varijabli: diskretne slučajne varijable i kontinuirane slučajne varijable. Slučajne varijable čiji se skup mogućih vrijednosti može napisati ili kao konačni niz $x_1, \ldots, x_n$ ili kao beskonačni niz x_1, ... mogu se nazvati diskretne. Diskretne slučajne varijable su varijable koje imaju diskretan skup mogućih vrijednosti. Njihove vrijednosti čine konačan skup ili ih možemo navesti u nizu u kojem postoji prvi element, drugi element i tako dalje („prebrojiv" beskonačan skup). Primjer beskonačnog prebrojivog skupa je skup prirodnih brojeva. Kontinuirane slučajne varijable su varijable koje imaju kontinuirani skup mogućih vrijednosti.

2.2. FUNKCIJA GUSTOĆE VJEROJATNOSTI

Funkcija gustoće vjerojatnosti može se definirati kao vjerojatnost da će diskretna slučajna varijabla biti točno jednaka nekoj određenoj vrijednosti. Drugim riječima, funkcija gustoće vjerojatnosti dodjeljuje određenu vjerojatnost svakoj mogućoj vrijednosti diskretne slučajne varijable.

Neka je *X* diskretna slučajna varijabla s mogućim vrijednostima $x_1, x_2, \ldots, x_n$. Funkcija gustoće vjerojatnosti mora zadovoljavati sljedeće uvjete.

- Zbroj svih vjerojatnosti mora biti 1:

$$\sum_{i=1,\ldots,n} p(x_i) = p(x_1) + p(x_2) + \cdots p(x_n) = 1.$$

- Svaka vjerojatnost mora biti nenegativna:

$$p(x_i) \geq 0 \qquad za\ sve\ x_i.$$

Funkcija gustoće vjerojatnosti *p(x)* diskretne slučajne varijable *X* određena je izrazom:

$$p(x) = P(X = x),$$

gdje *P* označava vjerojatnost.

Možemo reći da funkcija gustoće vjerojatnosti slučajne varijable *X* daje vjerojatnost da je ishod pokusa jednak upravo *x*, za svaku vrijednost *x*. Funkcija gustoće vjerojatnosti slučajne varijable *X* pokazuje kako je ukupna vjerojatnost 1 raspodijeljena između mogućih vrijednosti varijable *X*.

Ako *X* može poprimiti neku od vrijednosti $x_1, x_2, \ldots, x_n$ onda je

$$p(x_i) > 0 \qquad i = 1,2,\ldots,n$$

,a za sve ostale vrijednosti *x* je $p(x) = 0$

Funkciju gustoće vjerojatnosti *p(x)* možemo prikazati računski - tablicom, grafički - histogramom ili analitički - formulom.

2.3. FUNKCIJA DISTRIBUCIJE VJEROJATNOSTI

Funkcija distribucije vjerojatnosti *F(x)* diskretne slučajne varijable *X* predstavlja vjerojatnost da je vrijednost slučajne varijable *X* manja ili jednaka *x*.

Funkciju distribucije vjerojatnosti *F(x)* prikazujemo izrazom:

$$F(x) = P(X \leq x) = \textstyle\sum_{x_i \leq x} p(x_i).$$

Ako je *X* diskretna slučajna varijabla čije su moguće vrijednosti $x_1, x_2, x_3, \ldots$, gdje je $x_1 < x_2 < x_3$ tada je njena funkcija distiucije vjerojatnosti *F(x)* funkcija skoka. To znači da je vrijednost *F(x)* konstantna u intervalima $[x_{i-1}, x_i)$ te se zatim pomakne (ili skoči) za veličinu $f(x_i)$ na x_i.

Za bilo koja dva broja *a* i *b* gdje je $a \leq b$, vrijedi:

$$P(a \leq X \leq b) = F(b) - F(a^{-}),$$

gdje simbol a^{-} označava najveću moguću vrijednost *X* koja je strogo manja od *a*.

Ako su jedine moguće vrijednosti slučajne varijable *X* cijeli brojevi i ako su *a* i *b* također cijeli brojevi vrijedi:

$$P(a \leq X \leq b) = P(X = a \; ili \; a + 1 \; ili \ldots ili \; b)$$

$$= F(b) - F(a - 1).$$

Funkciju distribucije vjerojatnosti možemo opisati i kao kumulativnu funkciju gustoće vjerojatnosti.

2.4. OČEKIVANJA DISKRETNE SLUČAJNE VARIJABLE

Jedan od najvažniji koncepata u teoriji vjerojatnosti jest onaj o očekivanju slučajne varijable. Očekivanje slučajne varijable označava se kao *E(X),* μ_x, ili samo μ i predstavlja težinski prosjek svih njenih vrijednosti.

Ako je *X* diskretna slučajna varijabla sa skupom mogućih vrijednosti *D* i funkcijom vjerojatnosti *p(x), E(X)* definira se kao:

$$E(X) = \mu_x = \sum_{x \in D} x * p(x).$$

Drugim riječima očekivana vrijednost *X* predstavlja prosjek svih mogućih vrijednosti koje *X* može imati, pri čemu je svaka vrijednost pomnožena sa svojom vrijednosti.

Ako je *X* diskretna slučajna varijabla s funkcijom vjerojatnosti *p(x)* tada za bilo koju funkciju *g*, očekivanje *g(X)* je definirano kao:

$$E[g(X)] = \sum_x g(x)p(x).$$

Na primjer, ako su *a* i *b* konstantni i $g(x) = ax + b$, onda za diskretne slučajne varijable vrijedi:

$$E[aX + b] = \sum_x (ax + b)p(x)$$

$$= a \sum_x xp(x) + b \sum_x p(x)$$

$$= aE[X] + b.$$

2.5. VARIJANCA

Uzimajući u obzir slučajnu varijablu *X* zajedno s njenom funkcijom gustoće vjerojatnosti, bilo bi izuzetno korisno kada bismo mogli sažeti bitna svojstva vjerojatnosti pomoću određenih prikladno definiranih mjera. Jedna takva mjera je *E(X)*, očekivana vjerojatnost *X*. Međutim, dok *E(X)* daje težinski prosjek mogućih vrijednosti *X,* ne govori ništa o raspršenosti tih vrijednosti. Varijanca predstavlja mjeru raznolikosti, odnosno raspršenosti vrijednosti slučajne varijable oko njene očekivane vrijednosti (srednje vrijednosti). U kontekstu slučajnih varijabli varijanca predstavlja odstupanje od očekivane vrijednosti slučajne varijable.

Neka je *X* diskretna slučajna varijabla s vrijednostima u skupu *D* i funkcijom gustoće vjerojatnosti *p(x),* tada varijancu označavamo s *VAR(X)* i definira se kao:

$$VAR(X) = \sum_{x(\in)D} (x - E(X))^2 * p(x) = E[(X - E(X))^2].$$

Možemo koristiti i drugu formulu, [1]:

$$VAR(X) = E[X^2] - (E[X])^2.$$

2.6. STANDARDNA DEVIJACIJA

Standardna devijacija je, slično kao i varijanca, mjera raspršenosti za distribuciju slučajne varijable koja određuje stupanj u kojem se vrijednosti razlikuju od očekivane.

Standardna devijacija slučajne varijable *X* označava se σ ili σ*X*. Standardna devijacija definira se kao pozitivni kvadratni korijen iz varijance, tj vrijedi

$$VAR(X) = \sigma^2.$$

Važno je za razumjeti interpretaciju standardne devijacije.

- Mala standardna devijacija znači da je distribucija slučajne varijable usko koncentrirana oko srednje vrijednosti tj očekivanja.
- Velika standardna devijacija znači da je distribucija raširena, te imamo vrijednosti slučajne varijable s relativno velikom vjerojatnosti koje su dosta udaljene od očekivanja, [3].

3. BERNOULLIJEVA DISTRIBUCIJA

3.1. DEFINICIJA

Bernoullijeva distribucija je najjednostavnija vrsta distribucije diskretne slučajne varijable. Predstavlja diskretnu slučajnu distribuciju vjerojatnosti slučajne varijable X koja ima samo dva moguća ishoda, a to su:

- Neuspjeh (oznaka 0) koji se prikazuje vjerojatnošću $(1 - p)$,
- Uspjeh (oznaka 1) koja se prikazuje vjerojatnošću p.

Drugim riječima, može se smatrati modelom za skup mogućih ishoda eksperimenta koji postavlja pitanje da-ne. Neki primjeri Bernoullijeve distribucije uključuju:

- Polažete ispit gdje je mogućnost ili prolaz *(X=1)* ili pad *(X=0),*
- Bacate novčić gdje ishod može biti ili glava ili pismo,
- Dijete je rođeno, spol može biti ili muško ili žensko.

Funkcija gustoće vjerojatnosti Bernoullijeve slučajne varijable poprima samo dvije nenegativne vrijednosti:

$$P(X = 1) = p,$$

$$P(X = 0) = 1 - p,$$

gdje je p, $0 \leq p \leq 1$, vjerojatnost da je pokus „uspješan“, [4].

Očekivana vrijednost Bernoullijeve slučajne varijable iznosi

$$E[X] = 1 * P\{X = 1\} + 0 * P\{X = 0\} = p.$$

To znači da je očekivanje Bernoullijeve slučajne varijable vjerojatnost da slučajna varijabla bude jednaka 1.

Varijanca Bernoullijeve distribucije jednaka je:

$$VAR(X) = pq = p(1 - p).$$

Primarna svrha Bernoullijeve distribucije je da služi kao matematički model za eksperimente, događaje i procese koji imaju točno dva različita ishoda. Ukratko, Bernoullijeva distribucija daje nam jednostavan, ali moćan alat za razumijevanje binarnih ishoda. Dodjeljivanjem vjerojatnosti ishodima Bernoullijeva distribucija pomaže u predviđanju nastanka događaja, pružajući ključne podatke za donošenje odluka u znanosti, poslovanju i svakodnevnom životu. Razumijevanje principa ove distribucije predstavlja osnovu za primjenu u stvarnim situacijama.

3.2. PRIMJERI

PRIMJER 1

Testiramo žarulju novog proizvođača. Radi li ova žarulja? Odgovor može biti ili da (uspjeh) ili ne (neuspjeh). Ako je vjerojatnost da žarulja radi 0,8, znamo da je vjerojatnost kvara 1-0,8=0,2. Izračunat ćemo varijancu i očekivanu vrijednost.

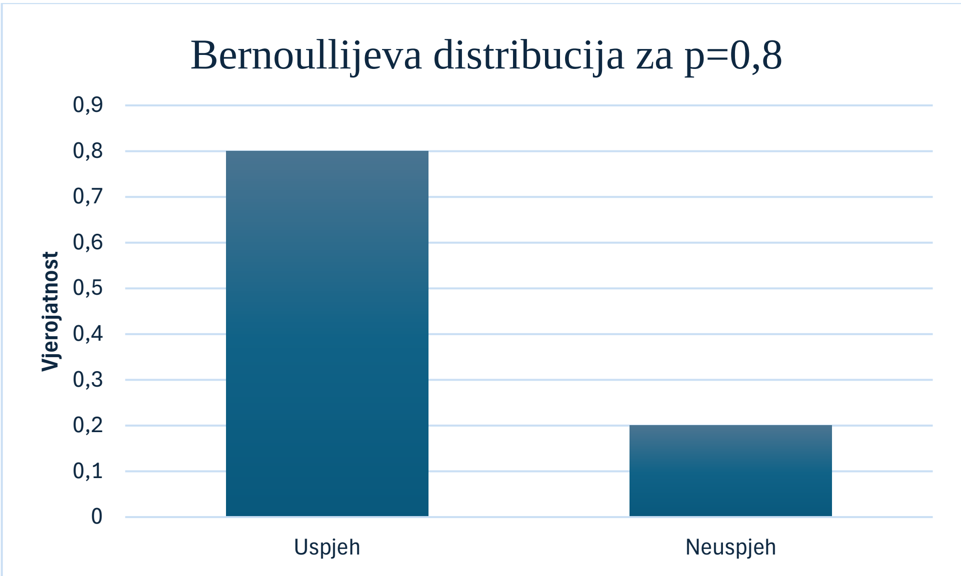
Očekivanje:

$$E[X] = 1 * P\{X = 1\} + 0 * P\{X = 0\} = p = 0{,}8$$

Varijanca:

$$VAR(X) = pq = p(1 - p) = 0{,}8(1 - 0{,}8) = 0{,}16$$

Na Slici 1. prikazan je grafički prikaz Bernoullijeve distribucije za *p=0,8* :



Slika 1. Grafički prikaz Bernoullijeve distribucije za p=0,8

PRIMJER 2

Nogometaš izvede 7 neovisnih slobodnih udarca, s vjerojatnošću 0,65 da postigne gol pri svakom udarcu. Odredite:

a) Broj pokusa,
b) Vjerojatnost da ne postigne gol pri svakom udarcu,
c) Očekivanu vrijednost i varijancu.

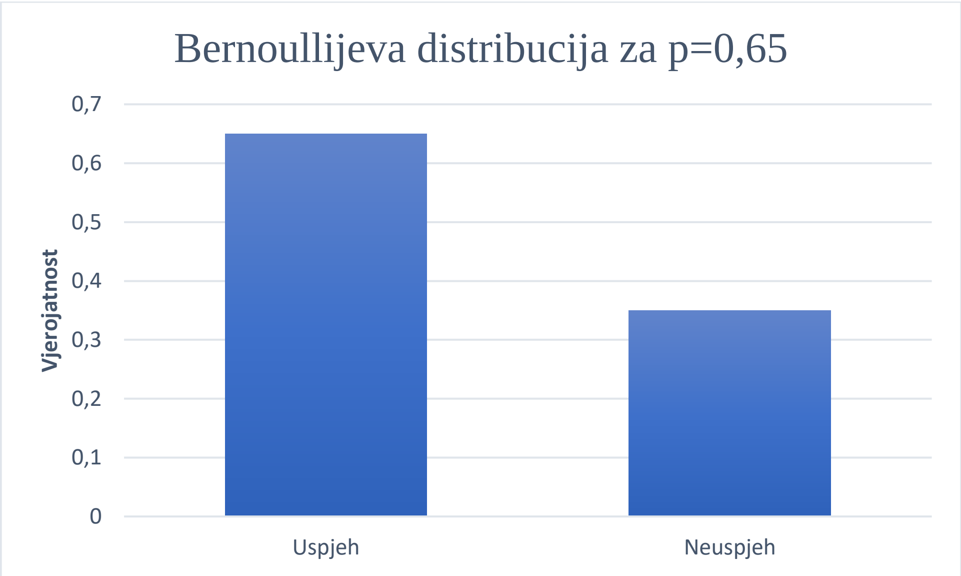
RIJEŠENJE

a) Broj pokusa odgovara broju neovisnih slobodnih udarca, u ovom slučaju 7.
b) Vjerojatnost neuspjeh je $1 - p = 1 - 0{,}65 = 0{,}35$.
c) Očekivanje i varijanca iznose redom:

$$E[X] = p = 0{,}65,$$

$$VAR(X) = p(1 - p) = 0{,}65(1 - 0{,}65) = 0{,}2275.$$

Grafički prikaz Bernoullijeve distribucije za p=0,65 prikazan je na Slici 2.



Slika 2. Grafički prikaz Bernoullijeve distribucije za p=0,65

4. BINOMNA DISTRIBUCIJA

4.1. DEFINICIJA

Binomna distribucija pojavljuje se kada postoji više eksperimenata koji točno ili približno odgovaraju sljedećim zahtjevima.

- Eksperiment se sastoji od *n* manjih eksperimenata, koji se nazivaju pokusi, gdje je broj pokusa *n* određen prije početka eksperimenta.
- Svaki pokušaj može rezultirati jednim od dva moguća ishoda, koje općenito označavamo uspjehom *(S)* ili neuspjehom *(F)*. To znači da je svaki pokušaj jedna Bernoullijeva slučajna varijabla.
- Ispitivanja su neovisna, tako da ishod bilo kojeg pojedinog pokušaja ne utječe na ishod bilo kojeg drugog pokušaja.
- Vjerojatnost uspjeha *P(S)* je konstanta iz pokušaja u pokušaj, te tu vjerojatnost označavamo sa p.

Ako eksperiment zadovoljava sva četiri zahtjeva tada taj eksperiment nazivamo binomni eksperiment.

Pretpostavimo da treba izvesti *n* neovisnih pokusa, od kojih svaki rezultira „uspjehom" s vjerojatnošću p i „neuspjehom" s vjerojatnošću $1-p$. Ako *X* predstavlja broj uspjeha koji se pojavljuju u *n* pokusa, tada za *X* kažemo da je to slučajna binomna varijabla s parametrima (n, p).

Oznaka: $X \sim B(n, p)$.

Funkcija gustoće vjerojatnosti binomne slučajne varijable s parametrima *n* i *p* dana je izrazom:

$$P(X=i) = \binom{n}{i} p^i (1-p)^{n-1}, \qquad i = 0,1,\dots,n.$$

Tumačenje članova gornjeg izraza je sljedeće:

- $\binom{n}{i}$ predstavlja binomni koeficijent, odnosno broj načina na koji se može odabrati *i* uspjeha iz *n* pokusa,
- p predstavlja vjerojatnost uspjeha u svakom pokusu,
- $(1-p)$ je vjerojatnost neuspjeha u svakom pokusu,
- i je broj uspjeha u *n* pokusa, [2].

Funkcija distribucije vjerojatnosti binomne slučajne varijable s parametrima n i p dana je izrazom:

$$P(X \leq x) = \sum_{i=0}^{x} P(X = i).$$

Budući da binomna slučajna varijabla *X*, s parametrima *n* i *p*, predstavlja broj uspjeha u *n* neovisnih pokusa, od kojih svaki ima vjerojatnost uspjeha *p*, možemo *X* predstaviti na sljedeći način:

$$X = \sum_{i=1}^{n} X_i,$$

gdje je :

$$X_i = \begin{cases} 1 & ako\ i\ u\ pokusu\ predstavlja\ uspjeh \\ 0 & ako\ je\ neuspjeh \end{cases}$$

Kako su $X_i = 1, \dots, n$ neovisne Bernoullijeve slučajne varijable, imamo:

$$E[X_i] = P\{X_i = 1\} = p$$

$$VAR(X_i) = E[X_i^2] - p^2$$

$$= p(1-p).$$

Posljednja jednakost slijedi $X_i^2 = X_i$ pa je stoga $E[X_i^2] = E[X_i] = p$.

Koristeći formulu iznad možemo izračunati očekivanje i varijancu od *X*:

$$E[X] = \sum_{i=1}^{n} E[X_i]$$

$$= np,$$

$$VAR(X) = \sum_{i=1}^{n} VAR(X_i) \text{ pošto su } X_i$$

$$= np(1-p).$$

Ako su X_1 i X_2 neovisne binomne slučajne varijable s odgovarajućim parametrima (n_1, p) i (n_2, p) tada je i njihova suma binomna sa parametrima $(n_1 + n_2, p)$. Budući da $X_i, i = 1,2$, predstavlja broj uspjeha u n_i neovisnih pokusa, pri čemu je vjerojatnost uspjeha p, suma $X_1 + X_2$ predstavlja broj uspjeha u $n_1 + n_2$ neovisnih pokusa, pri čemu je vjerojatnost uspjeha također p.

4.2. PRIMJERI

Za izračun i grafički prikaz binomne distribucije navedenih primjera korišten je program Microsoft Excel. U Microsoft Excelu binomnu distribuciju računamo pomoću funkcije

$$BINOM.DIST(i, n, p, FALSE)$$

što nam predstavlja funkciju gustoće vjerojatnosti, dok nam funkcija

$$BINOM.DIST(i, n, p, TRUE)$$

predstavlja računanje funkcije distribucije vjerojatnosti.

PRIMJER 1

Jedna telefonska kompanija izvještava da može riješiti probleme svojih korisnika istog dana kada su oni prijavljeni u 60% slučajeva. Pretpostavimo da je danas prijavljeno 15 kvarova.

a) Kolika je vjerojatnost da će 9 problema biti riješeno danas?
b) Kolika je vjerojatnost da će 9 ili 10 problema biti riješeno danas?
c) Kolika je vjerojatnost da će najviše 8 problema biti riješeno danas?
d) Kolika je vjerojatnost da će više od 10 problema biti riješeno danas?

RIJEŠENJE:

a) Parametri binomne distribucije su $n = 15$ i $p = 0,60$

Koristimo formulu za funkciju gustoće vjerojatnosti

$$P(X = i) = \binom{n}{i} p^i (1 - p)^{n-1}.$$

Za $i = 9$ dobijemo

$$P(X = 9) = \binom{15}{9} 0,60^9 (1 - 0,60)^{15-9} = 0,2066$$

U Excelu računamo pomoću funkcije *BINOM.DIST* :

$$= BINOM.DIST(9{,}15{,}0{,}60, FALSE) = 0{,}2066.$$

Što znači da je vjerojatnost da će 9 problema biti riješeno danas jednaka 0,2066, odnosno 20,66%.

b)

$$P(9\ ili\ 10) = P(X = 9) + P(X = 10)$$

$$= \binom{15}{9} 0{,}60^{9}(1 - 0{,}60)^{15-9} + \binom{15}{10} 060^{10}(1 - 0{,}60)^{15-10}$$

$$= 0{,}3925.$$

U Excelu računamo:

$$= BINOM.DIST(9{,}15{,}0{,}60, FALSE) + BINOM.DIST(10{,}15{,}0{,}60, FALSE) = 0{,}3925$$

Što znači da je vjerojatnost da će 9 ili 10 problema biti riješeno danas jednaka 0,3925, odnosno 39,25%.

c)

$$P(x \leq 8) = P(0) + P(1) + P(2) + P(3) + P(4) + P(5) + P(6) + P(7) + P(8)$$

$$= 0{,}3902.$$

U Excelu računamo:

$$= BINOM.DIST(8{,}15{,}0{,}60, TRUE) = 0{,}3902.$$

Što znači da je vjerojatnost da je vjerojatnost da će najviše 8 problema biti riješeno danas jednaka 0,3902, odnosno 39,02%.

Rezultati dobiveni u Excelu za funkciju gustoće vjerojatnosti *p(x)* i funkciju distribucije vjerojatnosti *F(X)* prikazani su u Tablici 1.

x	p(x)	F(X)
0	1,07374E-06	1,07374E-06
1	2,41592E-05	2,52329E-05
2	0,000253672	0,000278904
3	0,001648865	0,001927769
4	0,007419892	0,009347661
5	0,024485642	0,033833303
6	0,061214105	0,095047408
7	0,118055774	0,213103183
8	0,177083662	0,390186844
9	0,206597605	0,59678445
10	0,185937845	0,782722294
11	0,126775803	0,909498098
12	0,063387902	0,972885999
13	0,021941966	0,994827965
14	0,00470185	0,999529815
15	0,000470185	1

Tablica 1. Dobivene vrijednosti p(x) i F(x)

d)

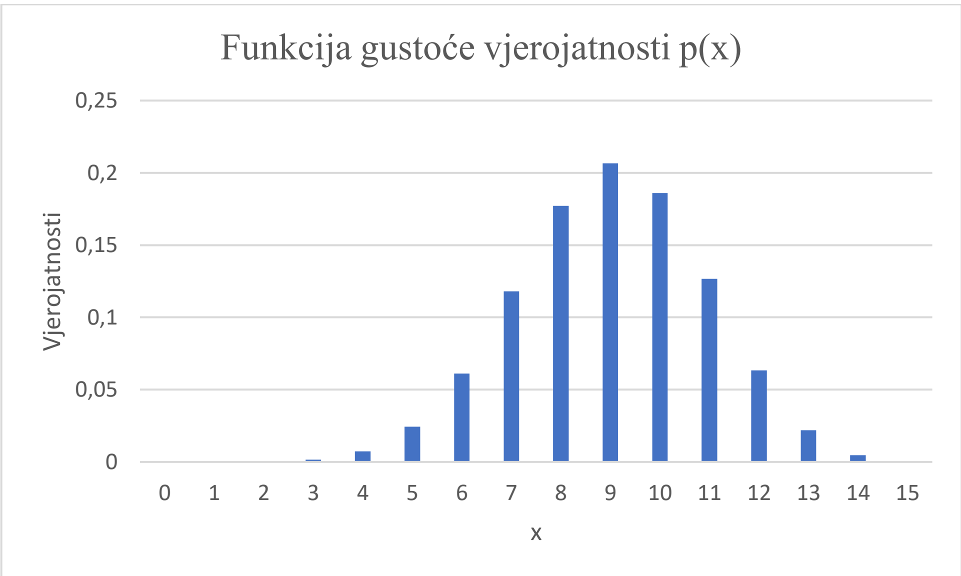
$$P(x > 10) = 1 - P(X \leq 10) = 1 - 0{,}7827 = 0{,}2173$$

U Excelu računamo na sljedeći način:

$$= 1 - BINOM.DIST(10{,}15{,}0{,}60, TRUE) = 0{,}2173$$

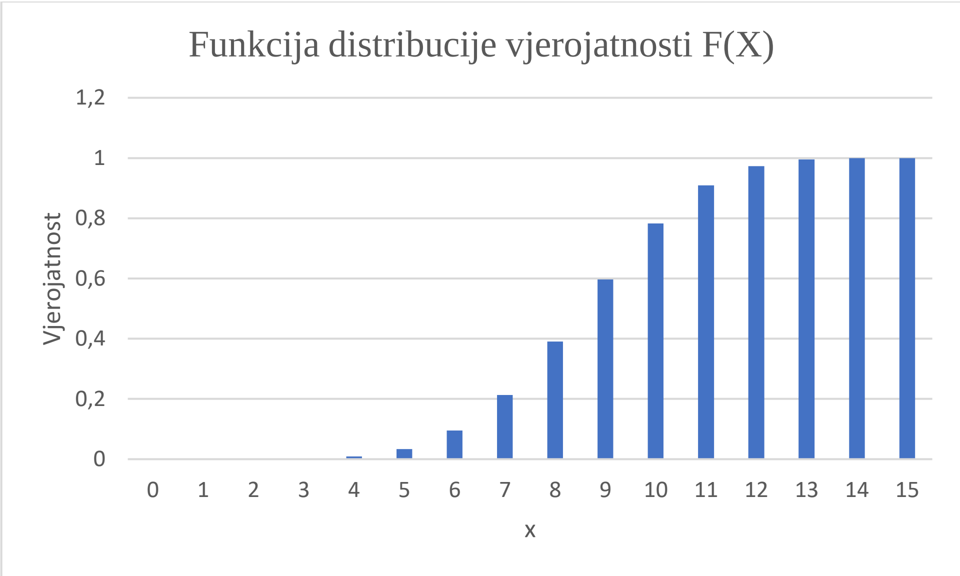
Vjerojatnost da će više od 10 problema biti riješeno jednaka je 0,2173, odnosno 21,73%.

Na Slici 3. prikazan je graf funkcije gustoće vjerojatnosti *p(x)* za primjer 1:



Slika 3. Funkcija gustoće vjerojatnosti p(x)

Na Slici 4. prikazan je graf funkcije distribucije vjerojatnosti *F(x)* za primjer 1:



Slika 4. Funkcija distribucije vjerojatnosti F(x)

PRIMJER 2

Poznato je da neki lijek uzrokuje negativne nuspojave kod 3 od svakih 100 pacijenata. Promatramo 5 osoba koje su koristile lijek. Koja je vjerojatnost sljedećih događaja?

a) Niti jedan od pet pacijenata nije imao nuspojave.
b) Najmanje dvoje pacijenata je imao nuspojave.
c) Koliki je očekivani broj pacijenata koji će imati nuspojave ako nasumično odaberemo 100 pacijenata?

RIJEŠENJE:

a)

$$P(X=0)=\binom{5}{0}0{,}03^{0}(0{,}97)^{5-0}=0{,}8687$$

U Excelu računamo pomoću funkcije *BINOM.DIST*:

$$=BINOM.DIST(0{,}5{,}0{,}03,FALSE)=0{,}8687$$

Vjerojatnost da niti jedan od pacijenata nije imao nuspojave jednaka je 0,8687, odnosno 86,87%.

b)

$$P(X \geq 2) = 1 - P(X < 2) = 1 - [P(X = 0) + P(X = 1)]$$
$$= 1 - [\binom{5}{0} 0{,}03^0 \left(0{,}97\right)^{5-0} + \binom{5}{1} 0{,}03^1 * 0{,}97^{5-1}\right] = 0{,}0085$$

U Excelu računamo:

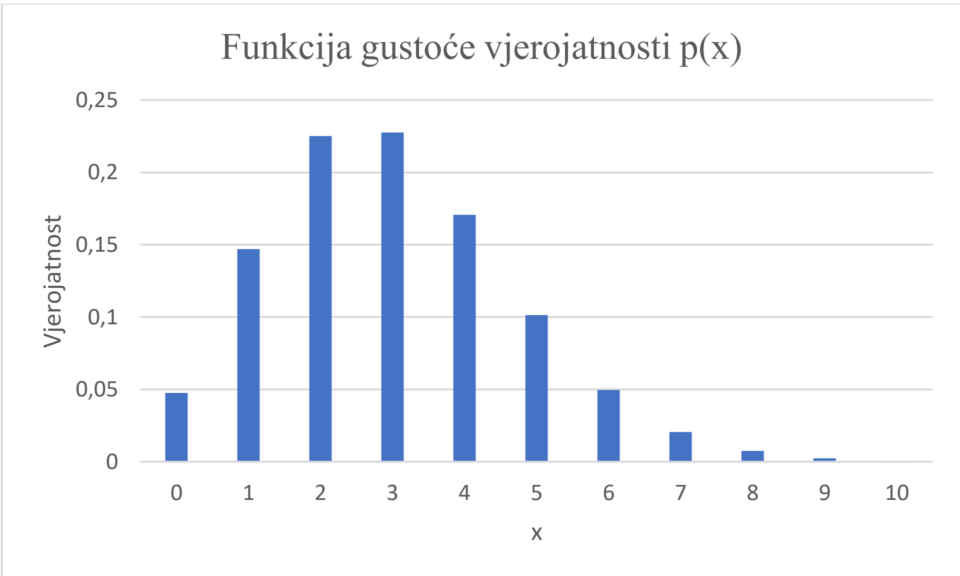
$$= 1 - BINOM.DIST(0{,}5{,}0{,}03, TRUE) = 0{,}0085$$

Vjerojatnost da je najmanje dvoje pacijenata imalo nuspojave jednaka je 0,0085, odnosno 0,85%.

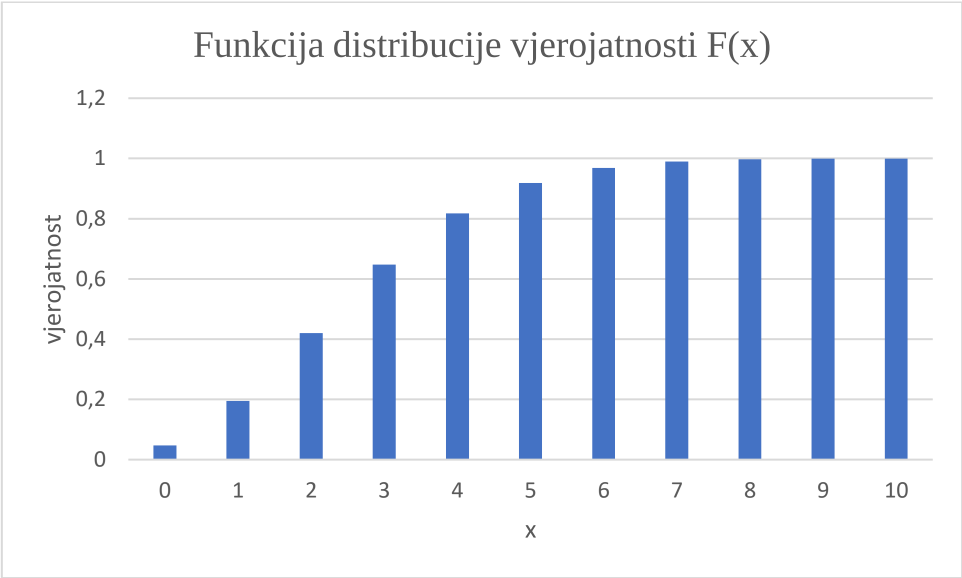
c) Sada promatramo 100 pacijenata pa je $n = 100$, a p je i dalje 0,03. Koristimo formulu za očekivanje:

$$E(X) = np = 100 * 0{,}03 = 3$$

Slika 5. i Slika 6. predstavljaju grafički prikaz funkcije gustoće vjerojatnosti *p(x)* i funkcije distribucije vjerojatnosti *F(x)* za primjer 2:



Slika 5.. Funkcija gustoće vjerojatnosti p(x)



Slika 6. Funkcija distribucije vjerojatnosti F(x)

5. POISSONOVA DISTRIBUCIJA

5.1. DEFINICIJA

Poissonova slučajna varijabla je poseban tip diskretne slučajne varijable. Daje vjerojatnost da će se događaj dogoditi određeni broj puta unutar zadanog intervala. Poissonova distribucija se primjenjuje na događaje koji se mogu dogoditi veliki broj puta, ali je svaki od tih događaja relativno rijedak. Poissonova slučajna varijabla postiže skup vrijednosti koji je beskonačan, ali prebrojiv.

Oznaka Poissonove varijable: $X \sim P(\lambda)$.

Za diskretnu slučajnu varijablu *X*, koja poprima jednu od vrijednosti $0,1,2,3,\dots$ kaže se da je Poissonova slučajna varijabla s parametrom $\lambda, \lambda > 0$. Taj parametar predstavlja očekivani broj događaja u promatranom intervalu. Njena funkcija gustoće vjerojatnosti dana je izrazom:

$$P\{X = i\} = e^{-\lambda}\frac{\lambda^i}{i!} \qquad i = 0,1,2,\dots$$

Simbol *e* predstavlja konstantu vrijednosti približno jednake 2,7183. To je poznata konstanta u matematici, imenovana po švicarskom matematičaru L. Euleru, [2].

Funkcija distribucije vjerojatnosti prikazana je formulom:

$$P(X \leq x) = \sum_{i=0}^{x} P(X = i).$$

Očekivanje i varijancu Poissonove slučajne varijable su jednaki i iznose

$$E(X) = VAR(X) = \lambda.$$

Standardna devijacija Poissonove distribucije računa se izrazom

$$\sigma = \sqrt{\lambda}.$$

Poissonova slučajna varijabla ima širok raspon primjena u raznim područjima zato što se može koristiti kao aproksimacija binomne slučajne varijable s parametrima (n, p) kada je *n* veliki, a *p* mali. Da bismo to vidjeli pretpostavimo da je *X* binomna slučajna varijabla s parametrima (n, p) i neka je $\lambda = np$. Tada je

$$P(X = i) \approx e^{-\lambda}\frac{\lambda^i}{i!}.$$

Drugim riječima, ako se provodi n neovisnih pokusa, od kojih svaki rezultira „uspjehom“ s vjerojatnošću p, tada kada je n veliki, a p mali broj uspjeha koji se dogodi približno slijedi Poissonovu slučajnu varijablu sa srednjom vrijednošću $\lambda = np$.

Ovu aproksimaciju možemo koristiti u praksi kada je $n > 50$, a $np < 5$.

Neki primjeri slučajnih varijabli koji obično slijede Poissonov zakon vjerojatnosti:

- Broj pogrešno biranih telefonskih brojeva u jednom danu,
- Broj kupaca koji ulaze u poštanski ured određenog dana,
- Broj a- čestica otpuštenih u nekom određenom vremenskom periodu iz radioaktivne čestice, [5].

Primjeri u kojima se radi o binomnoj distribuciji koju aproksimiramo Poissonovom: broj tiskarskih pogrešaka na strnici knjige i broj ljudi koji dožive 100 godina u zajednici.

Poissonova distribucija vizualno se može prikazati koristeći graf funkcije gustoće vjerojatnosti. Najvjerojatniji broj događaja prikazan je vrhom distribucije - modom.

- Kada λ nije cijeli broj, mod je najbliži cijeli broj od λ.
- Kada je λ cijeli broj, postoje dva moda λ i λ-1.

Poissonova distribucija koristi se za modeliranje pojave rijetkih događaja unutar fiksnog intervala uz konstantu prosječne brzine pojavljivanja i uz pretpostavku da se svaki događaj odvija neovisno. Poissonova distribucija ima sposobnost predviđanja vjerojatnosti zadanog broja događaja, na temelju poznate prosječne stope. To ju čini vrijednom u širokom rasponu primjena od analize protoka prometa do procjene rizika u financijama.

5.2. PRIMJERI

Primjere smo rješavali u programu Microsoft Excel pomoću funkcije

$$POISSON.DIST(n, \lambda, FALSE)$$

gdje nam *FALSE* ukazuje na to da je riječ o funkciji gustoće vjerojatnosti, te pomoću funkcije

$$POISSON.DIST(n, \lambda, TRUE)$$

gdje nam *TRUE* ukazuje na to da je riječ o kumulativnoj funkciji, odnosno funkciji distribucije vjerojatnosti.

PRIMJER 1

U banku u jednom danu u prosjeku uđe 5 osoba u vremenskom periodu od 10 minuta.

a) Koja je vjerojatnost da će u 10 minuta točno tri osobe ući u banku?
b) Koja je vjerojatnost da će u banku ući točno tri osobe u 20 minuta?
c) Koja je vjerojatnost da će najviše tri osobe ući u banku u 20 minuta?
d) Koja je vjerojatnost da će više od tri osobe ući u banku u 20 minuta?

RIJEŠENJE

a) Koristimo formulu za funkciju gustoće Poissonove slučajne varijable

$$P(X = x) = e^{-\lambda} \frac{\lambda^i}{i!}.$$

Kada promatramo vremenski period od 10 minuta vrijedi $\lambda = 5$.

$$P(X = 3) = \frac{e^{-5} 5^3}{3!} = 0{,}1404$$

U Excelu ovaj primjer računamo preko funkcije POISSON.DIST na sljedeći način:

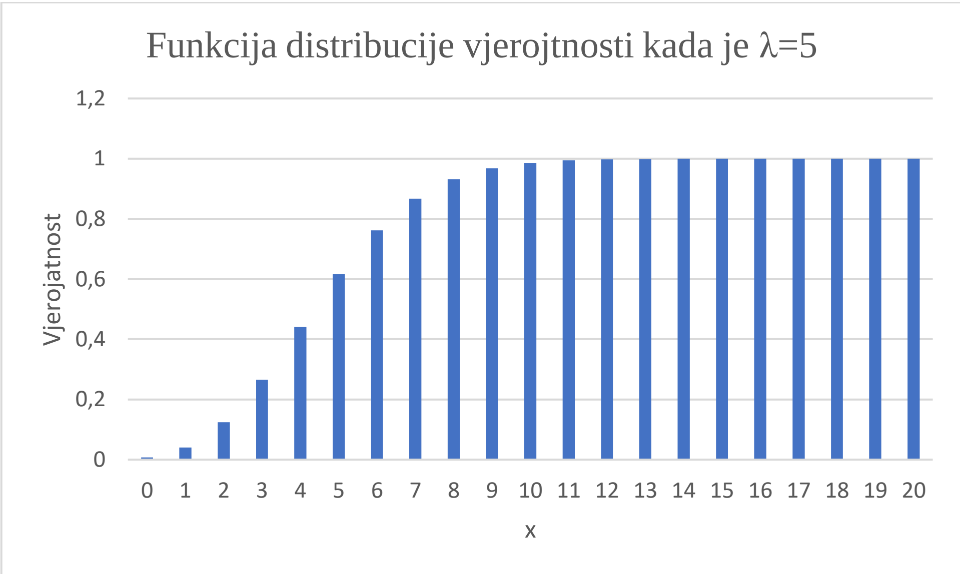
$$= POISSON.DIST(3{,}5, FALSE) = 0{,}1404$$

Vjerojatnost da će u 10 minuta u banku ući točno tri osobe jednaka je 0,1404, odnosno 14,04%.

Slika 7. prikazuje grafički prikaz funkcije gustoće vjerojatnosti Poissonove distribucije kada je λ*=5*

Slika 7. Funkcija gustoće vjerojatnosti Poissonove distribucije kada je λ=5

Slika 8. je grafički prikaz funkcije distribucije vjerojatnosti Poissonove distribucije kada je λ=5



Slika 8. Funkcija distribucije vjerojatnosti Poissonove distribucije kada je λ=5

b) Kada promatramo vremenski period od 20 minuta, parametar λ je dvostruko veći nego u a) dijelu zadatka.

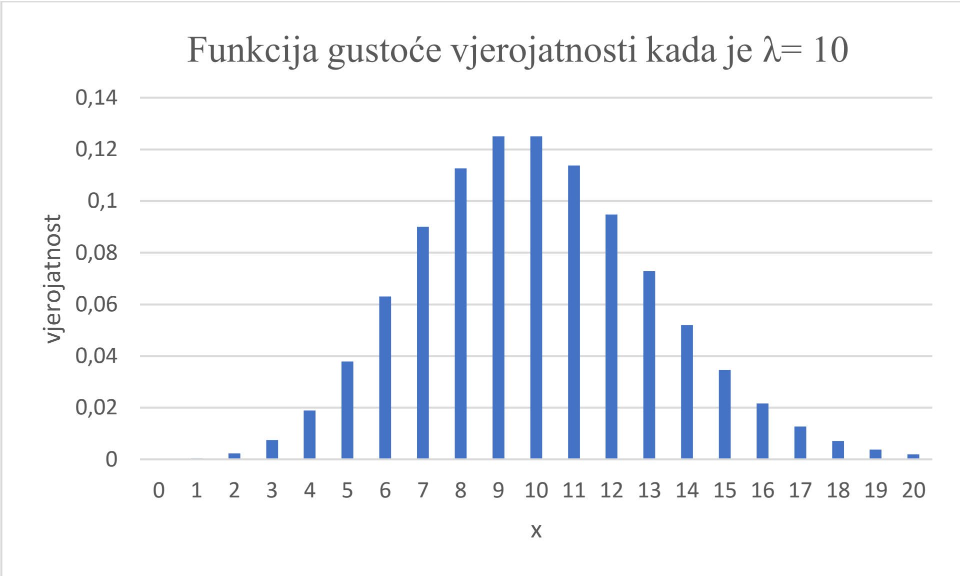
$$P(X = 3) = \frac{e^{-10} 10^3}{3!} = 0{,}0076$$

U Excelu:

$$= POISSON.DIST(3{,}10, FALSE) = 0{,}0076$$

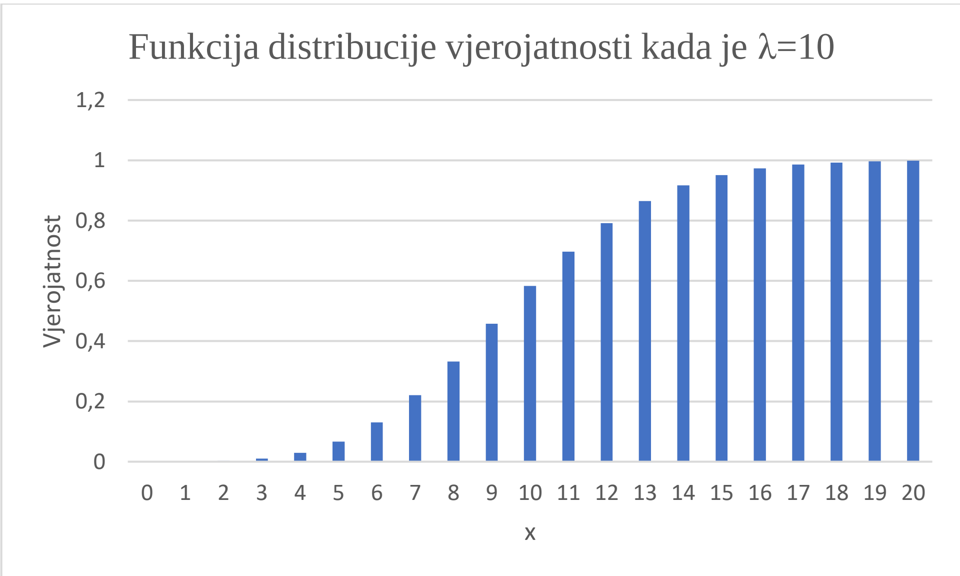
Vjerojatnost da će u banku ući točno tri osobe u vremenskom intervalu od 20 minuta jednaka je 0,0076, odnosno 0,76%.

Slika 9. prikazuje graf funkcije gustoće vjerojatnosti Poissonove distribucije kad je $\lambda = 10$



Slika 9. Funkcija gustoće vjerojatnosti Poissonove distribucije kada je $\lambda = 10$

Slika 10. je grafički prikaz funkcije distribucije vjerojatnosti Poissonove distribucije kada je λ=10



Slika 10. Funkcija distribucije vjerojatnosti Poissonove distribucije kada je λ=10

c)

$$P(X \leq 3) = P(0) + P(1) + P(2) + P(3)$$

$$= \frac{e^{-10}10^0}{0!} + \frac{e^{-10}10^1}{1!} + \frac{e^{-10}10^2}{2!} + \frac{e^{-10}10^3}{3!}$$

$$= e^{-10}\left(\frac{1}{0!} + \frac{10}{1!} + \frac{10^2}{2!} + \frac{10^3}{3!}\right) = 0{,}0103$$

U Excelu:

$$= POISSON.DIST(3{,}10, TRUE) = 0{,}0103$$

Vjerojatnost da će najviše tri osobe ući u banku u vremenskom intervalu od 20 minuta jednaka je 0,0103, odnosno 1,03%.

d)

$$P(X > 3) = 1 - P(X \leq 3)$$

$$= 1 - 0{,}0103 = 0{,}9897$$

U Excelu:

$$= 1 - POISSON.DIST(3{,}10, TRUE) = 0{,}9897$$

Vjerojatnost da će više od tri osobe ući u banku i vremenskom intervalu od 20 minuta jednaka je 0,9897, odnosno 98,97%.

PRIMJER 2

Ako je vjerojatnost loše reakcije na lijek 0,002, odredite šansu da će od 1000 osoba više od 3 imati lošu reakciju.

RIJEŠENJE

Ovdje se radi o binomnoj slučajnoj varijabli s parametrima $n = 1000$ i $p = 0{,}002$. Ako bismo računali direktno $\lambda = np = 2$

$$P(X > 3) = 1 - \{P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)\}$$

$$P(X = 0) = \frac{2^0 e^{-2}}{0!} = 1e^{-2}$$

$$P(X = 1) = \frac{2^1 e^{-2}}{1!} = 2e^{-2}$$

$$P(X = 2) = \frac{2^2 e^{-2}}{2!} = 2e^{-2}$$

$$P(X = 3) = \frac{2^3 e^{-2}}{3!} = 1{,}33e^{-2}$$

$$P(X > 3) = 1 - e^{-2} * (\frac{2^0}{0!} + \frac{2}{1!} + \frac{4}{2!} + \frac{8}{3!})$$

$$= 1 - 0{,}8571 = 0{,}1429$$

U Excelu računamo:

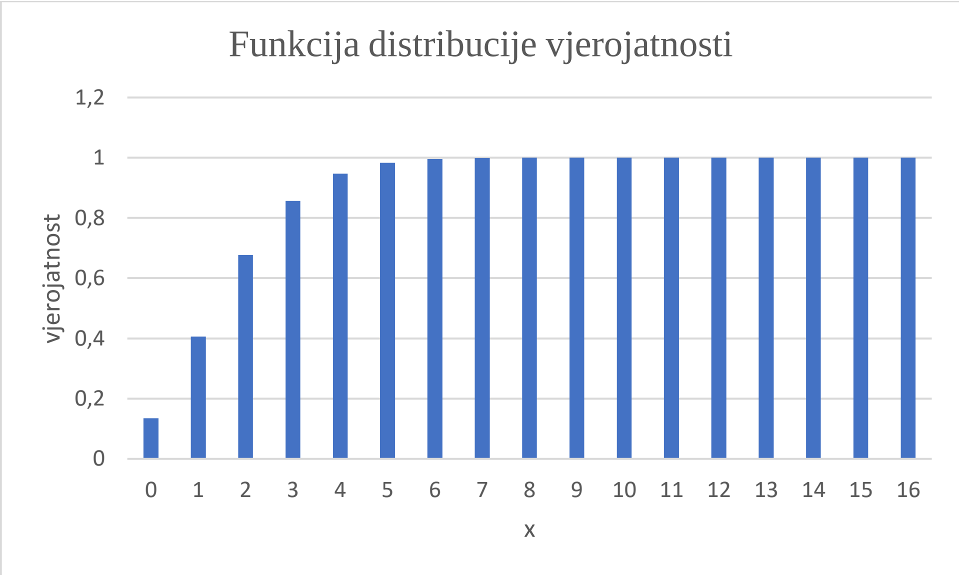
$$= 1 - POISSON.DIST(3{,}2{,}TRUE) = 0{,}1429$$

Što znači da je vjerojatnost da ćeviše od troje ljudi imati lošu reakciju na lijek jednaka 0,1429, odnosno 14,29%.

Slika 11. predstavlja grafički prikaz funkcije gustoće vjerojatnosti Poissonove distribucije kada je *λ=2*

Slika 11. Funkcija gustoće vjerojatnosti Poissonova distribucija kada je λ=2

Slika 12. predstavlja grafički prikaz funkcije distribucije vjerojatnosti Poissonove distribucije kada je *λ=2*



Slika 12. Funkcija distribucije vjerojatnosti Poissonove distribucije kada je λ=2

6. NEGATIVNA BINOMNA DISTRIBUCIJA

6.1. DEFINICIJA

U teoriji vjerojatnosti i statistici, negativna binomna distribucija je diskretna distribucija vjerojatnosti koja modelira broj neuspjeha u nizu neovisnih i identično raspodijeljenih Bernoulijevih pokušaja prije nego što se dogodi određeni (unaprijed zadan) broj uspjeha, označen kao *r*. Negativna binomna distribucija je distribucija broja pokušaja potrebnih za postizanje r-tog uspjeha. Negativna binomna distribucija je gotovo identična binomnoj distribuciji s jednom razlikom: u binomnoj distribuciji imamo fiksan broj pokušaja, dok u negativnoj binomnoj distribuciji imamo fiksan broj uspjeha, [6].

Oznaka negativne binomne distribucije:

$$X \sim NB(r,p).$$

Smatra se da slučajna varijabla *X* slijedi negativnu binomnu distribuciju ako je njena funkcija gustoće vjerojatnosti prikazana izrazom:

$$f(k;r,p) = \mathrm{P}(X=k) = \binom{k+r-1}{k}(1-p)^k p^r,$$

gdje je *r* broj uspjeha, *k* broj neuspjeha, a p vjerojatnost uspjeha svakog pokušaja. Od $k+r$ događaja znamo da je zadnji događaj uspjeh. Stoga u binomnom koeficjentu promatramo $k+r-1$ događaj.

Očekivanje negativne binomne distribucije prikazano je izrazom:

$$E(X) = \frac{r(1-p)}{p}.$$

Varijancu negativne binomne distribucije prikazujemo izrazom:

$$V(X) = \frac{r(1-p)}{p^2}.$$

Negativnu binomnu distribucije možemo primijeniti u različitim aspektima poslovne statistike kao što su: upravljanje projektima, kontrola kvalitete, službe za korisnike i drugi.

6.2. PRIMJERI

Primjeri su rađeni u programu Microsoft Excel pomoću funkcije

$$NEGBINOM.DIST(k, r, p, FALSE)$$

gdje F*ALSE* ukazuje na to da je riječ o funkcije gustoće vjerojatnosti.

PRIMJER 1

Emanuel se natječe u šahovskom turniru, gdje je njegova vjerojatnost pobjede jednaka 0,55. Nađite vjerojatnost da pobjeđuje četvrti put u svojoj sedmoj igri.

RIJEŠENJE

$$X \sim NB(4{,}0.55)$$

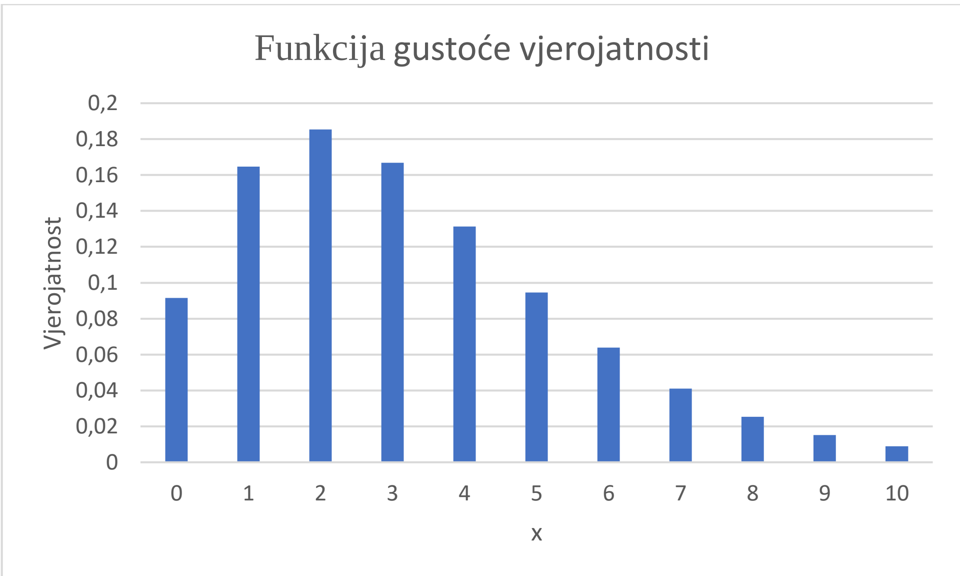
$$P(X = 7) = \binom{6}{3}(0{,}55)^4(0{,}45)^3 = 0{,}1668$$

U Excelu računamo:

$$= NEGBINOM.DIST(3{,}4{,}0{,}55, FALSE) = 0{,}1668$$

Vjerojatnost da pobjeđuje četvrti put u svojoj sedmoj igri jednaka je 0,1668, odnosno 16,68%,

Slika 13. je prikaz funkcije gustoće vjerojatnosti negativne binomne distribucije gdje je *r=4*, a *p=0,55*.



Slika 13. Funkcija gustoće vjerojatnosti negativne binomne distribucije gdje je r=4, a p=0,55

PRIMJER 2

Vjerojatnost da se rodi dječak je 0,5. Par želi imati djecu sve dok ne dobiju dvije djevojčice u obitelji.

a) Napišite funkciju gustoće vjerojatnosti slučajne varijable *X*, gdje *X* predstavlja broj dječaka prije nego što obitelj dobije dvije djevojčice.
b) Koja je vjerojatnost da obitelj ima četvero djece?
c) Koja je vjerojatnost da obitelj ima najviše četvero djece?
d) Koliki je očekivani broj dječaka koje će ova obitelj imati?
e) Koliki je očekivani broj djece koju će ova obitelj imati?

RIJEŠENJE

a)

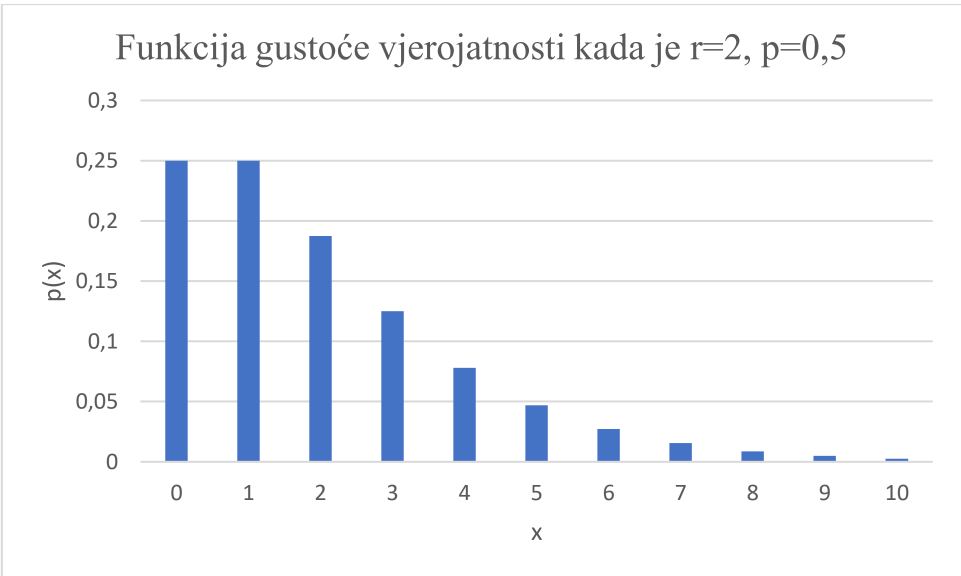
$$X \sim NB(2{,}0.5)$$

Funkcija gustoće raspodjele dana je izrazom:

$$P(X = x) = \binom{x+2-1}{x}(0{,}5)^2(0{,}5)^x \qquad x = 0{,}1{,}2{,}3, \dots$$

$$= \binom{x+1}{x}(0{,}5)^2(0{,}5)^x$$

Slika 14. prikazuje negativnu binomnu distribuciju gdje j r=2, a p=0,5.



Slika 14. Grafički prikaz funkcije gustoće vjerojatnosti negativne binomne distribucije gdje je r=2, a p=0,5

b) Ako obitelj ima četvero djece uz pretpostavku da imaju dvije djevojčice, zaključujemo da imaju i dva dječaka.

$$P(X = 2) = \binom{3}{2}(0{,}5)^2(0{,}5)^2 = 0{,}1875$$

Vjerojatnost da obitelj ima četvero djece jednaka je 0,1875, odnosno 18,75%.

c) Ako obitelj ima manje ili jednako od četvero djece, uz pretpostavku da imaju dvije djevojčice, mogu imati 0, 1 ili 2

$$P(X \leq 2) = \sum_{x=0}^{2} P(X = x)$$

$$= p(0) + p(1) + p(2) = 0{,}25 + 0{,}25 + 0{,}1875 = 0{,}6875$$

Vjerojatnost da obitelj ima najviše četvero djece jednaka je 0,6875, odnosno 68,75%.

d)

$$E(X) = \frac{r(1-p)}{p}$$

$$= \frac{2 * 0{,}5}{0{,}5} = 2$$

Očekivani broj dječaka koje će ova obitelj imati je 2.

e)

$$E(X+2) = E(X) + 2 = 2 + 2 = 4$$

Očekivani broj djece koje će ova obitelj imati je 4.

7. GEOMETRIJSKA DISTRIBUCIJA

7.1. DEFINICIJA

Geometrijska distribucija može se definirati kao diskretna distribucija vjerojatnosti koja predstavlja vjerojatnost broja uzastopnih neuspjeha prije nego što se postigne uspjeh u Bernoullijevom pokusu. Drugim riječima, u geometrijskoj distribuciji, Bernoullijev pokus se ponavlja dok se ne postigne uspjeh, a zatim se zaustavlja.

Geometrijska distribucija se temelji na tri važne pretpostavke.

- Ispitivanja koja se provode su neovisna.
- Postoje samo dva ishoda svakog mogućeg pokušaja- uspjeh ili neuspjeh.
- Vjerojatnost uspjeha, označena s p, ista je za svaki pokušaj, [7].

Funkcija gustoće vjerojatnosti geometrijske distribucije prikazana je izrazom:

$$P(X = k) = (1 - p)^{k-1}p,$$

gdje je k broj pokušaja $k = 1,2,3, \ldots$, a p je vjerojatnost uspjeha pokušaja.

Funkcija distribucije vjerojatnosti negativne binomne distribucije prikazana je izrazom:

$$F(x) = 1 - (1 - p)^{x+1}.$$

Očekivana vrijednost geometrijske distribucije slučajne varijable X:

$$E(X) = \frac{1}{p}.$$

Varijanca geometrijske distribucije slučajne varijable X:

$$V(X) = \frac{1 - p}{p^2}.$$

7.2. PRIMJERI

PRIMJER 1

Ako košarkaš ima 70% šanse da uspješno izvede slobodno bacanje, kolika je vjerojatnost da mu trebaju točno tri pokušaja da izvede svoje prvo uspješno slobodno bacanje?

RIJEŠENJE

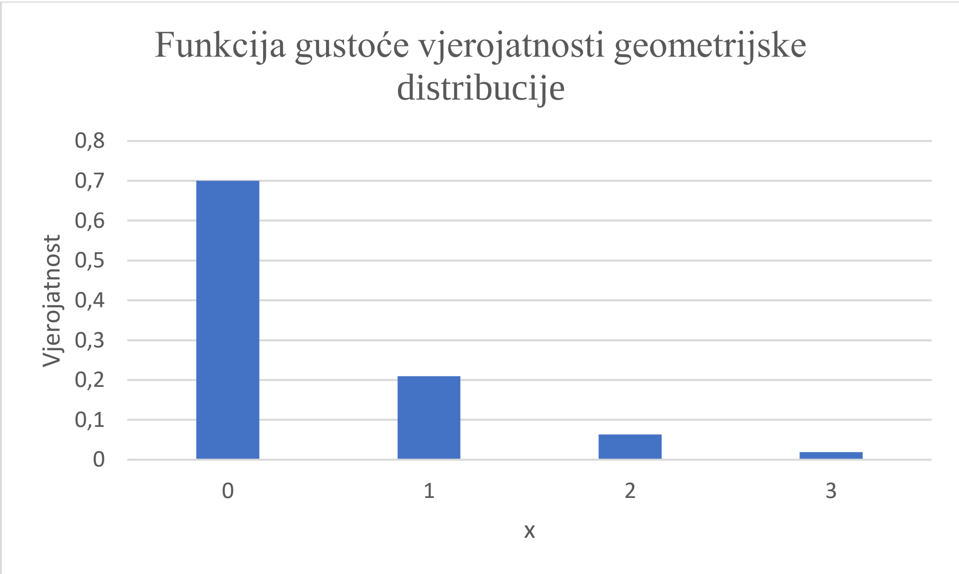
U primjeru vrijedi $p = 0{,}70, k = 3$. Koristimo formulu za funkciju gustoće vjerojatnosti geometrijske slučajne varijable.

$$P(X = k) = (1 - p)^{k-1}p$$

$$P(X = 3) = (1 - 0{,}70)^{3-1} * 0{,}70 = 0{,}0630$$

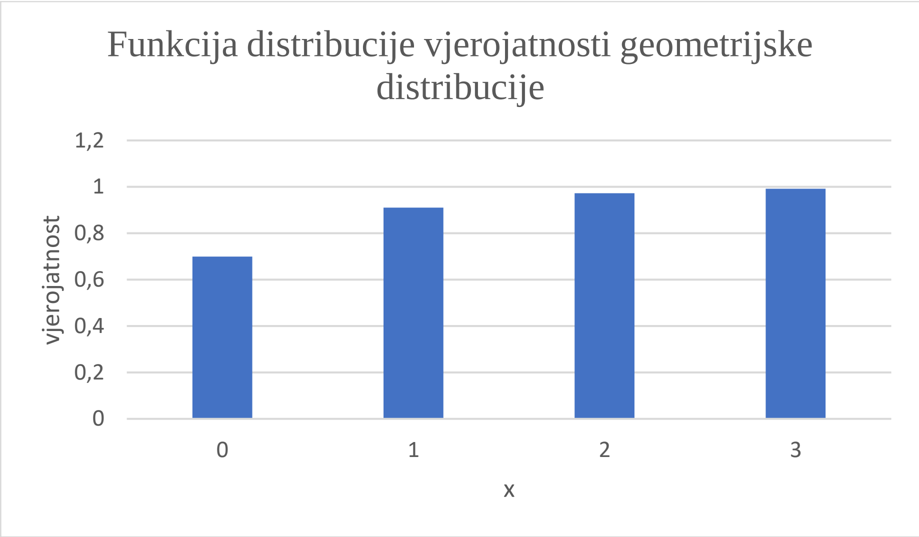
Vjerojatnost da će košarkašu trebati točno tri pokušaja da izvede svoje prvo uspješno slobodno bacanje jednaka je 0,0630, odnosno 6,30%.

Slika 15. prikazuje grafički prikaz funkcije gustoće vjerojatnosti geometrijske distribucije kada je *p=0,70*



Slika 15. Funkcija gustoće vjerojatnosti geometrijske distribucije kada je p=0,70

Slika 16. predstavlja grafički prikaz funkcije distribucije vjerojatnosti geometrijske distribucije kada je *p=0,70*



Slika 16. Funkcija distribucije vjerojatnosti geometrijske distribucije kada je p=0,7

PRIMJER 2

Softverska tvrtka objavljuje novu verziju svoje aplikacije. Kod čak 20% korisnika nadogradnja ne prođe bet problema. Koja je vjerojatnost da će peti korisnik koji pokuša izvršiti nadogradnju biti prvi koji je imao problema pri nadogradnji?

RIJEŠENJE

$$p = 0{,}2 \qquad k = 5$$

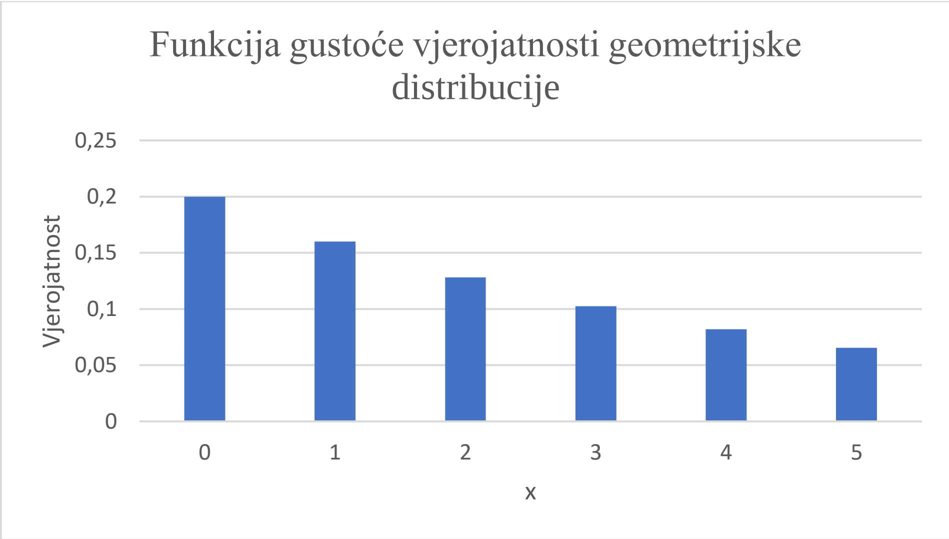
$$P(X = k) = (1 - p)^{k-1} p$$

$$P(X = 5) = (1 - 0{,}20)^{5-1} * 0{,}20$$

$$P(X = 5) = 0{,}0819$$

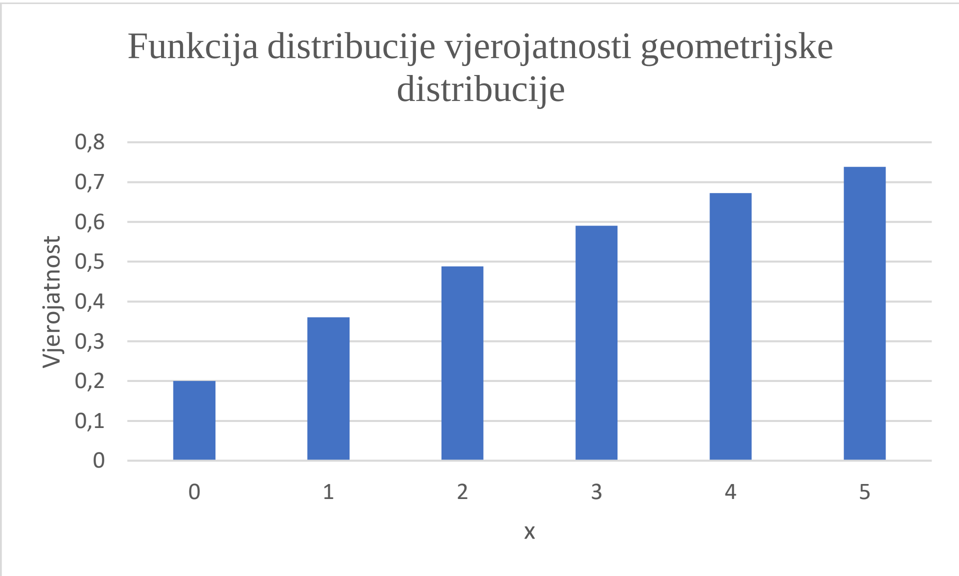
Vjerojatnost da će peti korisnik koji pokuša izvršiti nadogradnju biti prvi koji je imao problema pri nadogradnji jednaka je 0,0819, odnosno 8,19%.

Slika 17. je grafički prikaz funkcije gustoće vjerojatnosti geometrijske distribucije gdje je *p=0,2*



Slika 17. Funkcija gustoće vjerojatnosti geometrijske distribucije gdje je p=0,2

Slika 18. prikazuje funkciju distribucije vjerojatnosti geometrijske distribucije kada je *p=0,2*



Slika 18. Funkcija distribucije vjerojatnosti geometrijske distribucije kada je p=0,2

8. HIPERGEOMETRIJSKA DISTRIBUCIJA

8.1. DEFINICIJA

Hipergeometrijska distribucija je diskretna distribucija vjerojatnosti koja opisuje vjerojatnost da se postigne *k* uspjeha u *n* izvlačenja, bez vraćanja, iz konačne populacijske veličine *N* koja sadrži točno *K* objekata koji se smatraju uspjehom. Svako izvlačenje je ili uspjeh ili neuspjeh.

Hipergeometrijsku distribuciju karakteriziraju sljedeća svojstva.

- Rezultat svakog izvlačenja može se klasificirati u jednu od dvije međusobno isključive kategorije (zaposlen/nezaposlen, prošao/ nije prošao).
- Vjerojatnost uspjeha mijenja se svakim izvlačenjem, jer svako izvlačenje smanjuje populaciju, [8].

Oznaka: $X \sim Hipergeometrijska(N, K, n)$.

Slučajna varijabla slijedi hipergeometrijsku distribuciju ako njena funkcija gustoće vjerojatnosti glasi:

$$p_x k = \Pr(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}},$$

gdje je N veličina populacije, K je broj uspješnih stanja u populaciji, n je broj izvlačenja, k je broj opaženih uspjeha i $\binom{a}{b}$ binomni koeficjent.

Očekivanje je prikazano izrazom:

$$E(X) = \frac{nk}{N}.$$

Varijanca je prikazana izrazom:

$$VAR(X) = \frac{nk(n-k)*(N-n)}{N^2(N-1)}.$$

8.2. PRIMJERI

Primjeri su riješavani u Microsof Excelu pomoću funkcije

$$HYPGEOM.DIST(k, n, K, N, FALSE)$$

Gdje se *FALSE* koristi za funkciju gustoće vjerojatnosti, odnosno

$$HYPGEOM.DIST(k, n, K, N, TRUE)$$

Za funkciju distribucije vjerojatnosti.

PRIMJER 1

Proizvodni pogon proizvodi seriju od 30 elektroničkih komponenti, od kojih se očekuje da će 10% biti neispravno. Inspektor odabire 8 komponenti za testiranje. Kolika je vjerojatnost da su 2 od 8 komponenti neispravne?

RIJEŠENJE

Ukupni broj komponenti $N = 30$. Promatramo 8 komponenti pa je $n = 8$. S obzirom da nas zanimaju dvije od osam komponenti računamo

$$\mathrm{P}(X) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

$$\mathrm{P}(2) = \frac{\binom{3}{2}\binom{30-3}{8-2}}{\binom{30}{8}}$$

$$\mathrm{P}(2) = \frac{\binom{3}{2}\binom{27}{6}}{\binom{30}{8}}$$

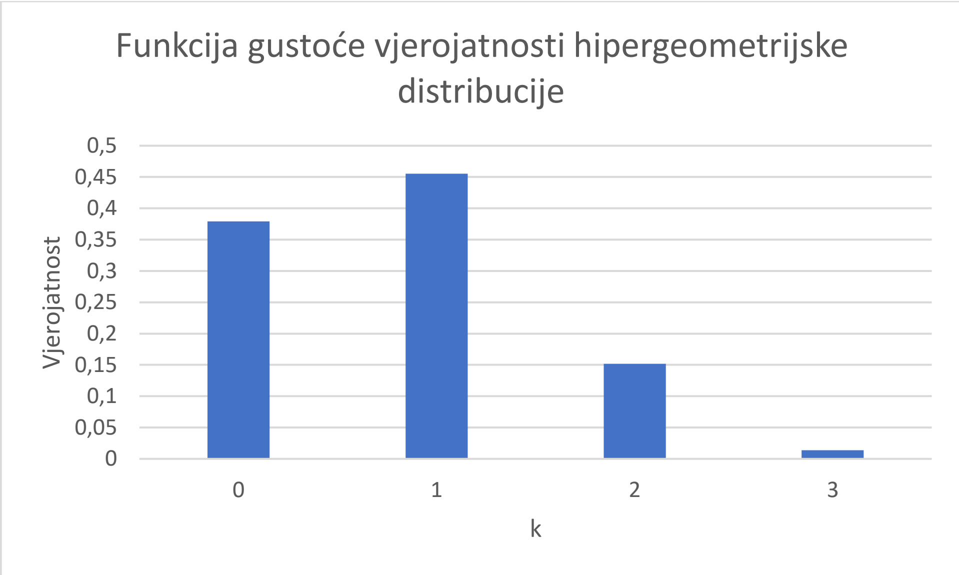
$$= \frac{3 * 296010}{5852925} = 0{,}1517$$

U Excelu računamo:

$$= HYPGEOM.DIST(2{,}8{,}3{,}30, FALSE) = 0{,}1517$$

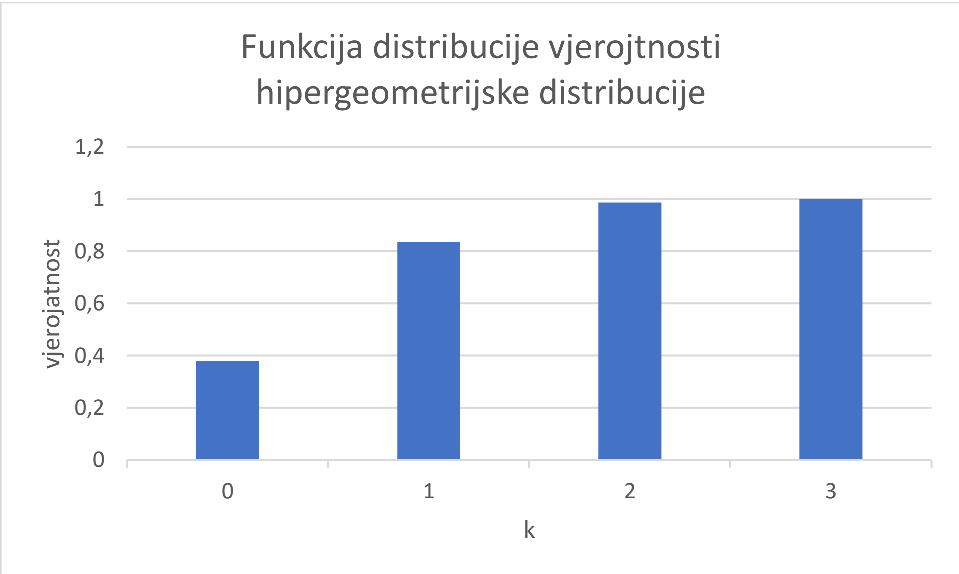
Vjerojatnost da su dvije od osam komponenata neispravne jednaka je 0,1517, odnosno 15,17%.

Slika 19. predstavlja grafički prikaz funkcije gustoće vjerojatnosti hipergeometrijske distribucije kada je *k=2, n=8, K=3 i N=30.*



Slika 19. Funkcija gustoće vjerojatnosti hipergeometrijske distribucije kada je k=2,n=8, K=3 i N=30

Slika 20. predstavlja grafički prikaz funkcije distribucije vjerojatnosti hipergeometrijske distribucije kada je *k=2, n=8, K=3 i N=30.*



Slika 20. Funkcija distribucije vjerojatnosti ipergeometrijske distribucije kada je k=2,n=8,K=3 i N=30

9. ZAKLJUČAK

U ovom završnom radu istražene su diskretne slučajne varijable i njihove primjene u različitim područjima, s velikim naglaskom na njihovu široku primjenu u statistici i inženjerstvu. Analizom različitih distribucija, kao što su Bernoullijeva, binomna, Poissonova, negativna binomna geometrijska te hipergeometrijska distribucija dobili smo uvid kako se te varijable ponašaju te kako njihove teorijske fenomene možemo primijeniti u stvarnim situacijama.

Bernoullijeva distribucija služi kao matematički model za eksperimente, događaje i procese koji imaju točno dva različita ishoda. Binomna distribucija predstavlja slučajan pokus, koji se sastoji od n manjih pokusa istog tipa, gdje rezultat svakog pokusa može biti uspjeh ili neuspjeh. Možemo reći da su to Bernoullijevi pokusi koji se ponavljaju n puta. Pogodna je za primjenu u analizi kvalitete kada želimo provjeriti određeni broj neispravnih stavki u seriji proizvoda ili u medicinskim istraživanjima za procjenu vjerojatnosti broja pacijenata koji će reagirati na određeni tretman. Poissonova distribucija se koristi za modeliranje broja događaja u fiksnom vremenskom ili prostornom intervalu kada su ti događaji rijetki i nezavisni. Pogodna je za primjenu u epidemiologiji, astronomiji, geologiji te za procjenu defekata u proizvodnoj liniji ili za procjenu kvarova na uređajima. Negativna binomna distribucija pogodna je za modeliranje broja neuspjeha koju su potrebni za postizanje određenog broja uspjeha u seriji. Može se koristiti za istraživanje kvalitete u proizvodnji, za klinička ispitivanja i marketing kompanije. Geometrijska distribucija se koristi u situacijama kada želimo saznati koliko pokušaja (neuspjeha) je potrebno do prvog uspjeha. Možemo ju koristiti za analizu kvalitete kada želimo znati koliko dijelova trebamo pregledati dok ne naiđemo na prvi neispravan. Pogodna je i za primjenu u korisničkim službama te testiranju pouzdanosti. Hipergeometrijska distribucija pogodna je za analizu u situacijama u kojima imamo uzorak iz ograničene populacije, bez vraćanja uzorka. Može se koristiti u epidemiologiji, genetici te za uzorkovanje anketa.

Praktična primjena distribucija prikazana je grafičkim prikazima i konkretnim primjerima, što je omogućilo bolje razumijevanje i primjenu.

Zaključno, razumijevanje i primjena diskretnih slučajnih varijabli i njihovih distribucija omogućuje rješavanje složenih problema. Proučavanje diskretnih slučajnih varijabli ne samo da poboljšava našu sposobnost predviđanja i analize ishoda, već podupire i složenije statističke metode i procese donošenja odluka. Daljnja istraživanja i razvoj ove oblasti mogu donijeti nove uvide i poboljšati metodologije u različitim znanstvenim i industrijskim disciplinama.

10.LITERATURA

[1] E. Begović Kovač, Numeričke i statističke metode, predavanja, Sveučilište u Zagrebu, Fakultet kemijskog inženjerstva i tehnologije

[2] J.L. Devore, Probability and Statistics or Engineering and the Sciences, Cengage Learning, 9th edition, 2014.

[3] https://nzmaths.co.nz/category/glossary/standard-deviation-discrete-random-variable

[4] https://en.wikipedia.org/wiki/Bernoulli_distribution

[5] https://www.geeksforgeeks.org/poisson-distribution/

[6] https://en.wikipedia.org/wiki/Negative_binomial_distribution

[7] https://en.wikipedia.org/wiki/Geometric_distribution

[8] https://en.wikipedia.org/wiki/Hypergeometric_distribution