

Parametarski zadane diskretne distribucije

Stehlik, Petra

Undergraduate thesis / Završni rad

2017

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Department of Mathematics / Sveučilište Josipa Jurja Strossmayera u Osijeku, Odjel za matematiku**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:126:688589>

Rights / Prava: [In copyright](#)

Download date / Datum preuzimanja: **2022-12-02**



Repository / Repozitorij:

[Repository of Department of Mathematics Osijek](#)



Sveučilište J. J. Strossmayera u Osijeku
Odjel za matematiku

Petra Stehlik

Parametarski zadane diskretne distribucije

Završni rad

Osijek, 2017.

Sveučilište J. J. Strossmayera u Osijeku
Odjel za matematiku

Petra Stehlik

Parametarski zadane diskretne distribucije

Završni rad

Voditelj: dr. sc. Slobodan Jelić

Osijek, 2017.

Sažetak

U ovom radu obrađuju se diskretne slučajne varijable, odnosno tipovi distribucije diskretnih slučajnih varijabli. U prvom dijelu rada će se navesti definicije osnovnih pojmova te glavni rezultati vezani za diskretne slučajne varijable. Definirat će se i numeričke karakteristike —očekivanje i varijanca, te navesti njihova svojstva. Zatim će se u drugom dijelu rada definirati i detaljno opisati pojedine distribucije te izračutati ili navesti njihova očekivanja i varijace. Konkretno, distribucije o kojima će se govoriti su: diskretna uniformna, Bernoullijeva, binomna, multinomna, Poissonova, geometrijska, negativna binomna, hipergeometrijska i negativna hipergeometrijska distribucija.

Ključne riječi

diskretna slučajna varijabla, distribucija, očekivanje, varijanca, diskretna uniformna distribucija, Bernoullijeva distribucija, binomna distribucija, multinomna distribucija, Poissonova distribucija, geometrijska distribucija, negativna binomna distribucija, hipergeometrijska distribucija, negativna hipergeometrijska distribucija

Abstract

In this paper, discrete random variables ie. types of distribution of discrete random variables are being studied. First chapter includes basic definitions and results related to discrete random variable. Also, numerical characteristics —expected value and variance, will be defined and their properties stated. Then, in chapter two, each of distributions will be defined, described in detail and their expected values and variances will be calculated. Namely, distributions that will be studied are: discrete uniform, Bernoulli, binomial, multinomial, Poisson, geometric, negative binomial, hypergeometric and negative hypergeometric distribution.

Key words

discrete random variable, distribution, expected value, variance, discrete uniform distribution, Bernoulli distribution, binomial distribution, multinomial distribution, Poisson distribution, geometric distribution, negative binomial distribution, hypergeometric distribution, negative hypergeometric distribution

Sadržaj

Uvod	1
1 Osnovne definicije	2
2 Parametarski zadane diskretne distribucije	6
2.1 Diskretna uniformna distribucija	6
2.2 Bernoullijeva distribucija	7
2.3 Binomna distribucija	8
2.4 Multinomna distribucija	11
2.5 Poissonova distribucija	12
2.6 Geometrijska distribucija	15
2.7 Negativna binomna distribucija	16
2.8 Hipergeometrijska distribucija	17
2.9 Negativna hipergeometrijska distribucija	19
Literatura	21

Uvod

U ovome radu se smatra da su usvojeni osnovni pojmovi teorije vjerojatnosti te da se znaju izračunati vjerojatnosti ishoda osnovnih pokusa.

Osnova ovog rada su modeli diskretnih slučajnih varijabli koji su nastali klasifikacijom raznih pokusa. Naime, ako promotrimo primjer bacanja novčića, zatim bacanja kockice pa čak i bacanja pravilnog poliedra sa numeriranim stranama, uočiti ćemo da kod pojedinog pokusa svi (međusobno različiti) ishodi imaju jednaku vjerojatnost realizacije. Za novčić smatramo da je simetričan, što znači da za npr. pismo ne postoji nikakav razlog niti karakteristika koja bi utjecala na ishod i time povećala (ili smanjila) vjerojatnost njegove realizacije. Analogno vrijedi za bacanje kockice ili pravilnog poliedra. Jedino po čemu se pokusi razlikuju je "broj strana", što nas navodi na kreiranje univerzalnog modela diskretnih slučajnih varijabli - diskretnu uniformnu distribuciju. Slično bi se za ostale distribucije mogli osmisliti motivacijski primjeri.

Također, ponekad se ne radi o samo jednom izvođenju pokusa, nego o nezavisnom ponavljanju istog pokusa, kao što je slučaj kod binomne distribucije pa je od interesa brojanje uspjeha u n izvođenja Bernoullijevog pokusa. Binomna distribucija služi i za aproksimaciju hipergeometrijske distribucije, ako su zadovoljeni određeni uvjeti, te je također njen granični slučaj upravo Poissonova distribucija, što će biti i pokazano.

U prvom dijelu ovog rada definirat će se osnovni pojmovi koji će se koristiti. Navest će se i korisni teoremi i tvrdnje koji će kasnije biti iskorišteni. Također će se definirati i numeričke karakteristike koje služe pri opisivanju slučajnih varijabli.

Zajedničko svim modelima je i to da imaju određene parametre koji nešto govore o samoj distribuciji ili je u potpunosti određuju.

U drugom dijelu bit će detaljno opisane sljedeće distribucije: diskretna uniformna, Bernoullijeva, binomna, multinomna, Poissonova, geometrijska, negativna binomna, hipergeometrijska i negativna hipergeometrijska redom, te će biti popraćene odgovarajućim primjerima i zadacima.

Dodatno će se i opisati i izvesti veza između binomne i Poissonove, te binomne i hipergeometrijske distribucije. Na kraju glavnog dijela će se spomenuti sličnosti i razlike između četiri distribucije, pregledno prikazane u tablici.

Poglavlje 1

Osnovne definicije

Slijede, za ovaj rad, ključne definicije i rezultati koji se koriste pri definiranju modela slučajnih varijabli. Počinje općenitom definicijom slučajne varijable, a nakon definiranja diskretne slučajne varijable sve naredne definicije i teoremi vezani su isključivo za diskretnu slučajnu varijablu.

Definicija 1.1. (Slučajna varijabla)

Neka je (Ω, \mathcal{F}, P) vjerojatnosni prostor i $\mathcal{B}_{\mathbb{R}}$ σ -algebra na \mathbb{R} koja sadrži otvorene podskupove od \mathbb{R} i nazivamo je Borelova σ -algebra. Svaka funkcija $X : \Omega \rightarrow \mathbb{R}$ takva da je skup $\{\omega \in \Omega : X(\omega) \in B\}$ događaj iz \mathcal{F} za svaki $B \in \mathcal{B}_{\mathbb{R}}$, to jest $X^{-1}(B) \in \mathcal{F}$ za svaki $B \in \mathcal{B}_{\mathbb{R}}$ nazivamo slučajna varijabla.

Definicija 1.2. (Slika slučajne varijable)

Skup svih vrijednosti koje može poprimiti slučajna varijabla naziva se slika slučajne varijable i označava s $\mathcal{R}(X)$.

Definicija 1.3. (Diskretna slučajna varijabla)

Slučajna varijabla $X : \Omega \rightarrow \mathbb{R}$ na vjerojatnosnom prostoru (Ω, \mathcal{F}, P) je diskretna ako postoji diskretni skup $D \subset \mathbb{R}$ tako da je $P(X \in D) = 1$, tj. ako joj je slika diskretni skup.

Na diskretnom vjerojatnosnom prostoru $(\Omega, \mathcal{P}(\Omega), P)$, (gdje je $\mathcal{P}(\Omega)$ partitivni skup skupa Ω), svaka funkcija $X : \Omega \rightarrow \mathbb{R}$ je slučajna varijabla i to diskretna.

Osim diskretne, u teoriji vjerojatnosti se još obrađuje i neprekidna slučajna varijabla. Kod te vrste slučajne varijable, slika nije diskretni skup, nego nekakav interval realnih brojeva ili cijeli skup \mathbb{R} . Vjerojatnost da ova varijabla poprima neku vrijednost se izračunava integriranjem njene funkcije gustoće kojom je neprekidna slučajna varijabla u potpunosti zadana, dok je diskretna slučajna varijabla u potpunosti zadana svojom tablicom distribucije čija definicija slijedi.

Definicija 1.4. (Tablica distribucije)

Diskretnu slučajnu varijablu $X : \Omega \rightarrow \mathbb{R}$ na vjerojatnosnom prostoru $(\Omega, \mathcal{P}(\Omega), P)$ kojoj je skup vrijednosti $\mathcal{R}(X) = \{x_i, i \in \mathbb{N}\}$ i pripadni niz vjerojatnosti $(p_i, i \in \mathbb{N})$ takav da vrijedi $p_i = P(X = x_i)$, prikazujemo u obliku tablice

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

i taj prikaz zovemo tablica distribucije, odnosno distribucija.

Pri tome mora biti zadovoljeno sljedeće:

- 1) $x_i \neq x_j \quad \forall i \neq j$
- 2) $0 \leq p_i = P(X = x_i) \leq 1$
- 3) $\sum_{i \in \mathbb{N}} p_i = 1$

U slučaju diskretne slučajne varijable, sve informacije potrebene za opis slučajne varijable nalaze se u njenoj tablici distribucije. Zato se kaže da je diskretna slučajna varijabla u potpunosti zadana svojom tablicom distribucije.

Također, uvode se neki karakteristični brojevi koji daju dodatne informacije o slučajnim varijablama, odnosno distribucije čine prepoznatljivima. Nazivaju se numeričkim karakteristikama i koriste se uglavnom matematičko očekivanje i varijanca, čije definicije slijede.

Definicija 1.5. (Matematičko očekivanje)

Neka je $(\Omega, \mathcal{P}(\Omega), P)$ diskretan vjerojatnosni prostor i X slučajna varijabla na njemu. Ako red $\sum_{\omega \in \Omega} X(\omega)P(\omega)$ apsolutno konvergira, onda kažemo da slučajna varijabla X ima matematičko očekivanje i broj

$$EX = \sum_{\omega \in \Omega} X(\omega)P(\omega)$$

zovemo matematičko očekivanje, odnosno očekivanje slučajne varijable X .

Teorem 1.1. Neka je $(\Omega, \mathcal{P}(\Omega), P)$ diskretan vjerojatnosni prostor i

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

slučajna varijabla na njemu. Redovi $\sum_{\omega \in \Omega} X(\omega)P(\omega)$ i $\sum_{i \in \mathbb{N}} x_i p_i$ istovremeno ili apsolutno konvergiraju ili apsolutno divergiraju i u slučaju apsolutne konvergencije sume su im jednake, odnosno, vrijedi

$$EX = \sum_{\omega \in \Omega} X(\omega)P(\omega) = \sum_{i \in \mathbb{N}} x_i p_i.$$

Teorem 1.2. Neka je $(\Omega, \mathcal{P}(\Omega), P)$ diskretan vjerojatnosni prostor,

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

slučajna varijabla na njemu i $g : \mathcal{R}(X) \rightarrow \mathbb{R}$ funkcija takva da postoji očekivanje $Eg(X)$. Tada vrijedi

$$Eg(X) = \sum_{i \in \mathbb{N}} g(x_i) p_i.$$

Matematičko očekivanje ima sljedeća važna svojstva:

- 1) $E(aX + b) = aEX + b$, $a, b \in \mathbb{R}$
- 2) Ako je $X(\omega) \geq 0, \forall \omega \in \Omega$, onda je $EX \geq 0$
- 3) Neka su X i Y slučajne varijable na istom vjerojatnosnom prostoru $(\Omega, \mathcal{P}(\Omega), P)$ takve da je $X(\omega) \leq Y(\omega), \forall \omega \in \Omega$. Tada je $EX \leq EY$.

Linearnost matematičkog očekivanja dana je teoremom:

Teorem 1.3. Neka su X i Y slučajne varijable na istom vjerojatnosnom prostoru $(\Omega, \mathcal{P}(\Omega), P)$ takve da postoje EX i EY . Tada za proizvoljne $a, b \in \mathbb{R}$ slučajna varijabla $aX + bY$ također ima matematičko očekivanje i vrijedi $E(aX + bY) = aEX + bEY$.

Definicija 1.6. (Varijanca)

Ako postoji broj $E(X - EX)^2$, onda taj pozitivan realan broj nazivamo varijanca slučajne varijable X i označavamo sa $VarX$.

Iako ovdje nisu definirani, u terminima r -tog momenta i r -tog centralnog momenta slučajne varijable, varijanca je drugi centralni moment od X ; odnosno, riječima, varijanca je očekivano kvadratno odstupanje slučajne varijable od njenog očekivanja.

Uz definiciju varijance obično se navode dva svojstva koja se često koriste:

- 1) $VarX = EX^2 - (EX)^2$
- 2) $Var(aX + b) = a^2VarX$,

a lako slijede iz definicije varijance i nekih svojstava matematičkog očekivanja.

Također koristi se pojam standardne devijacije σ_X , koja označava korijen iz varijance, tj. $\sigma_X = \sqrt{VarX}$.

Definicija 1.7. (Nezavisnost slučajnih varijabli)

Diskretne slučajne varijable X i Y su nezavisne ako su događaji $\{X = x\}$ i $\{Y = y\}$ nezavisni za svaki x i y .

Lema 1.1. *Ako su X i Y nezavisne, onda je $E(XY) = EX \cdot EY$.*

Definicija 1.8. *(Nekorelirane slučajne varijable)*

Za X i Y se kaže da su nekorelirane ako $E(XY) = EX \cdot EY$.

Teorem 1.4. *Za slučajne varijable X i Y vrijedi:*

(a) $Var(aX) = a^2VarX$, za $a \in \mathbb{R}$

(b) $Var(X + Y) = VarX + VarY$, ako su X i Y nekorelirane.

Poglavlje 2

Parametarski zadane diskretne distribucije

Već je spomenuto da su diskretne slučajne varijable u potpunosti određene svojom tablicom distribucije, te da se pri opisivanju slučajne varijable mogu koristiti numeričke karakteristike specifične za tu slučajnu varijablu odnosno njenu distribuciju.

Ovdje će se navesti devet parametarski zadanih diskretnih distribucija i objasniti kako se koja koristi i za kakve probleme.

2.1 Diskretna uniformna distribucija

Definicija 2.1. Za slučajnu varijablu X kažemo da ima diskretnu uniformnu distribuciju ako joj je slika $\mathcal{R}(X) = \{x_1, x_2, \dots, x_n\} \subset \mathbb{R}$ i vjerojatnosti realizacija te slučajne varijable su međusobno jednake,

$$P(X = x_i) = \frac{1}{n}.$$

Tablica distribucije:

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}$$

Pri izračunavanju očekivanja (koristeći Teorem 1.1.) dobije se:

$$EX = \sum_{i=1}^n x_i p_i = \sum_{i=1}^n x_i \frac{1}{n} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n,$$

gdje \bar{x}_n zapravo označava aritmetičku sredinu elemenata slike.

Varijancu izračunamo koristeći Definiciju 1.6.:

$$Var X = E(X - EX)^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2 \frac{1}{n} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

Ovakav tip distribucije se koristi za modelirane slučajnih varijabli pridruženih pokusima čiji su ishodi međusobno jednako vjerojatni. To su npr.:

– bacanje simetričnog novčića, s vjerojatnosti pojedinog ishoda $\frac{1}{2}$

– bacanje simetrične kockice, s vjerojatnosti pojedinog ishoda $\frac{1}{6}$

– izvlačenje jedne numerirane kuglice iz kutije sa n numeriranih kuglica, s vjerojatnosti pojedinog ishoda $\frac{1}{n}$, i slično.

2.2 Bernoullijeva distribucija

Slučajna varijabla koja može poprimiti točno dvije vrijednosti, često uspjeh – 1 ili neuspjeh – 0, modelira se Bernoullijevom distribucijom. Njen jedini parametar je vjerojatnost uspjeha što se obično označava sa p , te je očito da vrijedi $q = 1 - p$, gdje je q vjerojatnost neuspjeha.

Tablica ovakve distribucije jednostavnog je oblika: $X = \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$

Lako se dobije da je očekivanje $EX = p$.

Varijanca se dobije kad se iskoristi svojstvo $Var X = EX^2 - (EX)^2$ i činjenica da su X i transformirana¹ X^2 jednako distribuirane² pa je

$$Var X = EX^2 - (EX)^2 = p - p^2 = p(1 - p) = pq.$$

Primjer 2.1. *Na stolu se nalazi kutija sa 15 čokoladi. 7 je bijelih, a 8 tamnih čokoladi. U jednom izvlačenju jedne čokolade, uspjehom se smatra izvučena tamna čokolada. Slučajna varijabla pridružena ovakvom pokusu modelirana je Bernoullijevom distribucijom kojoj je pridružena sljedeća tablica:*

$$X = \begin{pmatrix} 1 & 0 \\ \frac{8}{15} & \frac{7}{15} \end{pmatrix}.$$

Obično se pokus sa dva moguća ishoda naziva Bernoullijev pokus i bitan je za sljedeće modele slučajnih varijabli.

¹Po definiciji transformacije diskretne slučajne varijable, X^2 je kompozicija $g \circ X$, gdje je $g(x) = x^2$ funkcija, pa će slika od X biti jednaka slici od X^2 , kao i pripadne vjerojatnosti.

²Po definiciji, diskretne slučajne varijabe su jednako distribuirane ako imaju jednake tablice distribucije.

2.3 Binomna distribucija

Kada bi se n puta nezavisno ponovio Bernoullijev pokus, s vjerojatnošću uspjeha p , tada bi se prirodno nametnulo pitanje kolika je vjerojatnost da se uspjeh realizirao točno k puta? Slučajna varijabla koja tako broji uspjehe modelira se binomnom distribucijom, definiranom na sljedeći način:

Definicija 2.2. *Neka je $p \in (0, 1)$, $n \in \mathbb{N}$. Za slučajnu varijablu X koja poprima vrijednosti iz skupa $\{0, 1, \dots, n\}$ s vjerojatnostima*

$$p_i = P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, \quad i \in \{1, \dots, n\},$$

kažemo da ima binomnu distribuciju s parametrima n i p i pišemo $X \sim \mathcal{B}(n, p)$.

Značenje parametara:

n – broj nezavisnih ponavljanja Bernoullijevog pokusa

p – vjerojatnost uspjeha u pojedinoj izvedbi pokusa

Dakle, vjerojatnost da se uspjeh realizirao točno k puta jednaka je

$$\binom{n}{k} p^k (1-p)^{n-k}.$$

Očekivanje:

$$EX = \sum_{k=0}^n k p_k = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k}$$

Varijanca (Teorem 1.2.):

$$Var X = EX^2 - (EX)^2, \quad EX^2 = \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k}$$

Da bismo izračunali matematičko očekivanje i varijancu, uočimo sljedeće:

$$g(t) = (at + b)^n \stackrel{(**)}{=} \sum_{k=0}^n \binom{n}{k} a^k t^k b^{n-k}, \quad a, b \in \mathbb{R} \quad (1)$$

(**) korišten je binomni teorem

Deriviranjem funkcije g , odnosno izraza (1), dobije se, s jedne strane

$$g'(t) = n(at + b)^{n-1}a,$$

a s druge strane

$$g'(t) = \sum_{k=0}^n \binom{n}{k} k a^k t^{k-1} b^{n-k},$$

pa izjednačavanjem strana i uvrštavanjem $t = 1$ slijedi

$$an(a + b)^{n-1} = \sum_{k=0}^n k \binom{n}{k} a^k b^{n-k}. \quad (2)$$

Analogno, ponovnim deriviranjem i uvrštavanjem $t = 1$, dobijemo

$$a^2n(n - 1)(a + b)^{n-2} = \sum_{k=0}^n k(k - 1) \binom{n}{k} a^k b^{n-k}. \quad (3)$$

Sada, ako se uzme $a = p$ i $b = 1 - p$, pa iskoristi (2), dobije se da je očekivanje

$$EX = \sum_{k=0}^n k \binom{n}{k} p^k (1 - p)^{n-k} = pn(p + 1 - p)^{n-1} = pn.$$

Zatim se iskoristi (3) i dobije da je:

$$\begin{aligned} EX^2 &= \sum_{k=0}^n k^2 \binom{n}{k} p^k (1 - p)^{n-k} = \sum_{k=0}^n (k^2 - k + k) \binom{n}{k} p^k (1 - p)^{n-k} \\ &= \sum_{k=0}^n k(k - 1) \binom{n}{k} p^k (1 - p)^{n-k} + \underbrace{\sum_{k=0}^n k \binom{n}{k} p^k (1 - p)^{n-k}}_{=EX=np} \\ &= p^2n(n - 1)(p + 1 - p)^{n-2} + pn = p^2n(n - 1) + pn \\ &= p^2n^2 - p^2n + pn, \end{aligned}$$

pa se lako izračuna varijanca:

$$VarX = p^2n^2 - p^2n + pn - (pn)^2 = pn(1 - p).$$

Zadatak 2.1. Student je na pismenom ispitu iz nekog kolegija ostvario 60 bodova. Na usmenom dijelu ispita mora odgovoriti na 5 izvučenih pitanja pri čemu je vjerojatnost da je izvukao pitanje na koje zna odgovor jednaka $\frac{2}{3}$, te mu svaki točan odgovor donosi 20 bodova koji se pribrajaju bodovima s pismenog ispita. Zanima nas distribucija slučajne varijable koja broji ukupan broj bodova koje je student ostvario na usmenom i pismenom ispitu zajedno. Koliki je očekivani broj bodova koje će student ostvariti?

Rješenje:

Bernoullijev pokus se ovdje sastoji od izvlačenja pitanja i mogući ishodi su

0 —student ne zna odgovor na izvučeno pitanje

1 —student zna odgovor na izvučeno pitanje

pri čemu je $p(0) = \frac{1}{3}$, a $p(1) = \frac{2}{3}$.

Sa X označimo slučajnu varijablu koja broji izvučena pitanja na koja je student znao odgovoriti: $X \sim \mathcal{B}(5, \frac{2}{3})$, slika joj je $\mathcal{R}(X) = \{0, 1, 2, 3, 4, 5\}$, a pripadne vjerojatnosti

$$p_i = \binom{5}{k} \left(\frac{2}{3}\right)^k \left(\frac{1}{3}\right)^{5-k}.$$

Pripadna tablica distribucije:

$$X = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ \frac{1}{243} & \frac{10}{243} & \frac{40}{243} & \frac{80}{243} & \frac{80}{243} & \frac{32}{243} \end{pmatrix}$$

Pripadno očekivanje:

$$EX = n \cdot p = 5 \cdot \frac{2}{3} = \frac{10}{3}$$

Neka je sada Y slučajna varijabla koja broji ukupne bodove: $Y = 20X + 60$, slika joj je $\mathcal{R}(Y) = \{60, 80, 100, 120, 140, 160\}$, a pripadna tablica distribucije:

$$Y = \begin{pmatrix} 60 & 80 & 100 & 120 & 140 & 160 \\ \frac{1}{243} & \frac{10}{243} & \frac{40}{243} & \frac{80}{243} & \frac{80}{243} & \frac{32}{243} \end{pmatrix}$$

Pripadno očekivanje, koje se traži u zadatku:

$$EY = E(20X + 60) = 20EX + 60 = 126.67,$$

što znači da se očekuje da će student ostvariti ukupno oko 126 bodova.

2.4 Multinomna distribucija

Binomnom distribucijom se modeliralo brojanje uspjeha u n nezavisnih ponavljanja Bernoullijevog pokusa koji ima dva moguća ishoda s vjerojatnostima p i $1 - p$. Kada bi se n puta nezavisno ponovio pokus koji ima tri moguća ishoda s pripadnim vjerojatnostima p , q i $1 - p - q$, tada bi se vjerojatnost da se prvi ishod dogodio r puta, drugi w puta i treći $n - r - w$ puta, računala na sljedeći način:

$$p(r, w, n - r - w) = \frac{n!}{r! w! (n - r - w)!} \cdot p^r q^w (1 - p - q)^{n - r - w}.$$

Slučajna varijabla koja broji takve uspjehe se modelira takozvanom *trinomnom distribucijom*.

Analogno, kada bi se radilo o n nezavisnih ponavljanja pokusa koji ima konačno mnogo (k) ishoda, imali bi:

$$p(n_1, \dots, n_k) = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \cdot \dots \cdot p_k^{n_k}.$$

Dakle, $p(n_1, \dots, n_k)$ je vjerojatnost da se i -ti ishod dogodio n_i puta.

k – broj ishoda pojedinog pokusa

p_i – vjerojatnost i -tog ishoda u pojedinom pokusu

n_i – koliko puta se pojavio i -ti ishod, pri čemu vrijedi $\sum_{i=1}^k n_i = n$.

Analogno se pripadajuća distribucija naziva *multinomna distribucija*.

Primjer 2.2. *Nezavisno bacimo 4 simetrične kocke. Kolika je vjerojatnost da padnu točno dvije jedinice i točno jedna dvojka?*

Rješenje:

Ovdje nezavisno bacamo istu kocku 4 puta odnosno bacimo 4 kocke odjednom, svejedno je. Od interesa su nam jedinice i dvojke, pri čemu se jedinica pojavi dva puta, $n_1 = 2$, a dvojka jednom, $n_2 = 1$. Kako imamo 4 kocke, $n = 4$, a $n_1 + n_2 = 3$, pa se na jednoj preostaloj kocki može pojaviti bilo što osim jedinice i dvojke i $n_3 = 1$. Imamo:

$$p(1) = \frac{1}{6}, n_1 = 2$$

$$p(2) = \frac{1}{6}, n_2 = 1$$

$$p(3, 4, 5 \text{ ili } 6) = \frac{2}{3}, n_3 = 1$$

Dakle, ukupno imamo $k = 3$ ishoda. Vjerojatnost događaja koji nas zanima je:

$$p(n_1, n_2, n_3) = p(2, 1, 1) = \frac{4!}{2! 1! 1!} \left(\frac{1}{6}\right) \left(\frac{1}{6}\right) \left(\frac{2}{3}\right) = \frac{1}{27}.$$

2.5 Poissonova distribucija

Poissonovom distribucijom modeliramo slučajnu varijablu koja broji uspjehe u nekom jediničnom intervalu vremena, površine, mase i sl. Pri tome pokus mora zadovoljiti sljedeća tri uvjeta:

- 1) Vjerojatnost pojavljivanja uspjeha ne ovisi o tome u kojem će se jediničnom intervalu on dogoditi.
- 2) Broj uspjeha u jednom, neovisan je o broju uspjeha u bilo kojem drugom jediničnom intervalu.
- 3) Očekivani broj uspjeha isti je za sve jedinične intervale i dan je pozitivnim realnim brojem λ .

Definicija 2.3. Slučajna varijabla X ima Poissonovu distribuciju s parametrom $\lambda > 0$ ako prima vrijednosti iz skupa $\{0, 1, 2, \dots\} = \mathbb{N}_0$ s pripadnim vjerojatnostima

$$p_i = P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}, \quad i \in \mathbb{N}_0.$$

Pišemo $X \sim \mathcal{P}(\lambda)$.

Očekivanje:

$$EX = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}}_{e^{\lambda}} = \lambda$$

Varijanca:

Kako je $Var X = EX^2 - (EX)^2$, a zbog linearnosti očekivanja je

$$EX^2 = E(X^2 - X) + EX,$$

te vrijedi (Teorem 1.2.):

$$E(X^2 - X) = E(X(X-1)) = \sum_{k=2}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda^2 e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda^2,$$

dobijemo

$$Var X = \lambda.$$

Primjer 2.3. *Pretpostavimo da neku slastičarnicu tijekom jednog sata u prosjeku posjeti 17 ljudi. Posjetitelje broji slučajna varijabla s Poissonovom distribucijom čiji je parametar $\lambda = 17$, tj. $X \sim \mathcal{P}(17)$. Broj posjetitelja je iz skupa $\mathcal{R}(X) = \{0, 1, \dots\}$, a pripadne vjerojatnosti su*

$$p_i = \frac{17^i}{i!} e^{-17}.$$

Ako bi nas zanimalo kolika je vjerojatnost da je između 16 i 17 sati slastičarnu posjetilo točno 22 ljudi, to bi jednostavno izračunali:

$$P(X = 22) = \frac{17^{22}}{22!} e^{-17} = 0.0433,$$

a vjerojatnost da je tada broj posjetitelja bio veći od 22 je:

$$P(X > 22) = 1 - P(X \leq 22) = 1 - \sum_{i=0}^{22} \frac{17^i}{i!} e^{-17} = 0.095.$$

Napomenimo da je nebitno u kojem vremenskom intervalu računamo vjerojatnost, jer je kod Poissonove distribucije vjerojatnost neovisna o tome o kojem se intervalu radi, pa su pripadne vjerojatnosti iste za npr. interval između 12 i 13 sati.

Napomena 2.1. *Poissonovu distribuciju možemo promatrati kao granični slučaj binomne s parametrima n i p , kada n teži u beskonačno, a p u nulu, tako da produkt np ostaje konstantan.*

Uzmimo da je

$$\lambda = np, \quad p_i(n) = \binom{n}{i} p^i q^{n-i}$$

i pokažimo da je

$$\lim_{n \rightarrow \infty} p_i(n) = \frac{\lambda^i}{i!} \cdot e^{-\lambda}.$$

Imamo:

$$\begin{aligned} p_i(n) &= \binom{n}{i} p^i q^{n-i} = \binom{n}{i} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n!}{i!(n-i)!} \cdot \frac{\lambda^i}{n^i} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \frac{\overbrace{n(n-1)(n-2)\dots(n-(i-1))(n-i)!}^{i \text{ brojeva, svaki dijelimo sa } n}}{i!(n-i)!} \cdot \frac{\lambda^i}{\underbrace{n \cdot n \cdot \dots \cdot n}_{i \text{ puta}}} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= 1 \cdot \left(1 - \frac{\lambda}{n}\right) \cdot \left(1 - \frac{\lambda}{n}\right) \cdot \dots \cdot \left(1 - \frac{\lambda}{n}\right) \cdot \frac{\lambda^i}{i!} \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-i}. \end{aligned}$$

Kada $n \rightarrow \infty$ faktor $\left(1 - \frac{\lambda}{n}\right)^n$ će konvergirati u $e^{-\lambda}$, a preostali faktori u zagradama će konvergirati u 1.

Slijedi:

$$\lim_{n \rightarrow \infty} p_i(n) = \frac{\lambda^i}{i!} \cdot e^{-\lambda}$$

što je jednako vjerojatnosti slučajne varijable s Poissonovom distribucijom.

Primjer 2.4. U velikoj kutiji nalazi se $p = 1\%$ neispravnih proizvoda. Ako uzmemo uzorak od $n = 100$ proizvoda iz kutije, slučajna varijabla $X =$ broj neispravnih proizvoda u uzorku ima binomnu distribuciju $X \sim \mathcal{B}(n, p)$. Budući da je n velik u odnosu na mali p i $np = 1$, slučajnu varijablu X možemo aproksimirati Poissonovom distribucijom $X \sim \mathcal{P}(np)$. Izračunajte vjerojatnost da se pojavi barem jedan neispravan proizvod u uzorku.

Rješenje:

Računamo po Poissonovoj distribuciji: $X \sim \mathcal{P}(\lambda)$, $\lambda = np = 1$

$$\begin{aligned} P(X \geq 1) &= 1 - P(X < 1) = 1 - P(X = 0) \\ &= 1 - \frac{\lambda^0}{0!} \cdot e^{-\lambda} = 0.63212 \end{aligned}$$

Ako računamo po binomnoj distribuciji: $X \sim \mathcal{B}(n, p)$, $n = 100$, $p = \frac{1}{10}$

$$\begin{aligned} P(X \geq 1) &= 1 - P(X < 1) = 1 - P(X = 0) \\ &= 1 - \binom{100}{0} \left(\frac{99}{100}\right)^{100} = 0.63397 \end{aligned}$$

Uspoređivanjem rezultata uočavamo da su približno jednaki; dakle Poissonova distribucija se zaista pod određenim uvjetima može aproksimirati binomnom.

2.6 Geometrijska distribucija

Ovaj tip distribucije također se koristi pri nezavisnom ponavljanju Bernoullijevog pokusa. Međutim, ovdje nas zanima koliko puta trebamo ponoviti pokus do pojave prvog uspjeha? Ako je vjerojatnost pojave uspjeha u jednom izvođenju pokusa jednaka p (dogadjaj A), onda je vjerojatnost da se uspjeh prvi put pojavi u k -tom izvođenju pokusa opisana na sljedeći način:

$$\begin{aligned} P(X = k) &= P(\underbrace{A^c \cap A^c \cap \dots \cap A^c}_{k-1 \text{ puta neuspjeh}} \cap \underbrace{A}_{\text{uspjeh}}) \\ &= P(A^c) \cdot P(A^c) \cdot \dots \cdot P(A^c) \cdot P(A) \\ &= (1 - p)^{k-1} p \end{aligned}$$

Definicija 2.4. *Slučana varijabla X ima geometrijsku distribuciju s parametrom $p \in \langle 0, 1 \rangle$ ako prima vrijednosti iz skupa $\{1, 2, 3, \dots\}$ s vjerojatnostima*

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

$$\text{Očekivanje: } EX = \frac{1}{p}$$

$$\text{Varijanca: } VarX = \frac{1 - p}{p^2}$$

Primjer 2.5. *Doktor treba prepisati lijek protiv depresije pacijentu kojemu je upravo postavio dijagnozu. Pretpostavimo da je, od svih dostupnih lijekova, vjerojatnost da će mu pojedini od njih pomoći jednaka $p = 0.6$. Pacijent će uzimati jedan lijek, no ako mu ne pomogne, prestat će i početi uzimati drugi i tako redom dok ne pronađe odgovarajući i počne ga uzimati za stalno. Kolika je vjerojatnost da prvi lijek koji pacijentu pomogne bude ujedno i prvi koji je isprobao? Kolika je vjerojatnost da to bude drugi, treći po redu, i tako dalje? Koliko će očekivano pacijent isprobati lijekova dok ne pronađe onaj koji će mu pomoći?*

Rješenje:

Vjerojatnost da će pacijentu pomoći prvi probani lijek je $p_1 = p(1 - p)^{1-1} = 0.6$, dok je vjerojatnost da će to biti drugi $p_2 = p(1 - p)^{2-1} = 0.6 \cdot 0.4 = 0.24$, a treći $p_3 = p(1 - p)^{3-1} = 0.6 \cdot 0.4^2 = 0.096$. Slično se dobije za ostale.

Očekivani broj probanih lijekova je $EX = \frac{1}{p} = \frac{1}{0.6} = 1.667$, što znači da se očekuje da će pacijentu pomoći prvi ili drugi isprobani lijek.

2.7 Negativna binomna distribucija

Slično kao kod geometrijske distribucije, broji se koliko puta bi trebali ponoviti pokus da bi se ostvarilo, ne jedan, nego r uspjeha. Vjerojatnost da se r -ti uspjeh pojavio u k -tom ponavljanju pokusa iznosi

$$P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, \dots$$

Parametri: r – broj uspjeha, p – vjerojatnost uspjeha u pojedinom pokusu

Slučajna varijabla X distribuirana ovom distribucijom zapravo je suma r nezavisnih slučajnih varijabli distribuiranih istom geometrijskom distribucijom:

$$X = X_1 + X_2 + \dots + X_r,$$

gdje je X_i slučajna varijabla koja broji pokušaje do i -tog uspjeha, $i = 1, \dots, r$.

Očekivanje (linearnost očekivanja):

$$EX = E(X_1 + X_2 + \dots + X_r) = EX_1 + EX_2 + \dots + EX_r = \frac{1}{p} \cdot r = \frac{r}{p}$$

Varijanca (Teorem 1.4.):

$$\text{Var}X = \text{Var}(X_1 + X_2 + \dots + X_r) = \text{Var}X_1 + \text{Var}X_2 + \dots + \text{Var}X_r = r \cdot \frac{1-p}{p^2}$$

Primjer 2.6. *Ivan je odličan nogometaš. U 70% slučajeva, ako Ivan dobije priliku, on zabije gol. Dakle, vjerojatnost uspjeha je $p = 0.7$. Tijekom sezone, kolika je vjerojatnost da je Ivan zabio peti po redu gol u sedmoj prilici koju je dobio?*

Dakle, treba nam vjerojatnost da je peti uspjeh (gol) postignut u sedmom ponavljanju (prilici):

$$P(X = 7) = \binom{7-1}{5-1} p^5 (1-p)^{7-5} = \binom{6}{4} \cdot 0.7^5 \cdot 0.3^2 = 0.2269.$$

2.8 Hipergeometrijska distribucija

Zadan je skup S od N elemenata, ali tako da je M elemenata iz tog skupa prve vrste, a preostalih $(N - M)$ druge vrste. Slučajna varijabla koja broji elemente prve vrste u slučajnom uzorku od n elemenata iz skupa S modelira se hipergeometrijskom distribucijom.

Definicija 2.5. *Diskretna slučajna varijabla X ima hipergeometrijsku distribuciju s parametrima N, M i $n, N, M, n \in \mathbb{N}$ ako prima vrijednosti iz skupa $\mathcal{R}(X) = \{k \in \mathbb{N} : \max(0, n - N + M) \leq k \leq \min(n, M)\}$ s vjerojatnostima*

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

Očekivanje:

$$\begin{aligned} EX &= \sum_{k=0}^n k \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \stackrel{(*)}{=} n \frac{M}{N} \sum_{k=1}^n \frac{\binom{M-1}{k-1} \binom{N-M}{n-k}}{\binom{N-1}{n-1}} \\ &= n \frac{M}{N} \sum_{j=0}^{n-1} \frac{\binom{M-1}{j} \binom{N-M}{n-1-j}}{\binom{N-1}{n-1}} \\ &\stackrel{(**)}{=} n \frac{M}{N} \end{aligned}$$

(*) Općenito vrijedi: $\binom{n}{k} = \frac{n}{k} \binom{n-1}{k-1}$

(**) (Vandermondeov) identitet:

$$\sum_{j=0}^p \binom{a}{j} \binom{b}{p-j} = \binom{a+b}{p}, \quad a, b \in \mathbb{N}$$

Varijanca: Slično kao u izračunavanju očekivanja, koristeći Teorem 1.2. te (*) i (**) imamo:

$$\begin{aligned}
 E(X(X-1)) &= \sum_{k=0}^n k(k-1) \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \\
 &\stackrel{(*)}{=} n(n-1) \frac{M(M-1)}{N(N-1)} \sum_{k=2}^n \frac{\binom{M-2}{k-2} \binom{N-M}{n-k}}{\binom{N-2}{n-2}} \\
 &= n(n-1) \frac{M(M-1)}{N(N-1)} \sum_{j=0}^{n-2} \frac{\binom{M-2}{j} \binom{N-M}{n-2-j}}{\binom{N-2}{n-2}} \\
 &\stackrel{(**)}{=} n(n-1) \frac{M(M-1)}{N(N-1)}
 \end{aligned}$$

Sada je:

$$\begin{aligned}
 \text{Var}X &= EX^2 - (EX)^2 = E(X(X-1)) + EX - (EX)^2 \\
 &= n(n-1) \frac{M(M-1)}{N(N-1)} + n \frac{M}{N} - n^2 \frac{M^2}{N^2} \\
 &= n(n-1) \frac{nM(N-M)(N-n)}{N^2(N-1)} \\
 &= \frac{nM}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}
 \end{aligned}$$

Zadatak 2.2. Iz vaze koja sadrži 4 crvene i 6 bijelih ruža izvlačimo 3 ruže. Slučajna varijabla X koja broji izvučene crvene ruže (uspjeh) modelirana je hipergeometrijskom distribucijom. Kako izgleda tablica distribucije i koliki je očekivani broj izvučenih crvenih ruža?

Rješenje:

Slika ove slučajne varijable je $\mathcal{R}(X) = \{0, 1, 2, 3\}$, a pripadne vjerojatnosti izračunamo po formuli pa je

$$P(X=0) = \frac{\binom{4}{0} \binom{6}{3}}{\binom{10}{3}} = \frac{1}{6}, \quad P(X=1) = \frac{1}{2}, \quad P(X=2) = \frac{3}{10}, \quad P(X=3) = \frac{1}{30}.$$

Tablica je sljedeća:

$$X = \begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{6} & \frac{1}{2} & \frac{3}{10} & \frac{1}{30} \end{pmatrix}$$

Očekivani broj crvenih ruža: $EX = 3 \cdot \frac{4}{10} = 1.2$.

Primjedba 2.1. Neka je X slučajna varijabla koja ima hipergeometrijsku distribuciju s parametrima N, M i n . Ako $N \rightarrow \infty$ i $M \rightarrow \infty$ tako da $\frac{M}{N} \rightarrow p$, gdje je p pozitivna konstanta, tada

$$\lim_{N \rightarrow \infty} \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} = \binom{n}{k} p^k (1-p)^{n-k}, \quad \forall k \in \{0, 1, \dots, n\}.$$

Dakle, hipergeometrijska distribucija se može aproksimirati binomnom distribucijom, u slučaju kad je n malen u odnosu na N i M .

To se može opravdati time da je vjerojatnost odabira n elemenata iz N -članog skupa bez vraćanja prethodno izvučenih, za velik N , približno jednaka vjerojatnosti odabiranja n elemenata s vraćanjem izvučenih elemenata u skup.

2.9 Negativna hipergeometrijska distribucija

Kod hipergeometrijske se broje elementi prve vrste u slučajnom uzorku od n elemenata iz skupa sa ukupno N elemenata od kojih je M prve vrste. Kod negativne hipergeometrijske imamo isti skup, a pokus je sljedeći:

Izvlačimo jedan element i bilježimo koje je vrste; zatim bez vraćanja prvog izvlačimo drugi element i također bilježimo koje je vrste i tako nastavljamo dok zabilježenih elemenata druge vrste ne bude r .

Tada prestajemo s izvlačenjem i prebrojimo elemente prve vrste koji su se pojavili u tom pokusu.

Zanima nas vjerojatnost da smo izvukli točno k elemenata prve vrste.

To se može izračunati tako da razmatramo vjerojatnost pojave k elemenata prve vrste u $(k+r-1)$ izvlačenja, te da se u $(k+r)$ -tom izvukao element druge vrste. Prva vjerojatnost se dobije primjenom hipergeometrijske distribucije, a vjerojatnost da se u $(k+r)$ -tom izvukao element druge vrste je jednostavno broj preostalih takvih $(N-M-(r-1))$ podijeljen sa brojem svih preostalih elemenata $(N-(k+r-1))$.

Množenjem tih dvaju vjerojatnosti dobili smo vjerojatnost da se do zabilježenog r -tog elementa druge vrste pojavilo točno k elemenata prve vrste:

$$\frac{\binom{M}{k} \binom{N-M}{k+r-1-k}}{\binom{N}{k+r-1}} \cdot \frac{N-M-(r-1)}{N-(k+r-1)} = \frac{\binom{k+r-1}{k} \binom{N-r-k}{M-k}}{\binom{N}{M}} = P(X=k)$$

Slučajna varijabla koja na opisani način broji elemente prve vrste (uspjehe) poprima vrijednosti iz skupa $\{0, 1, \dots, M\}$ sa gore navedenim pripadnim vjerojatnostima $p_k = P(X = k)$ i kaže se da je modelirana negativnom hipergeometrijskom distribucijom.

Parametri:

N –ukupan broj elemenata skupa

M –ukupan broj elemenata prve vrste

r –broj elemenata druge vrste koje moramo izvući

Očekivanje:

$$EX = r \cdot \frac{M}{N - M + 1}$$

Varijanca:

$$VarX = r \cdot \frac{M(N + 1)(N - M - r + 1)}{(N - M + 1)^2(N - M + 2)}$$

Napomena 2.2. *Negativna hipergeometrijska distribucija (kao i hipergeometrijska) razmatra pokuse u kojima se izvlačenja vrše bez vraćanja izvučenih elemenata u skup, tako da je vjerojatnost uspjeha (elemenata prve vrste) različita u svakom od izvlačenja. Suprotno, negativna binomna distribucija (kao i binomna) razmatra pokuse u kojima se izvlačenja vrše s vraćanjem, tako da je vjerojatnost uspjeha ista u svakom izvlačenju i izvlačenja su međusobno nezavisna. Sljedeća tablica ilustrira navedenu razliku među distribucijama:*

	s vraćanjem	bez vraćanja
<i>Broj uspjeha u n izvlačenja</i>	binomna	hipergeometrijska
<i>Broj uspjeha sa točno r neuspjeha</i>	negativna binomna	negativna hipergeometrijska

Tablica 2.1: klasifikacija: s vraćanjem/bez vraćanja

Literatura

- [1] G. R. GRIMMETT, D. R. STIRZAKER, *Probability and Random processes*, Oxford Univ. Press, Oxford, 2001.
- [2] M. BENŠIĆ, N. ŠUVAK *Uvod u vjerojatnost i statistiku*, Odjel za matematiku, Osijek, 2013.
- [3] N. SARAPA, *Teorija vjerojatnosti*, Školska knjiga, Zagreb, 2002.