Centralni limitni izrek

Naum Kostoski, Andrej Matevc, Lucija Matijašić
Mentor: Žan Hafner Petrovski

Povzetek

Pri našem projektu smo se ukvarjali z verjetnostjo in spoznali osnovne pojme statistike. Z enostavnimi primeri, kot je met kovanca, smo si postavili vprašanja, na katera je pomagal odgovoriti centralni limitni izrek. Čeprav smo ta izrek uporabili na Marsu, smo prepričani, da nam lahko pomaga tudi na Zemlji.

1 Uvod

Vsakdo je že kdaj metal kovance. Morda ste se že vprašali, kolikšna je verjetnost, da pade grb. To zveni zelo preprosto. Sedaj pa si predstavljajte, da morate vreči 1000 kovancev in vas zanima, kolikšna je verjetnost, da pade grb med 750 in 800-krat. Pri tem si lahko pomagate s centralnim limitnim izrekom. Čeprav se vam taka vprašanja ne porajajo pogosto, je ta izrek zelo pomemben v statistiki.

2 Osnovni pojmi

Začnimo s splošno definicijo verjetnosti.

Definicija 1. *Verjetnost je število, ki pove, kolikšna je možnost, da se zgodi dogodek.*

Naj bo A dogodek sestavljen iz m izidov, vseh možnih izidov pa naj bo n. Klasična definicija verjetnosti pravi, da je verjetnost, da se zgodi dogodek A, razmerje med številom ugodnih izidov in številom vseh možnih izidov. To velja pod pogojem, da so vsi izidi enako verjetni. Pišemo

$$P(A) = \frac{m}{n}.$$

Zgled 1. *Pri metu poštene kocke je možnih šest izidov. Dogodek, da pade šestica sestavlja en izid, vsi izidi pa so enako verjetni. Z upoštevanjem zgornje definicije dobimo, da je verjetnost $P(\textit{pade šestica}) = \frac{1}{6}$.*

Zapišimo nekaj lastnosti verjetnosti:

- veljati mora $P(\{\text{vsi možni izidi}\}) = 1$,
- če sta A in B disjunktna dogodka, je $P(A \cup B) = P(A) + P(B)$,
- za poljubna dogodka A in B je $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Definicija 2. *Dogodka A in B sta neodvisna, če velja*

$$P(A \cap B) = P(A)P(B).$$

Zgled 2. *Imamo vrečo z desetimi oštevilčenimi kroglicami. Iz nje naključno ter brez vračanja vzamemo tri kroglice. Zanima nas, kolikšna je verjetnost, da izvlečemo žogice oštevilčene z 1, 2 in 3, pri tem pa zaporedje ni pomembno.*

Oglejmo si najprej število, ki nam pove, koliko različnih k-teric kroglic lahko izvlečemo iz vrečke z n kroglicami. Izračunamo ga lahko z binomskim simbolom $\binom{n}{k}$, ki je definiran s funkcijo fakulteta na sledeč način

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Za izračun iskane verjetnosti opazimo, da so vsi izmed možnih $\binom{10}{3}$ dogodkov enako verjetni, mi pa iščemo natanko enega izmed njih. Zapišemo lahko

$$P(\textit{izvlečemo kroglice oštevilčene z } 1, 2, 3) = \frac{1}{\binom{10}{3}}.$$

Oglejmo si še primer, pri katerem nimajo vsi izidi enake verjetnosti.

Zgled 3. *Kovanec vržemo desetkrat. Meti so med seboj neodvisni, verjetnost, da pade grb naj bo* $p \in [0, 1]$. *Kolikšna je verjetnost, da grb pade v prvih treh metih, potem pa ne več? To lahko zapišemo kot dogodek*

$$\omega = GGGCCCCCCC,$$

kjer G *predstavlja grb,* C *pa cifro. Moč množice vseh izidov je enaka*

$$| \{G, C\}^{10} | = 2^{10}.$$

Tu moramo biti pozorni, saj razen v izjemnem primeru, ko je $p = \frac{1}{2}$, *izidi med seboj niso enako verjetni. Prava verjetnost dogodka* ω *je*

$$P(\omega) = p^3(1-p)^{10-3}.$$

Razmislimo zdaj še, kolikšna je verjetnost, da grb pade trikrat, torej ne nujno na prvih treh mestih. Preštetí je treba, na koliko različnih načinov lahko vržemo tri grbe. To naredimo tako, da izmed desetih zaporednih metov izberemo tri, pri katerih bo padel grb, to pa lahko naredimo natanko na $\binom{10}{3}$ *načinov. Velja torej*

$$P(padejo\ trije\ grbi) = \binom{10}{3} p^3(1-p)^7.$$

Zgled 4. *Zdaj bomo met kovanca predstavili z diskretno slučajno spremenljivko* X, *ki se uporablja kot indikator tega dogodka. Predpišemo, da ima vrednost* 1, *če pade grb in* 0, *če pade cifra. Verjetnost, da pade grb, je* p, *verjetnost, da pade cifra, pa* $1-p$. *To drugače zapišemo kot*

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

Bolj splošno lahko diskretno slučajno spremenljivko predstavimo takole:

$$\begin{pmatrix} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix},$$

pri čemer so a_i možni izidi, p_i pa je verjetnost, da se zgodi a_i. Pri tem mora veljati

$$\sum_i p_i = 1.$$

Definicija 3. *Porazdelitev, podano s predpisom*

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

imenujemo binomska porazdelitev. Označimo

$$X \sim Bin(n, p).$$

Definirajmo, kdaj sta dve slučajni spremenljivki neodvisni.

Definicija 4. *Dve slučajni spremenljivki X in Y sta neodvisni, kadar velja*

$$P(X = x)P(Y = y) = P(X = x, Y = y),$$

za vse možne pare x, y.

3 Pričakovana vrednost

Definicija 5. *Pričakovano vrednost slučajne diskretne spremenljivke X definiramo kot*

$$E(X) = \sum_k kP(X = k).$$

Trditev 1. *Pričakovana vrednost slučajne spremenljivke je linearna. To pomeni, da veljata naslednji enakosti:*

- *$E(aX) = aE(X)$ za vsak $a \in \mathbb{R}$,*
- *$E(X + Y) = E(X) + E(Y)$.*

Definicija 6. *Naj bo $f : \mathbb{R} \to \mathbb{R}$ in X diskretna slučajna spremenljivka. Pričakovano vrednost slučajne spremenljivke $f(X)$ definiramo kot*

$$E(f(X)) = \sum_k f(k)P(X = k).$$

Oglejmo si preprost zgled izračuna pričakovane vrednosti po definiciji.

Zgled 5. *Naj bo indikator I definiran tako kot zgoraj. Po definiciji izračunajmo njegovo pričakovano vrednost.*

$$\begin{aligned}
E(X) &= \sum_k kP(X=k) \\
&= \sum_{k=0}^{1} kP(X=k) \\
&= 1 \cdot p + 0 \cdot (1-p) \\
&= p
\end{aligned}$$

Število grbov pri metu n-tih kovancev je porazdeljeno binomsko. Vsak od teh kovancev lahko prispeva en grb ali nič grbov, tako da lahko število grbov predstavimo tudi z vsoto n-tih indikatorjev.

Zapišimo to še simbolično. Naj bo $X \sim Bin(n,p)$, potem velja

$$X = \sum_{i=1}^{n} I_i.$$

Sledi, da je pričakovana vrednost X enaka

$$E(X) = \sum_{i=1}^{n} E(I_i) = np.$$

4 Varianca

Varianca meri, kako razpršena je porazdelitev oziroma kako daleč od pričakovane vrednosti se gibljejo vrednosti.

Definicija 7. *Varianco diskretne slučajne spremenljivke X definiramo kot*

$$var(X) = \sum_k (k - E(X))^2 P(X=k).$$

Trditev 2. *Za neodvisni slučajni spremenljivki X in Y velja*

$$var(X+Y) = var(X) + var(Y).$$

Dokaz prepustimo bralcu.

Trditev 3. *Če je $a \in \mathbb{R}$ konstanta, potem za slučajno spremenljivko X velja, da je*

$$var(aX) = a^2 var(X).$$

Dokaz:

$$\begin{aligned}
var(aX) &= \sum_k (ak - aE(X))^2 P(X = k) \\
&= \sum_k a^2 (k - E(X))^2 P(X = k) \\
&= a^2 \sum_k (k - E(X))^2 P(X = k) \\
&= a^2 var(x)
\end{aligned}$$

□

Trditev 4. *Za varianco slučajne spremenljivke X velja*

$$var(X) = E(X^2) - E(X)^2.$$

Dokaz:

$$\begin{aligned}
var(X) &= \sum_k (k - E(X))^2 P(X = k) \\
&= E((X - E(X))^2) \\
&= E(X^2 - 2XE(X) + E(X)^2) \\
&= E(X^2) - 2E(X)^2 + E(X)^2 \\
&= E(X^2) - E(X)^2
\end{aligned}$$

□

Zgled 6. *Naj bo indikator I definiran tako kot zgoraj. Po definiciji izračunajmo njegovo varianco.*

$$\begin{aligned}
var(I) &= \sum_k (k - E(I))^2 P(I = k) \\
&= (1 - p)^2 p + p^2 (1 - p) \\
&= p - p^2
\end{aligned}$$

Iz trditve 2 in zgleda 6 sledi, da je za $X \sim Bin(n, p)$ varianca

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

5 Kumulativna porazdelitev

Kovanec vržemo n-krat. Naj slučajna spremenljivka X predstavlja število grbov. Podano imamo verjetnost p, da pade grb in neko število k med 0 in n. Kolikšna je verjetnost, da pade največ k grbov? To verjetnost zapišemo kot

$$P(X \leq k) = \sum_{l=0}^{k} \binom{n}{l} p^l (1-p)^{n-l}.$$

Ta primer prikazuje kumulativno porazdelitev.

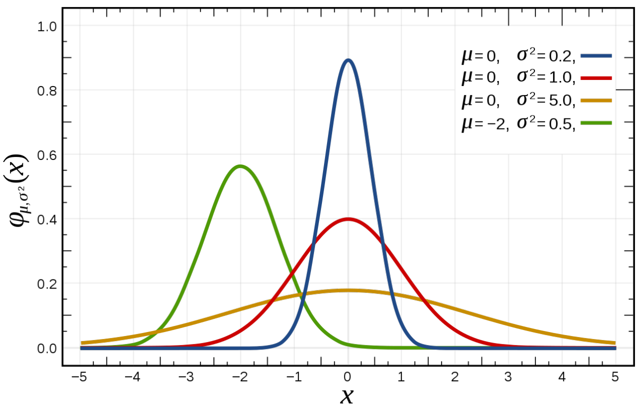
Če postavimo spodnjo mejo a in zgornjo mejo b, potem lahko verjetnost, da ima X vrednost med a in b, zapišemo kot

$$P(a \leq X \leq b) = P(X \leq b) - P(X < a).$$

6 Standardna normalna porazdelitev

Standardna normalna porazdelitev je poseben primer normalne oziroma Gaussove porazdelitve, kjer sta $E(X) = 0$ in $var(X) = 1$. Gostota standardne normalne porazdelitve je

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

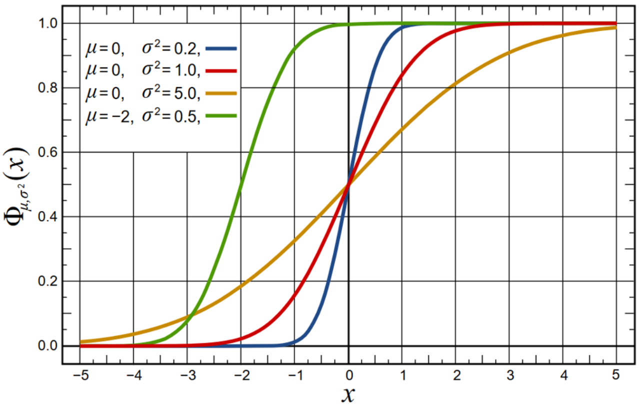


Slika 1: Graf gostote standardne normalne porazdelitve je prikazan z rdečo barvo

Njeno kumulativno porazdelitev označimo s

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t) \ \mathrm{dt},$$

kar je enako ploščini med absciso in grafom gostote standardne normalne porazdelitve.



Slika 2: Graf kumulativne porazdelitve

7 Pričakovana vrednost 0 in varianca 1

Če želimo, da imata binomska porazdelitev in standardna normalna porazdelitev enaki pričakovani vrednosti in enaki varianci, binomski porazdelitvi odštejemo njeno pričakovano vrednost in jo delimo s korenom njene variance. Naj bo torej $X \sim Bin(n, p)$. Vemo

$$\begin{aligned} E(X) &= np, \\ var(X) &= np(1-p). \end{aligned}$$

Definirajmo novo slučajno spremenljivko

$$Y := \frac{X - np}{\sqrt{np(1-p)}}.$$

Varianco Y izračunamo s pomočjo naslednje lastnosti variance:

$$var(aX) = a^2 var(X),$$

kjer je a konstanta.

V našem primeru je $a = \frac{1}{\sqrt{var(X)}}$. Računajmo

$$\begin{aligned} var(Y) &= var(\frac{X - np}{\sqrt{var(X)}}) \\ &= \frac{1}{var(X)} var(X - np) \\ &= \frac{1}{var(X)} var(X) \\ &= 1. \end{aligned}$$

Pri računanju pričakovane vrednosti binomske porazdelitve uporabimo linearnost pričakovane vrednosti. Dobimo

$$\begin{aligned} E(Y) &= E\left(\frac{X - np}{\sqrt{var(X)}}\right) \\ &= E\left(\frac{X}{\sqrt{var(X)}}\right) - E\left(\frac{np}{\sqrt{var(X)}}\right) \\ &= \frac{1}{\sqrt{var(X)}}(E(X) - E(np)) \\ &= 0. \end{aligned}$$

8 Centralni limitni izrek

Brez dokaza navedimo zdaj glavni izrek tega članka.

Izrek 1. *Naj bodo $X_1, X_2, \ldots, X_n$ neodvisne in enako porazdeljene slučajne spremenljivke ter naj bo n dovolj veliko naravno število. Naj bodo $S_n = \sum_i X_i$, $\mu = E(S_n)$, $\sigma^2 = var(S_n)$ in $Z \sim N(0,1)$. Potem za $a < b$ velja aproksimacija:*

$$\begin{aligned} P(a < S_n < b) &\approx P(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}) \\ &= \Phi(\frac{b - \mu}{\sigma}) - \Phi(\frac{a - \mu}{\sigma}) \end{aligned}$$

Uporabimo ta izrek na primeru metanja kovanca.

Zgled 7. *Kovanec vrzimo 30-krat. Podana je verjetnost $p = 0.3$, da pade grb. Naj bodo $X_1, X_2, \ldots, X_{30}$ indikatorji definirani kot*

$$X_i = I_i = \begin{cases} 1; & \textit{pade grb} \\ 0; & \textit{pade cifra} \end{cases}$$

Kolikšna je verjetnost, da bo grb padel vsaj 8-krat in največ 12-krat? Pišimo

$$\begin{aligned} S_{30} &= I_1 + I_2 + \ldots + I_{30} \\ \mu &= 30p \\ \sigma^2 &= 30p(1-p) \end{aligned}$$

Pri računanju verjetnosti si pomagamo s centralnim limitnim izrekom, ki ga lahko uporabimo, saj je $n = 30$ *in to smatramo za dovolj veliko vrednost, točen rezultat pa recimo, da nas trenutno ne zanima. Za aproksimacijo na intervalu* $(8, 12)$ *dobimo*

$$\begin{aligned} P(8 \leq S_{30} \leq 12) &\approx \Phi(\frac{12-9}{\sqrt{6.3}}) - \Phi(\frac{8-9}{\sqrt{6.3}}) \\ &= \Phi(\frac{3}{\sqrt{6.3}}) - \Phi(\frac{-1}{\sqrt{6.3}}) \\ &\approx 0,5366 \end{aligned}$$

Da bi aproksimacijo izboljšali, povečamo interval (a, b) *za* 0.5 *v vsako smer, to pomeni, da gledamo interval* $(a-0.5, b+0.5)$ *oziroma v našem konkretnem primeru interval* $(7.5, 12.5)$. *Če upoštevamo to izboljšavo, dobimo približek* 0.6436. *Spodaj si bomo ogledali še program, ki izračuna točno vrednost in kako se ta primerja s približkom.*

9 Program

Napisali smo tudi Python program, ki izračuna vsoto binomskih vrednosti med a in b in točno izračuna verjetnost $P(a \leq S \leq b)$. Uporabili smo ga na prej opisanem primeru meta 30-tih kovancev.

```
def fakulteta(n):
    a = 1
    for i in range (n):
        a = (i+1)*a
    return a

def binomski(n, k):
    a = fakulteta(n)/(fakulteta(k)*fakulteta(n-k))
    return a

def verjetnost(a, b, n, p):
    x = 0
    for i in range (a, b+1):
        f=(p**i)*((1-p)**(n-i))*binomski(n, i)
```

```
        x = x+f
    return x

print(verjetnost(8, 12, 30, 0.3))
```

Program izpiše vrednost 0.6342. Spomnimo se, da smo za aproksimacijo s centralnim limitnim izrekom za $P(7.5 \leq S_{30} \leq 12.5)$ dobili 0.6436. S približkom smo zagotovo lahko zadovoljni.

Literatura

[1] Zapiski s predavanj predmetov Verjetnost in Statistika v šolskem letu 2018/19 na UL FMF, izvajatelja dr. Mihael Perman in dr. Martin Raič