5. Diskretne vjerojatnosne razdiobe

Binomna razdioba

U praksi se često susrećemo s problemima u kojima se isti pokus ponavlja više puta neovisno, i pod jednakim uvjetima:
– niz uzastopnih bacanja novčića,
– proizvodnja na nekom stroju, koji proizvodi 5% defektnih proizvoda,
– ispitivanje javnog mijenja o ulasku Hrvatske u EU.
U svi tim pokusima imamo dva moguća elementarna ishoda:

P/G, Ispravan/Neispravan proizvod, Da/Ne.

Jednog od njih ćemo zvati **uspjeh** (U), a drugog **neuspjeh** (N).
Vjerojatnost uspjeha je u svakom pokusu jednaka, i označavat ćemo je s $p \in [0, 1]$:

$$P(U) = p.$$

Stoga za vjerojatnost neuspjeha imamo da je

$$P(N) = 1 - p = q.$$

(Primijetimo da uspjeh i neuspjeh predstavljaju suprotne događaje.)

Pretpostavimo da se pokus ponavlja n puta.
Tada broj opažanja uspjeha u n ponavljanja pokusa predstavlja slučajnu varijablu X koja može poprimiti vrijednosti

$$X = 0, 1, 2, \ldots, n.$$

Zanima nas kolika je vjerojatnost da slučajna varijabla X poprimi neku od gornjih vrijednosti

$$p_k = P\{X = k\}.$$

Za određivanje p_k proučit ćemo sve moguće serije od n pokusa u kojima je uspjeh nastupio k puta, a neuspjeh $n - k$ puta.
Ishod takve jedne serije je

$$\underbrace{UUU.\ldots U}_{k}\underbrace{NN\ldots N}_{n-k}$$

Kolika je vjerojatnost njenog pojavljivanja?
Budući da su pokusi neovisni, zaključujemo da može nastupiti s vjerojatnošću

$$p^k q^{n-k}.$$

Ostaje nam prebrojiti koliko ima takvih serija (kombinatorika!).
Problem se svodi na pitanje na koliko načina možemo izabrati k pokusa (koji su za ishod imali uspjeh) među njih n.
Odgovor je izražen binomnim koeficijentom

$$\binom{n}{k},$$

gdje je

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k \cdot (k-1) \cdot \ldots \cdot 1},$$

dok je za $n \in \mathbf{N}$, $n! = n \cdot (n-1) \cdot \ldots \cdot 1$, te $0! = 1$.

Dakle, imamo $\binom{n}{k}$ serija u kojima je broj uspjeha jednak k, svaka od nih ima jednaku vjerojatnost pojavljivanja $p^k q^{n-k}$, te je stoga

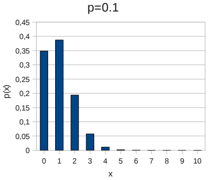
$$p_k = P\{X = k\} = \binom{n}{k} p^k q^{n-k}.$$

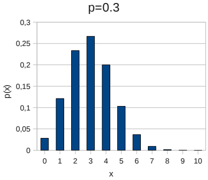
Za diskretnu varijablu X koja poprima vrijednosti iz skupa $\{0, 1, 2, \ldots, n\}$ s pripadnim vjerojatnostima

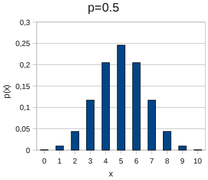
$$p_k = P\{X = k\} = \binom{n}{k} p^k q^{n-k},$$

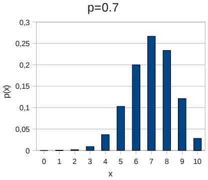
kažemo da ima **binomnu razdiobu** s parametrima n i p $(n \in N, p \in [0, 1])$.

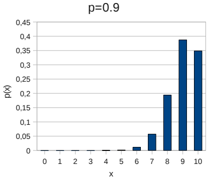
Oznaka $X \sim B(n, p)$.
Koji je njezin oblik?
Za $n = 10$ nacrtani su grafovi binomne razdiobe za razne vrijednosti vjerojatnosti p.


p=0.1
p(x)
0 0,05 0,1 0,15 0,2 0,25 0,3 0,35 0,4 0,45
0 1 2 3 4 5 6 7 8 9 10
x


p=0.3
p(x)
0 0,05 0,1 0,15 0,2 0,25 0,3
0 1 2 3 4 5 6 7 8 9 10
x


p=0.5
p(x)
0 0,05 0,1 0,15 0,2 0,25 0,3
0 1 2 3 4 5 6 7 8 9 10
x


p=0.7
p(x)
0 0,05 0,1 0,15 0,2 0,25 0,3
0 1 2 3 4 5 6 7 8 9 10
x


p=0.9
p(x)
0 0,05 0,1 0,15 0,2 0,25 0,3 0,35 0,4 0,45
0 1 2 3 4 5 6 7 8 9 10
x

Vidimo da je razdioba simetrična za $p = 0.5$, lijevo (desno) asimetrična za $p < 0.5$ $(p > 0.5)$.
Logično, ako je $p = 0.5$ onda je najvjerojatnije da u n ponavljanja pokusa broj uspjeha bude $\frac{n}{2}$.
Lagano se provjeri da vrijedi

- $p_k \geqslant 0, \quad k = 0, 1, 2, \dots n\,,$
- $\sum_{k=0}^{n} p_k = 1,$

čime su zadovoljeni nužni uvjeti diskretne razdiobe.
Zaista,

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

pri čemu smo koristili **binomnu formulu**:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k \, b^{n-k}.$$

Numeričke karakteristike binomne razdiobe

Za očekivanje i varijancu slučajne varijable $X \sim B(n,p)$ vrijedi

$$\mu = np,\ \sigma^2 = np(1-p).$$

Dem.

$$\begin{aligned}
\mu = \sum_{k=0}^{n} k\, p_k &= \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k} \\
&= \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k q^{n-k} \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k} \\
&\overset{j=k-1}{=} np \sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-1-j)!} p^j q^{n-1-j} \\
&= np\,(p+q)^{n-1} = np.
\end{aligned}$$

Nadalje,

$$\sigma^2 = \sum_{k=0}^{n} k^2 p_k - \mu^2.$$

Primjenjujući sličan postupak kao gore, za drugi moment imamo

$$\begin{aligned}
\sum_{k=0}^{n} k^2 \binom{n}{k} p^k q^{n-k} &= \sum_{k=0}^{n} \Big(k(k-1)+k\Big) \binom{n}{k} p^k q^{n-k} \\
&= \sum_{k=2}^{n} k(k-1) \binom{n}{k} p^k q^{n-k} + \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k} \\
&= \sum_{k=2}^{n} n(n-1) \binom{n-2}{k-2} p^2 p^{k-2} q^{n-k} + np \\
&\overset{j=k-2}{=} n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j q^{n-2-j} + np \\
&= n(n-1)p^2 (p+q)^{n-2} + np \\
&= (np)^2 - np^2 + np.
\end{aligned}$$

Stoga je

$$\sigma^2 = (np)^2 - np^2 + np - (np)^2 = np(1-p).$$

Q.E.D.

Kada koristiti binomnu razdiobu?

1. Pokus ponavljamo n puta.
2. Svakom pokusu su pridružena dva elementarna događaja – uspjeh (U) i neuspjeh (N).
3. Vjerojatnost uspjeha je za svaki pokus jednaka – p.
4. Pokusi su međusobno neovisni.

Tada slučjna varijabla X koja označuje broj uspjeha u n ponavljanja pokusa ima binomnu razdiobu $B(n, p)$.
Primjeri:
– bacanje novčića $(p = 0.5)$,
– bacanje kocke (vjerojatnost da padne 6-ica je jednaka $p = \frac{1}{6}$),
– provjera ispravnosti proizvoda na stroju koji proizvodi 5% neispravne robe $(p = 0.95)$.

Pomoću binomne razdiobe možemo rješavati i složenije probleme.

Primjer 5.6. Na dan izbora na birališta je izašlo 80% birača. Petero nasumce izabranih ljudi s pravom glasa pitali ste da li su izašli na izbore.
a) Kolika je vjerojatnost da je točno 3 ispitanika izašlo na izbore?
b) Kolika je vjerojatnost da niti jedan ispitanik nije izašao na izbore?
c) Koji je najvjerojatniji broj ispitanika koji je izišao na izbore?

Rješenje:
X - broj ispitanika koji su izašli na izbore.
$X \sim B(5, 0.8)$

a) $P\{X = 3\} = p_3 = \binom{5}{3} 0.8^3 0.2^2 = 0.20480\,.$
b) $P\{X = 0\} = p_0 = \binom{5}{0} 0.8^0 0.2^5 = 0.00032\,.$
c) Lako se provjeri da najveću vjerojatnost u iznosu od 40.96% dobijemo za $k = 4$, što je upravo vrijednost matematičkog očekivanja

$$\mu = np = 5 \cdot 0.8 = 4.$$

∎

Općenito, matematičko očekivanje binomne razdiobe nam ukazuje na vrijednost s najvećom vjerojatnosti pojavljivanja. Tu vrijednost nazivamo **modom**.

Pri tom moramo biti oprezni, jer matematičko očekivanje ne mora uvijek imati cjelobrojnu vrijednost (npr. za 3 bacanja novčića $\mu = 3 \cdot 0.5 = 1.5$). Međutim, uvijek postoji vrijednost k binomne slučajne varijable kojoj pripada najveća vjerojatnost (**mod**) i za koju vrijedi

$$np - q \leqslant k \leqslant np + p.$$

Ukoliko je $np + p$ cijeli broj, onda postoje dvije takve vrijednosti (moda): $np + p$ i $np + p - 1 = np - q$.
Npr, za $B(3, 0.5)$ imamo:

$$P\{X = 0\} = p_0 = 0.5^3 = \frac{1}{8}$$
$$P\{X = 1\} = p_1 = 3 \cdot 0.5^2 \cdot 0.5 = \frac{3}{8}$$
$$P\{X = 2\} = p_2 = 3 \cdot 0.5 \cdot 0.5^2 = \frac{3}{8}$$
$$P\{X = 3\} = p_3 = 0.5^3 = \frac{1}{8}.$$

Vidimo da najveću vjerojatnost pojavljivanja imaju upravo vrijednosti $np + p = 2$ i $np - q = 1$.

Primjer 5.7. Što je vjerojatnije u igri s ravnopravnim protivnikom: dobiti 3 partije od 4 ili 5 partija od 8?

Rješenje:

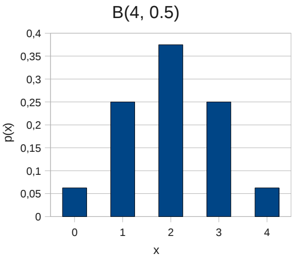
Broj dobivenih partija se ravna po binomnoj razdiobi. U prvom slučaju to je $B(4, \frac{1}{2})$, a u drugom $B(8, \frac{1}{2})$.

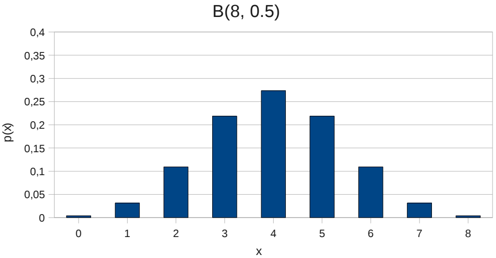
$$P\{3 \text{ partije od } 4\} = \binom{4}{3} \frac{1}{2^3} \frac{1}{2} = \frac{8}{32},$$

$$P\{5 \text{ partija od } 8\} = \binom{8}{5} \frac{1}{2^5} \frac{1}{2^3} = \frac{7}{32}.$$

∎

Rezultat nam može bitio jasniji ukoliko grafički prikažemo gornje razdiobe.





Primijetimo također da vjerojatnost dobitka 3 od 4 partije, nije jednaka vjerojatnosti dobitka 6 od 8 partija.

Primjer 5.8. U Zaljevskom ratu 1990. Amerikanci su koristili rakete Patriot za uništavanje iračkih projektila Scud. Proizvođač tvrdi da raketa Patriot ima 60%-tnu uspješnost pogađanja mete. Pretpostavimo da je uočen projektil SCUD, te da su na njega ispaljena 4 Patriota.
a) Nađite vjerojatnost da sva 4 pogode cilj.
b) Nađite vjerojatnost da će meta biti pogođena.

Rješenje:
X - broj raketa koje su pogodile metu.
$X \sim B(4, 0.6)$

a) $P\{X = 4\} = p_4 = 0.1296\,.$

b) $$\begin{aligned} P\{X \geqslant 1\} &= p_1 + p_2 + p_3 + p_4 \\ &= 1 - P\{X < 0\} \\ &= 1 - p_0 = 1 - 0.0256 = 0.9744\,. \end{aligned}$$

∎

Zadatak 5.1. Uz uvjete prethodnog zadatka odredite minimalni broj raketa koje treba ispaliti da meta bude uništena s vjerojatnošću većom od 99%.

Poissonova razdioba

– dobije se graničnim prijelazom iz binomne razdiobe kad $n \to \infty$, pri tom zadržavajući očekivanje np konstantnim.
Zanima nas čemu pri tom teže pripadne vjerojatnosti, odnosno čemu je jednak

$$\lim_n p_k = \lim_n \binom{n}{k} p^k q^{n-k} = \lim_n \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k \cdot (k-1) \cdot \ldots \cdot 1} p^k (1-p)^{n-k}.$$

Korištenjem oznake $\lambda = np$ imamo da je

$$\begin{aligned} p_k &= \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k \cdot (k-1) \cdot \ldots \cdot 1} \frac{\lambda^k}{n^k} (1 - \frac{\lambda}{n})^{n-k} \\ &= \frac{1 \cdot (1 - \frac{1}{n}) \cdot (1 - \frac{2}{n}) \cdot \ldots \cdot (1 - \frac{k-1}{n})}{k!} \lambda^k \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}. \end{aligned}$$

Stoga za traženi limes imamo da je

$$\lim_n p_k = \frac{\lambda^k}{k!} \lim_n \left(1 - \frac{\lambda}{n}\right)^n = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Na osnovu prethodnih razmatranja možemo izreći sljedeću definiciju.

Za diskretnu varijablu X koja poprima vrijednosti iz skupa $\{0, 1, 2, \ldots, n, \ldots\}$ s pripadnim vjerojatnostima

$$p_k = P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda},$$

kažemo da ima **Poissonovu razdiobu** s parametrom $\lambda \in \mathbf{R}^+$.

Oznaka $X \sim P(\lambda)$.
Gornja definicija uistinu zadaje diskretnu vjerojatnosnu razdiobu, odnosno vrijedi

- $p_k \geqslant 0, \quad k = 0, 1, 2, \ldots,$
- $\sum_{k=0}^{\infty} p_k = 1.$

Zaista,

$$\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1,$$

pri čemu smo koristili da je $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$.

Primijetimo da smo pri izvodu Poissonove razdiobe promatrali granični slučaj kad $n \to \infty$, pri čemu je $\lambda = np$ konstantno.
Očito je to moguće samo ako p istovremeno teži k nuli kad n teži u beskonačnost.
Stoga Poissonova razdioba mjeri broj pojavljivanja događaja koji imaju malu vjerojatnost, tzv. *rijetkih* događaja.
Primjeri takvih, *rijetkih* događaja su

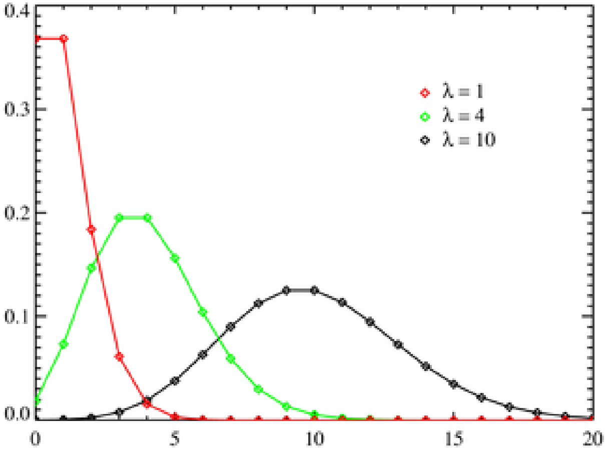
- broj dana u danoj godini u kojima je zabilježena 10%-tna promjena indeksa Crobex,
- broj nesreća na hrvatskim autocestama u jednom mjesecu,
- broj defekata na novom Honda autu,
- broj četvorki rođenih u nekoj zemlji tijekom jedne godine,
- broj virusa na pojedinom računalu, ...

Kada koristiti Poissonovu razdiobu?

Da bi mogli koristiti Poissonovu razdiobu moraju biti ispunjeni sljedeći uvjeti:

1. pokus se sastoji od mjerenja broja povoljnih događaja po nekoj mjernoj jedinici (vremenskoj, prostornoj, ...)
2. vjerojatnost nastanka događaja je za svaku mjernu jedinicu jednaka;
3. ishodi pokusa su međusobno neovisni (pojava događaja u jednoj mjernoj jedinici ne ovisi o njegovom pojavljivanju u ostalim jedinicama);
4. prosječan broj povoljnih događaja po mjernoj jedinici je jednak λ.

Tada slučajna varijabla X koja označuje broj pojavljivanja povoljnih događaja ima Poissonovu razdiobu s parametrom λ.
Na idućem slajdu grafički su prikazane Poissonove razdiobe za niz vrijednosti parametra λ.



(http://en.wikipedia.org/wiki/Poisson_distribution)

Primjećujemo s grafova da se vrijednost s najvećom vjerojatnosti pojavljivanja podudara s parametrom Poissonove razdiobe, λ.
To vrijedi i općenito za Poissonove razdiobe:
mod ili vrijednost s najvećom vjerojatnosti je $\lfloor \lambda \rfloor$ ($\lfloor \cdot \rfloor$ je funkcija najveće cijelo), dok u slučaju kad je $\lambda \in \mathbf{N}$, imamo dva moda: λ i $\lambda - 1$.
Kao i kod binomne razdiobe, matematičko očekivanje nam daje mod (barem približno).

Za očekivanje i varijancu slučajne varijable $X \sim P(\lambda)$ vrijedi

$$\mu = \sigma^2 = \lambda.$$

Dem.

$$\begin{aligned}\mu = \sum_{k=0}^{\infty} k p_k &= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\ &\overset{j=k-1}{=} \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\ &= \lambda e^{-\lambda} e^{\lambda} = \lambda.\end{aligned}$$

Slično, za naći varijancu

$$\sigma^2 = \sum_{k=0}^{\infty} k^2 p_k - \mu^2$$

računamo drugi moment

$$\begin{aligned}
\sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} &= e^{-\lambda} \sum_{k=0}^{\infty} \Big(k(k-1) + k \Big) \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=2}^{\infty} k(k-1) \frac{\lambda^k}{k!} + e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda \\
&= e^{-\lambda} \lambda^2 e^{\lambda} + \lambda = \lambda^2 + \lambda.
\end{aligned}$$

Stoga za varijancu imamo

$$\sigma^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Q.E.D.

A sada primjeri

Primjer 5.9. Automatska telefonska centrala tijekom jedne minute primi prosječno 10 poziva.
a) Nađite vjerojatnost da će centrala tokom jedne minute primiti točno 6 poziva.
b) Nađite vjerojatnost da će tijekom jedne minute biti više od 2 poziva.

Rješenje:
$X \sim P(10)$
a) $P\{X = 6\} = p_6 = \frac{10^6}{6!} e^{-10} = 0.063\,.$
b) $P\{X > 2\} = p_3 + p_4 + p_5 + \ldots$

$$= 1 - P\{X \leqslant 2\}$$
$$= 1 - (p_0 + p_1 + p_2) = 1 - 0.00277 = 0.99723\,.$$

■

Najčešće srednju vrijednost ne znamo.
Tada Poissonovu razdiobu moramo konstruirati na osnovu empiričkih podataka.

Zadatak 5.2. Računalni inženjer je analizirao broj virusa na računalima na jednom učilištu. Dobiveni su sljedeći rezultati, pri čemu x predstavlja broj virusa, a $f(x)$ pripadnu frekvenciju, odnosno broj računala s x virusa u uzorku:

x	$f(x)$
0	5
1	15
2	22
3	23
4	17
5	10
6	5
7	3

Konstruirajte Poissonovu razdiobu za ove podatke, te izračunajte:
a) vjerojatnost da je nasumično izabrano računalo bez virusa,
b) vjerojatnost da je nasumično izabrano računalo zaraženo s više od četiri virusa.

Rješenje:

$$\lambda := \bar{x} = \frac{297}{100} = 2.97\,.$$

Dalje nastavite sami. ∎