Normalna ili Gaussova razdioba

Ako je područje vrijednosti kontinuirane slučajne varijable X čitav skup $\mathbb{R}$, a funkcija gustoće je dana sa

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}, \qquad \sigma,\ m \in \mathbb{R}, \quad \sigma > 0,$$

onda kažemo da X ima **Normalnu razdiobu.** Normalna razdioba je potpuno određena s parametrima σ, m, pa pišemo

$$X : \mathcal{N}\left(m,\ \sigma^2\right).$$

Može se pokazati da je u ovom slučaju

$$\mu = E[X] = m \quad \text{(očekivanje)},$$

$$V[X] = \sigma^2 \quad \text{(varijanca)},$$

$$\sqrt{V[X]} = \sigma \quad \text{(standardna devijacija)}.$$

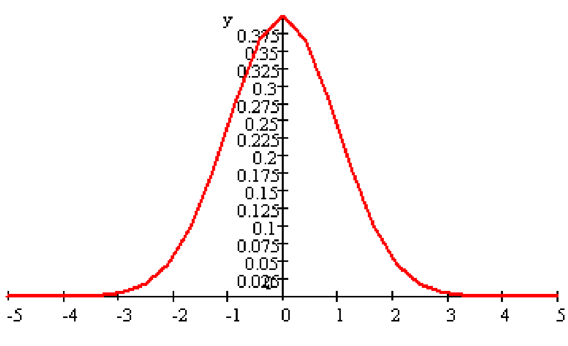
Primjer 2.22 Ako je $X : \mathcal{N}(5,\ 9)$, to znači da je $x \in (-\infty, +\infty)$ i

$$f(x) = \frac{1}{3\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-5}{3}\right)^2}.$$

Ako je $X : \mathcal{N}(0,\ 1)$ kažemo da kontinuirana slučajna varijabla X ima jediničnu Normalnu razdiobu, tj. $x \in (-\infty, +\infty)$ i

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

Ova funkcija je parna pa joj je graf simetričan s obzirom na $y-$os

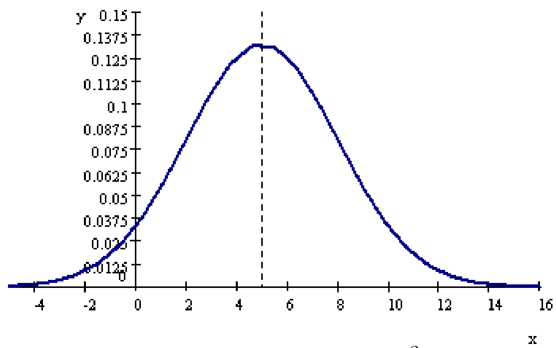


$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

Općenito, ako je $X : \mathcal{N}(m,\ \sigma^2)$ onda je funkcija gustoće

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

simetrična s obzirom na pravac $x = m$ (pomak), a o σ ovisi spljoštenje.



$f(x) = \frac{1}{3\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-5}{3}\right)^2}$

Za $X : \mathcal{N}(m,\ \sigma^2)$ imamo

$$\begin{aligned}
\mathcal{P}\{x_1 < X < x_2\} &= \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx, \\
\mathcal{P}\{X < x_2\} &= \int_{-\infty}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx, \\
\mathcal{P}\{x_1 < X\} &= \int_{x_1}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx.
\end{aligned}$$

Međutim ovi integrali nisu elemetarno rješivi. Ako uvedemo supstituciju

$$u = \frac{x-m}{\sigma},$$

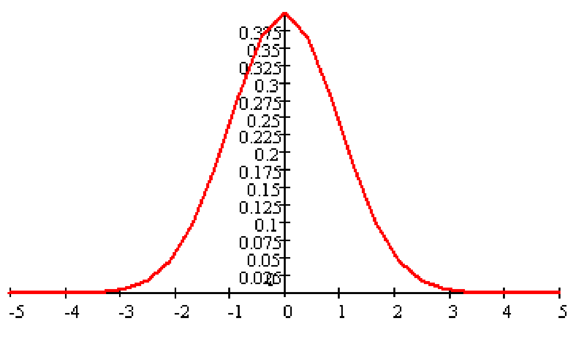
onda je $u_1 = \frac{x_1 - m}{\sigma}$, $u_2 = \frac{x_2 - m}{\sigma}$ i $du = \frac{1}{\sigma} dx$, pa je

$$\begin{aligned}
\mathcal{P}\left\{x_1 < X < x_2\right\} &= \frac{1}{\sqrt{2\pi}} \int\limits_{u_1}^{u_2} e^{-\frac{1}{2}u^2} du = \mathcal{P}\left\{u_1 < U < u_2\right\}, \\
\mathcal{P}\left\{X < x_2\right\} &= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{u_2} e^{-\frac{1}{2}u^2} du = \mathcal{P}\left\{U < u_2\right\}, \\
\mathcal{P}\left\{x_1 < X < x_2\right\} &= \frac{1}{\sqrt{2\pi}} \int\limits_{u_1}^{\infty} e^{-\frac{1}{2}u^2} du = \mathcal{P}\left\{u_1 < U\right\},
\end{aligned}$$

gdje je slučajna varijabla $U : \mathcal{N}(0,\ 1)$.
Funkciju oblika

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int\limits_{0}^{u} e^{-\frac{1}{2}u^2} du, \qquad u > 0,$$

nazivamo **Gaussov integral** (vrijednosti za $\Phi(u)$ nalazimo u tablicama).



$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}, \quad \Phi(u) = \int\limits_0^u \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$

Očito vrijedi: $\Phi(-u) = \Phi(u)$ za svaki $u > 0$.
Koristeći definiciju Gaussovog integrala, sada vjerojatnost da slučajna varijabla $X : \mathcal{N}(m,\ \sigma^2)$ poprimi vrijednost iz bilo kojeg intervala, možemo izraziti pomoću njega.

Uočimo: ako je $U : \mathcal{N}(0,\ 1)$ onda je očito, iz geometrijske interpretacije,

$$\mathcal{P}\left\{0 < U\right\} = \frac{1}{\sqrt{2\pi}} \int\limits_0^{\infty} e^{-\frac{1}{2}u^2} du = 0.5,$$

tj. površina između osi u i grafa funkcije gustoće $y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$ za $u > 0$, je, zbog simetričnosti, jednaka 0.5. Slično, $\mathcal{P}\{U < 0\} = 0.5$.

Primjer 2.23 Imamo:

a)

$$\begin{aligned} \mathcal{P}\{x_1 < X < x_2\} &= \left[u = \frac{x-m}{\sigma}, \;\; U : \mathcal{N}(0,\ 1) \right] = \\ &= \frac{1}{\sqrt{2\pi}} \int_{u_1}^{u_2} e^{-\frac{1}{2}u^2} du = \mathcal{P}\{u_1 < U < u_2\} = (*) \end{aligned}$$

a1) $0 \leq u_1 \leq u_2 \implies (*) = \Phi(u_2) - \Phi(u_1)\,;$

a2) $u_1 \leq 0,\ 0 \leq u_2 \implies (*) = \Phi(-u_1) + \Phi(u_2)\,;$

a3) $u_1 \leq u_2 \leq 0 \implies (*) = \Phi(-u_1) - \Phi(-u_2)\,;$

b)

$$\begin{aligned} \mathcal{P}\{X < x_2\} &= \left[u = \frac{x-m}{\sigma}, \;\; U : \mathcal{N}(0,\ 1) \right] = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u_2} e^{-\frac{1}{2}u^2} du = \mathcal{P}\{U < u_2\} = (**) \end{aligned}$$

b1) $0 \leq u_2 \implies (**) = 0.5 + \Phi(u_1)\,;$

b2) $u_2 \leq 0 \implies (**) = 0.5 - \Phi(-u_1)\,;$

c)

$$\begin{aligned} \mathcal{P}\{x_1 < X\} &= \left[u = \frac{x-m}{\sigma}, \;\; U : \mathcal{N}(0,\ 1) \right] = \\ &= \frac{1}{\sqrt{2\pi}} \int_{u_1}^{\infty} e^{-\frac{1}{2}u^2} du = \mathcal{P}\{u_1 < U\} = (***) \end{aligned}$$

c1) $0 \leq u_1 \implies (***) = 0.5 - \Phi(u_1)\,;$

c2) $u_1 \leq 0 \implies (***) = 0.5 + \Phi(-u_1)$;

Primjer 2.24 Ako je $X : \mathcal{N}(20,\ 16)$ odredite očekivanje $E[X]$, varijancu $V[X]$, te $\mathcal{P}\{X < 15\}$.

Rješenje: Imamo $m = 20$, $\sigma^2 = 16$, pa je $\sigma = 4$, $\mu = E[X] = m = 20$, $V[X] = \sigma^2 = 16$. Odredimo $\mathcal{P}\{X < 15\}$. Uvedimo supstituciju

$$\begin{aligned}
\mathcal{P}\{X < 15\} &= \left[u = \frac{x-20}{4},\ u_2 = \frac{15-20}{4} = -1.25,\ U : \mathcal{N}(0,\ 1)\right] = \\
&= \mathcal{P}\{U < -1.25\} = 0.5 - \Phi(1.25) = 0.5 - 0.3964 = 0.1054.
\end{aligned}$$

Primjer 2.25 Ako je $X : \mathcal{N}(m,\ \sigma^2)$ odredimo:

a) $\mathcal{P}\{\mu - \sigma < X < \mu + \sigma\}$,

b) $\mathcal{P}\{\mu - 2\sigma < X < \mu + 2\sigma\}$,

c) $\mathcal{P}\{\mu - 3\sigma < X < \mu + 3\sigma\}$.

Funkcija gustoće je

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2},$$

očekivanje $\mu = E[X] = m$, varijanca $V[X] = \sigma^2$, a σ standardna devijacija. Supstitucijom $\frac{x-m}{\sigma} = u$, dobivamo

a)

$$\mathcal{P}\{\mu - \sigma < X < \mu + \sigma\} = [U : \mathcal{N}(0,\ 1),\ \mu = m] =$$

$$= \mathcal{P}\left\{\frac{m-\sigma-m}{\sigma} < U < \frac{m+\sigma-m}{\sigma}\right\} = \mathcal{P}\{-1 < U < 1\} =$$

$$= 2\Phi(1) = 2 \cdot 0.34134 = \mathbf{0.68268}$$

b)

$$\mathcal{P}\{\mu - 2\sigma < X < \mu + 2\sigma\} = [U : \mathcal{N}(0,\ 1),\ \mu = m] =$$

$$= \mathcal{P}\left\{\frac{m-2\sigma-m}{\sigma} < U < \frac{m+2\sigma-m}{\sigma}\right\} = \mathcal{P}\{-2 < U < 2\} =$$

$$= 2\Phi(2) = 2 \cdot 0.47725 = \mathbf{0.95450}$$

c)

$$\mathcal{P}\left\{\mu - 3\sigma < X < \mu + 3\sigma\right\} = \left[U : \mathcal{N}\left(0,\ 1\right),\ \mu = m\right] =$$
$$= \mathcal{P}\left\{\frac{m - 3\sigma - m}{\sigma} < U < \frac{m + 3\sigma - m}{\sigma}\right\} = \mathcal{P}\left\{-3 < U < 3\right\} =$$
$$= 2\Phi\left(3\right) = 2 \cdot 0.49865 = \mathbf{0.99730}$$

Komentar: vjerojatnost da će slučajna varijabla $X : \mathcal{N}\left(m,\ \sigma^2\right)$ poprimiti vrijednosti:

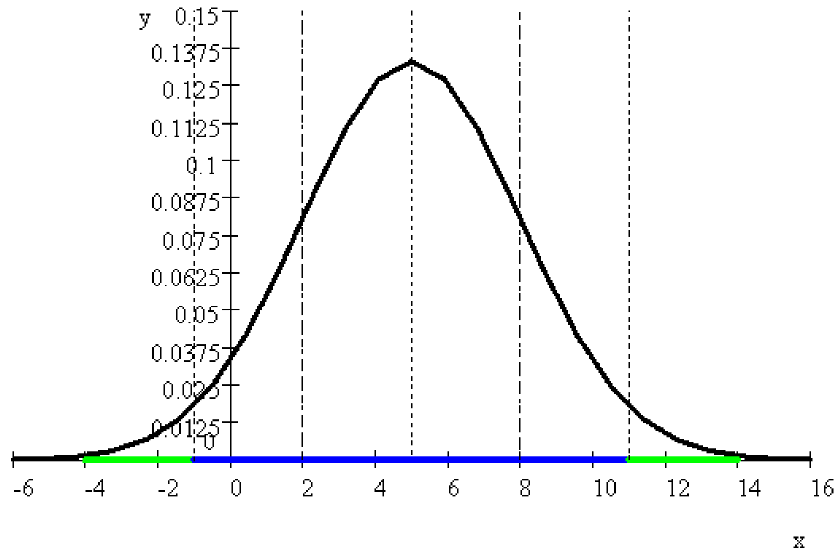
a) iz intervala $(\mu - \sigma,\ \mu + \sigma)$ (interval oko očekivane vrijednosti μ širine 2σ ($\sigma-$stand. dev.)) je **0.68268** ili oko **68.27**% vrijednosti od X pada u interval $\left(\mu - \sigma,\ \mu + \sigma\right)$;

b) iz intervala $(\mu - 2\sigma,\ \mu + 2\sigma)$ (interval oko očekivane vrijednosti μ širine 4σ) je **0.95450** ili oko **95.45**% vrijednosti od X pada u interval $\left(\mu - 2\sigma,\ \mu + 2\sigma\right)$;

c) iz intervala $(\mu - 3\sigma,\ \mu + 3\sigma)$ (interval oko očekivane vrijednosti μ širine 6σ) je **0.99730** ili oko **99.73**% vrijednosti od X pada u interval $\left(\mu - 3\sigma,\ \mu + 3\sigma\right)$.

Svojstva a), b), c) uočio je Gauss.
Ako kod mjerenja nastaju pogreške koje se pokoravaju zakonu Normalne razdiobe, gdje pogreškom smatramo odstupanje od prosjeka (očekivanja) μ, tj. veličinu $x - \mu$, onda su male pogreške (po apsolutnoj vrijednosti), "česte",

a velike "rijetke".



Drugim riječima, pogreške koje su veće od 2σ imaju vjerojatnost manju od 0.05 (5%), a one veće od 3σ imaju vjerojatnost manju od 0.003 (0.3%). Ovo bi značilo da npr. u 1000 mjerenja pogreška (po apsolutnoj vrijednosti) premašuje veličinu 3σ samo u 3 mjerenja. praktički skoro sve pogreške će se nalaziti između -3σ i 3σ. Ova osobina Normalne razdiobe naziva se pravilo tri sigme.

Aproksimacija Binomne razdiobe Normalnom

Neka je $X : \mathcal{B}(n,p)$. Tada je

$$R(X) = \{0, 1, 2, ..., n\},$$

$$p(x) = \binom{n}{x} p^x q^{n-x},$$

$$E[X] = np, \qquad V[X] = npq.$$

Ako je n jako velik $(n >> 0)$ i $p \approx 0.5$, onda je moguće Binomnu razdiobu aproksimirati Normalnom razdiobom $\mathcal{N}(m,\ \sigma^2)$, gdje je $m = np$ i $\sigma^2 = npq$, tj. tada je

$$\mathcal{B}(n,p) \approx \mathcal{N}(np,\ npq).$$

Aproksimacija će biti dobra ako je:

1) $\frac{1}{n+1} < p < \frac{n}{n+1}$,

2) $npq > 9$.

Ako su zadovoljeni uvjeti 1) i 2), onda je

$$p(x) = \binom{n}{x} p^x q^{n-x} \approx \mathcal{P}_{\mathcal{N}}\{x - 0.5 < X < x + 0.5\},$$

$$\mathcal{P}_{\mathcal{B}}\{x_1 \leq X \leq x_2\} = \sum_{x=x_1}^{x_2} \binom{n}{x} p^x q^{n-x} \approx \mathcal{P}_{\mathcal{N}}\{x_1 - 0.5 < X < x_2 + 0.5\},$$

$$\mathcal{P}_{\mathcal{B}}\{x_1 \leq X\} = \sum_{x=x_1}^{n} \binom{n}{x} p^x q^{n-x} \approx \mathcal{P}_{\mathcal{N}}\{x_1 - 0.5 < X\},$$

$$\mathcal{P}_{\mathcal{B}}\{X \leq x_2\} = \sum_{x=0}^{x_2} \binom{n}{x} p^x q^{n-x} \approx \mathcal{P}_{\mathcal{N}}\{X < x_2 + 0.5\}.$$

Primjer 2.26 Neka je $X : \mathcal{B}(100,\ 0.3)$. Odredite:

a) $\mathcal{P}\{20 \leq X \leq 40\}$,

b) $\mathcal{P}\{20 < X < 40\}$,

c) $\mathcal{P}\{20 < X \leq 40\}$,

d) $\mathcal{P}\{30 < X\}$,

e) $\mathcal{P}_{\mathcal{B}}\{X < 25\}$.

<u>Rješenje:</u>

a)

$$\mathcal{P}_{\mathcal{B}}\{20 \leq X \leq 40\} = \mathcal{P}_{\mathcal{B}}\{X = 20\} + \mathcal{P}_{\mathcal{B}}\{X = 21\} + ... + \mathcal{P}_{\mathcal{B}}\{X = 40\} =$$

$$= \binom{100}{20}(0.3)^{20}(0.7)^{100-20} + ... + \binom{100}{40}(0.3)^{40}(0.7)^{100-40} =$$

$$= \sum_{x=20}^{40}\binom{100}{x}(0.3)^{x}(0.7)^{100-x}$$

Vidimo da je ovo vrlo "naporno" za računati. Budući je:

1) $\frac{1}{n+1} = \frac{1}{101} \approx 0.009\,99 < 0.3 < \frac{100}{101} \approx 0.99$,

2) $npq = 100 \cdot 0.3 \cdot 0.7 = 21 > 9$,

onda je moguće Binomnu razdiobu aproksimirati Normalnom razdiobom. Budući je $np = 30$ i $npq = 21$ onda je $\mathcal{B}(100,\ 0.3) \approx \mathcal{N}(30, 21)$, pa imamo:

a)

$$\mathcal{P}_{\mathcal{B}}\left\{20 \leq X \leq 40\right\} \approx \mathcal{P}_{\mathcal{N}}\left\{20 - 0.5 < X < 40 + 0.5\right\} =$$

$$= \mathcal{P}_{\mathcal{N}}\left\{19.5 < X < 40.5\right\} = \left[u = \frac{x-30}{\sqrt{21}},\quad U : \mathcal{N}\left(0,1\right)\right] =$$

$$= \mathcal{P}_{\mathcal{N}}\left\{\frac{19.5-30}{\sqrt{21}} < U < \frac{40.5-30}{\sqrt{21}}\right\} = \mathcal{P}_{\mathcal{N}}\left\{-2.291\,3 < U < 2.291\,3\right\} =$$

$$\approx \mathcal{P}_{\mathcal{N}}\left\{-2.29 < U < 2.29\right\} = 2 \cdot \Phi\left(2.29\right) = 2 \cdot 0.48899 = 0.977\,98.$$

b)

$$\mathcal{P}_{\mathcal{B}}\left\{20 < X < 40\right\} = \mathcal{P}_{\mathcal{B}}\left\{21 \leq X \leq 39\right\} \approx$$

$$\approx \mathcal{P}_{\mathcal{N}}\left\{21 - 0.5 < X < 39 + 0.5\right\} = \left[u = \frac{x-30}{\sqrt{21}},\quad U : \mathcal{N}\left(0,1\right)\right] =$$

$$= \mathcal{P}_{\mathcal{N}}\left\{\frac{20.5-30}{\sqrt{21}} < U < \frac{39.5-30}{\sqrt{21}}\right\} = \mathcal{P}_{\mathcal{N}}\left\{-2.073\,1 < U < 2.073\,1\right\} =$$

$$\approx \mathcal{P}_{\mathcal{N}}\left\{-2.07 < U < 2.07\right\} = 2 \cdot \Phi\left(2.07\right) = 2 \cdot 0.48077 = 0.961\,54.$$

c)

$$\mathcal{P}_{\mathcal{B}}\left\{20 < X \leq 40\right\} = \mathcal{P}_{\mathcal{B}}\left\{21 \leq X \leq 40\right\} \approx$$

$$\approx \mathcal{P}_{\mathcal{N}}\left\{21 - 0.5 < X < 40 + 0.5\right\} = \left[u = \frac{x-30}{\sqrt{21}},\quad U : \mathcal{N}\left(0,1\right)\right] =$$

$$= \mathcal{P}_{\mathcal{N}}\left\{\frac{20.5-30}{\sqrt{21}} < U < \frac{40.5-30}{\sqrt{21}}\right\} = \mathcal{P}_{\mathcal{N}}\left\{-2.073\,1 < U < 2.291\,3\right\} =$$

$$\approx \mathcal{P}_{\mathcal{N}}\left\{-2.07 < U < 2.29\right\} = \Phi\left(2.07\right) + \Phi\left(2.29\right) = 0.48077 + 0.49861 = 0.979\,38.$$

d)

$$\mathcal{P}_{\mathcal{B}}\left\{30 < X\right\} = \mathcal{P}_{\mathcal{B}}\left\{31 \leq X\right\} \approx \mathcal{P}_{\mathcal{N}}\left\{31 - 0.5 < X\right\} =$$

$$= \left[u = \frac{x-30}{\sqrt{21}},\quad U : \mathcal{N}\left(0,1\right)\right] = \mathcal{P}_{\mathcal{N}}\left\{\frac{30.5-30}{\sqrt{21}} < U\right\} = \mathcal{P}_{\mathcal{N}}\left\{0.109\,11 < U\right\} \approx$$

$$= \mathcal{P}_{\mathcal{N}}\left\{0.11 < U\right\} = 0.5 - \Phi\left(0.11\right) = 0.5 - 0.04380 = 0.456\,2$$

e)

$$\mathcal{P}_{\mathcal{B}}\{X<25\}=\mathcal{P}_{\mathcal{B}}\{X\leq 24\}\approx\mathcal{P}_{\mathcal{N}}\{X<24+0.5\}=$$

$$=\left[u=\frac{x-30}{\sqrt{21}},\quad U:\mathcal{N}(0,1)\right]=\mathcal{P}_{\mathcal{N}}\left\{U<\frac{24.5-30}{\sqrt{21}}\right\}=\mathcal{P}_{\mathcal{N}}\{U<-1.200\,2\}\approx$$

$$\approx\mathcal{P}_{\mathcal{N}}\{U<-1.2\}=0.5-\Phi(0.12)=0.5-0.38493=0.115\,07$$

Primjer 2.28 Vjerojatnost rođenja dječaka je približno jednaka 0.515. Kolika je vjerojatnost da među 10000 novorođene djece bude više (ili jednako) djevojčica?

Rješenje: Definirajmo događaj A: rođen je dječak. Onda je događaj $\bar{A}$: nije rođen je dječak, tj. rođena je djevojčica. Sada se, 10000 novorođene djece može predočiti nizom

$$\underbrace{A,\ \bar{A},\ \bar{A},\ A,\ ...,\ A,\ \bar{A}}_{10000}.$$

Definirajmo slučajnu varijablu X : broj rođenih dječaka među 10000 novorođene djece. Tada je $X:\mathcal{B}(10000,\ 0.515)$. Tražena vjerojatnost onda je

$$\mathcal{P}_{\mathcal{B}}\{X<5000\}=\mathcal{P}_{\mathcal{B}}\{X\leq 4999\}=\sum_{x=0}^{4999}\binom{10000}{x}(0.515)^{x}\cdot(0.485)^{10000-x}.$$

Vidimo da je ovo vrlo "naporno" za računati.
Budući je:

1) $\frac{1}{n+1}=\frac{1}{10001}\approx 9.999\times 10^{-5}<0.515<\frac{10000}{10001}\approx 0.999\,9,$
2) $npq=10000\cdot 0.515\cdot 0.485=2497.75>9,$

onda je moguće Binomnu razdiobu aproksimirati Normalnom razdiobom. Budući je $np=5150$ i $npq=2497.75$ onda je

$$\mathcal{B}(10000,\ 0.515)\approx\mathcal{N}(5150,\ 2497.75),$$

pa imamo:

$$\mathcal{P}_{\mathcal{B}}\{X\leq 4999\}\approx\mathcal{P}_{\mathcal{N}}\{X<4999+0.5\}=\mathcal{P}_{\mathcal{N}}\{X<4999.5\}=$$

$$=\left[u=\frac{x-30}{\sqrt{21}},\quad U:\mathcal{N}(0,1)\right]=\mathcal{P}_{\mathcal{N}}\left\{U<\frac{4999.5-5150}{\sqrt{2497.75}}\right\}$$

$$=\mathcal{P}_{\mathcal{N}}\{U<-3.011\,4\}\approx\mathcal{P}_{\mathcal{N}}\{U<-3.01\}=0.5-\Phi(3.01)=0.5-0.49869=0.001\,31.$$

Prilagođavanje teoretske razdiobe empiričkim podacima

Statistički podaci koje dobivamo mjerenjem veličine X (numeričkog statističkog obilježja) čine skup **empiričkih podataka**, a njihova distribucija je **empirička distribucija** (raspodjela frekvencija).

Proučavanjem empiričkih distribucija pokazalo se da postoje zakoni, tj. analitički izrazi, koji, u nekim slučajevima sasvim dobro aproksimiraju dane empiričke podatke. Takav analitički izraz, ako postoji za neku empiričku distribuciju, predstavlja onda **teoretsku distribuciju**.

Prilagoditi teoretsku distribuciju frekvenciju empiričkim podacima zapravo znači naći analitički izraz na osnovi kojeg svakoj numeričkoj vrijednosti statističkog obilježja X, odnosno svakom razredu, s empiričkom frekvencijom f_x, možemo izračunati pripadnu teoretsku frekvenciju f_{tx}. Pri tome ćemo slučajnu varijablu tretirati kao statističko obilježje, a matematičko očekivanje kao aritmetičku sredinu.

Napomena: Za sve empiričke podatke nije moguće naći odgovarajuće teoretske zakone. Mi ćemo pokazati kako se Binomna, Poissonova i Normalna razdioba prilagođavaju empiričkim podacima.

Prilagođavanje Binomne razdiobe

Binomna razdioba $\mathcal{B}(n,\ p)$ je jednoznačno određena parametrima n i p. Zbog toga najprije, iz statističkih podataka, određujemo ta dva parametra. Iz definicije Binomne razdiobe vidimo da je n najveća moguća vrijednost koju može poprimiti slučajna varijabla X $(R(X) = \{0,\ 1, ..., n\})$. Dakle, n je teoretski najveća vrijednost koju može poprimiti diskretno statističko obilježje X, pa se n ne mora podudarati s najvećom vrijednošću za X empiričkih podataka. Budući je očekivanje kod Binomne razdiobe $E[X] = np$, onda, zamjenom $E[X]$ s aritmetičkom sredinom $\bar{x}$, imamo:

$$\bar{x} = np \implies p = \frac{\bar{x}}{n}.$$

Nakon što smo odredili veličine n i p, teoretske frekvencije računamo po formuli

$$f_{tx} = N \cdot p(x),$$

gdje je N ukupan broj statističkih podataka, a $p(x)$ vjerojatnost po Binomnoj razdiobi $\mathcal{B}(n,\ p)$, tj.

$$p(x) = \binom{n}{x} p^x q^{n-x}, \qquad q = 1 - p.$$

Primjer 2.29 Kontrolor kontrolira proizvodnju tako što u jednakim vremenskim razmacima uzima uzorak od 20 proizvoda i zabilježi broj neispravnih proizvoda u uzorku. Nakon 100 kontrola dobio je sljedeće podatke:

broj neispravinih proizvoda u uzorku x	0	1	2	3	4	5	6
frekvencija f_x	14	25	27	23	7	3	1

Prilagodite Binomnu razdiobu empiričkim podacima i nađite teoretske frekvencije. Za koju vrijednost $x-$a imamo najveće odstupanje teoretske od empiričke frekvencije.

Rješenje: Definirajmo događaj A: proizvod je neispravan. Onda je događaj $\bar{A}$: proizvod nije neispravan. Sada se, uzorak od 20 proizvoda može se predočiti nizom

$$\underbrace{A,\ \bar{A},\ \bar{A},\ A,\ ...,\ A,\ \bar{A}}_{20}.$$

Definirajmo slučajnu varijablu X : broj neispravnih proizvoda u uzorku. Sada je teoretski

$$R(X) = \{0, 1, 2, ..., 20\} \Longrightarrow n = 20.$$

Treba nam $\bar{x} = \frac{1}{N} \sum\limits_{x=0}^{6} x f_x$.

x	f_x	xf_x	$p(x)$	$f_{tx} = 100 \cdot p(x)$
0	14	0	0.1216	12.16
1	25	25	0.2701	27.01
2	27	54	0.2852	28.52
3	23	69	0.1901	19.01
4	7	28	0.0898	8.98
5	3	15	0.0319	3.19
6	1	6	0.0089	0.89
	$N = \sum = 100$	$\sum = 197$	$\sum = 0.9976$	$\sum = 99.76$

Sada je, uz identifikaciju $E[X] = np \equiv \bar{x}$, imamo

$$\bar{x} = \frac{1}{N}\sum_{x=0}^{6} x f_x = \frac{197}{100} = 1.97 \quad \Longrightarrow \quad p = \frac{\bar{x}}{n} = \frac{1.97}{20} = 0.098\,5 \approx 0.1$$

Dakle, $X : \mathcal{B}(20,\ 0.1)$, pa vjerojatnosti $p(x)$ imamo u tablicama ili ih računamo kao

$$p(x) = \binom{20}{x}(0.1)^x (0.9)^{n-x}, \qquad q = 1 - 0.1 = 0.9.$$

za $x = 0, 1, ..., 6$, ili pomoću rekurzivne formule.
Za vrijednost $x = 3$ imamo najveće odstupanje

$$\Delta_3 = |f_3 - f_{t3}| = 3.99$$

teoretske od empiričke frekvencije. Ovdje je

$$\begin{aligned} E[X] &= np = 1.97, \\ V[X] &= npq = 1.97 \cdot 0.9 = 1.\,773. \end{aligned}$$

Prilagođavanje Poissonove razdiobe

Želimo li nekom empiričkom skupu podataka, za koji naslučujemo da se pokorava Poissonovom zakonu, prilagoditi teoretsku Poissonovu razdiobu, potrebno je odrediti veličinu m (jer je Poissonova razdioba $\mathcal{P}_m$ jednoznačno određena s parametrom m).
Budući je za $X : \mathcal{P}_m$ očekivanje $E[X] = m$, onda za m uzimamo aritmetičku sredinu $\bar{x}$, tj.

$$m = \bar{x}.$$

Teoretske frekvencije računamo po formuli

$$f_{tx} = N \cdot p(x),$$

gdje je N ukupan broj podataka, a $p(x)$ vjerojatnost po Poissonovoj razdiobi $\mathcal{P}_m$, tj.

$$p(x) = \frac{m^x}{x!}e^{-m}.$$

Primjer 2.30 Empiričkim podacima zadanim u tablici

x	0	1	2	3	4	5
f_x	59	39	14	6	2	1

prilagodite Poissonovu razdiobu. Kolika je očekivana vrijednost $E[x]$ i varijanca $V[x]$? Izračunajte pripadne teoretske frekvencije, te vjerojatnost da x poprimi vrijednost veću od 1.

Rješenje:

x	f_x	xf_x	$p(x)$	$f_{tx} = 121 \cdot p(x)$
0	59	0	0.44933	54. 37
1	39	39	0.35946	43. 49
2	14	28	0.14379	17. 40
3	6	18	0.03834	4. 64
4	2	8	0.00767	0.93
5	1	5	0.00123	0.15
	$N = \sum = 121$	$\sum = 98$		

$$\bar{x} = \frac{\sum x_i f_x}{N} = \frac{98}{121} \approx 0.809\,92 \approx 0.8 = m$$

$$X : \mathcal{P}_{0.8} \Longrightarrow E[x] = V[x] = m = 0.8$$

Dakle, $X : \mathcal{P}_{0.8}$, pa vjerojatnosti $p(x)$ imamo u tablicama ili ih računamo kao

$$p(x) = \frac{(0.8)^x}{x!} e^{-0.8},$$

za $x = 0, 1, ..., 5$ (ili pomoću rekurzivne formule). Sada je

$$E[x] = V[x] = m = 0.8,$$

te

$$\mathcal{P}\{x > 2\} = 1 - \mathcal{P}\{x = 0\} - \mathcal{P}\{x = 1\} = 0.191\,21.$$