4 Slučajne varijable

4.1 Osnovni pojmovi

Mnogi eksperimenti čiji su ishodi "slučajni" podvrgavaju se istim zakonitostima. Cilj nam je definirati neke osnovne zakonitosti. Osnovni aparat koji koristimo su **slučajne varijable**.

Primjer 49
1. *Ponavljamo bacanje novčića $n-$puta uzastopce. Zanima nas broj padanja pisama.*
2. *Ponavljamo neki pokus $n-$puta u istim uvjetima. U svakom ponavljanju možemo imati uspješan rezultat i neuspješan rezultat pokusa sa određenom vjerojatnošću. Zanima nas broj uspješnih pokusa.*
3. *Zanima nas broj (učestalost) poziva nekoj telefonskoj centrali u nekoj jedinici vremena (dan, mjesec,...).*
4. *Zanima nas učestalost nesreća u nekom gradu u nekoj jedinici vremena.*
5. *Vrijeme do prvog poziva.*
6. *Proteklo vrijeme do prve nesreće (od nekog fiksnog trenutka).*
7. *Zanima nas distribucija (raspodjela, razdioba) visine u nekoj populaciji (žene, muškarci, studenti, u Zagrebu, Hrvatskoj...)*
8. *Isto za težinu, tlak,...*

Definicija 10 *Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnostni prostor. Svaku funkciju $X : \Omega \to \mathbf{R}$ nazivat ćemo* **slučajnom varijablom**.

Mogućnost da se neke metode iz klasične analize (kalkulusa) prenesu na izučavanje slučajnih varijabli daju nam pojmovi **funkcije distribucije** i **funkcije gustoće** slučajnih varijabli.

Definicija 11 *Neka je* $X : \Omega \to \mathbf{R}$ *slučajna varijabla. Funkciju* $F_X : \mathbf{R} \to \mathbf{R}$ *definiranu sa*

$$F_X(x) \overset{def}{=} P\left(\{\omega \in \Omega\ : X(\omega) \leq x\}\right) \overset{oznaka}{=} P\left(X \leq x\right)$$

nazivat ćemo funkcijom distribucije slučajne varijable X*.*

Primjer 50 *Bacamo 3 ispravna (pravilna) novčića. Slučajna varijabla* X *registrira broj palih pisama. Odredite sliku* $\mathrm{Im}X$*, funkciju distribucije* F_X*, te* $\Gamma(F_X)$ *slučajne varijable* X*.*

Rješenje:

$$\Omega = \{(PPP), (PPG), (PGP), (PGG), (GPP), (GPG), (GGP), (GGG)\}.$$

Očito je

$$\mathrm{Im}X = \{0, 1, 2, 3\}.$$

$$P(X = 0) = \frac{1}{8},\ P(X = 1) = \frac{3}{8},\ P(X = 2) = \frac{3}{8},\ P(X = 3) = \frac{1}{8}.$$

$$x < 0 : F_X(x) = P(X \leq x) = P(\emptyset) = 0,$$

$$0 \leq x < 1 :\ F_X(x) = P(X \leq x) = P(X = 0) = \frac{1}{8},$$

$$1 \leq x < 2 :\ F_X(x) = P(X \leq x) = P(X = 0) + P(X = 1) = \frac{4}{8} = \frac{1}{2},$$

$$2 \leq x < 3 :\ F_X(x) = P(X \leq x) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{7}{8},$$

$$3 \leq x :\ F_X(x) = P(X \leq x)$$

$$= P(\Omega) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = \frac{8}{8} = 1.$$

$\square$

Svojstva funkcije distribucije: Neka je X proizvoljna slučajna varijabla i F_X njezina funkcija distribucije.

1. F_X je rastuća funkcija.

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

3. $F_X(-\infty) = 0$, $F_X(+\infty) = 1$.

4.2 Diskretne slučajne varijable

Navedeni primjer spada u posebnu skupinu slučajnih varijabli koju nazivamo diskretnim slučajnim varijablama kod kojih je $\mathrm{Im}X$ diskretan (najčešće konačan) skup. Ako slučajna varijabla X može poprimiti bilo koju vrijednost iz nekog intervala (visina, težina, vrijeme do prvog poziva itd.) nazivat ćemo je neprekidnom.

U gornjem primjeru smo vidjeli da nakon što smo odredili $\mathrm{Im}X$, te $P(X = x_i)$, $x_i \in \mathrm{Im}X$ kod računanja daljnih vjerojatnosti ne koristimo konkretan kontekst tj. vjerojatnostni prostor.

Općenito ako za neku slučajnu varijablu X odredimo

a) $\mathrm{Im}X = \{a_1, a_2, ..., a_i, ...\}$

b) $p_i = P(X = a_i)$, $i = 1, 2, ...$ td. je $\sum_{i=1}^{\infty} p_i = 1$

govorit ćemo da je time određena razdioba (distribucija) slučajne varijable X.

Zapis:

$$X \sim \begin{pmatrix} a_1 & a_2 & ... & a_i & ... \\ p_1 & p_2 & ... & p_i & ... \end{pmatrix}.$$

Definicija 12 *Neka je $X : \Omega \to \mathbb{R}$ slučajna varijabla.* ***Funkcija gustoće vjerojatnosti*** *od X je funkcija $p_X :\ \mathrm{Im}X \to [0, 1]$ definirana s*

$$p_X(a_i) := P(X = a_i) = p_i$$

Definicija 13 ***Matematičko očekivanje*** *diskretne slučajne varijable je broj $\mathrm{E}[X]$ definiran s*

$$\mathrm{E}[X] = \sum_{a_i \in \mathrm{Im}X} a_i \cdot p_X(a_i)$$

Vrijedi:

(i) $\mathrm{E}[g(X)] = \sum_{a_i \in \mathrm{Im}X} g(a_i) \cdot p_X(a_i), \quad g : \mathbb{R} \to \mathbb{R}$

(ii) $\mathrm{E}[X + Y] = \mathrm{E}[X] + \mathrm{E}[Y]$

(iii) $\mathrm{E}[cX] = c\,\mathrm{E}[X], \quad c \in \mathbb{R}$

(iv) $\mathrm{E}[c] = c, \quad c \in \mathbb{R}$

Iz svojstva (ii) vidimo da je očekivanje aditivno, iz svojstva (iii) da je homogeno. Ta dva svojstva zajedno daju svojstvo linearnosti.

Definicija 14 *Broj*

$$\mathrm{Var}[X] := \sum_{a_i \in \mathrm{Im} X} (a_i - \mathrm{E}\, X)^2 \cdot p_X(a_i)$$

zove se ***varijanca*** *diskretne slučajne varijable X.*
Standardna devijacija *slučajne varijable X je broj*

$$\sigma_X := +\sqrt{\mathrm{Var}[X]}$$

Vrijedi:

(i) $\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}\, X)^2]$

(primijenimo svojstvo (i) od očekivanja za $g(x) = (x - \mathrm{E}\, X)^2$)

(ii) $$\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}\, X)^2] = \mathrm{E}[X^2 - 2X \cdot \mathrm{E}\, X + (\mathrm{E}\, X)^2]$$
$$= \mathrm{E}[X^2] - 2\,\mathrm{E}\, X \cdot \mathrm{E}\, X + (\mathrm{E}\, X)^2 = \mathrm{E}[X^2] - (\mathrm{E}\, X)^2$$
$$\Longrightarrow \mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}\, X)^2$$

(iii) $$\mathrm{Var}[aX + b] = \mathrm{E}[(aX + b - \mathrm{E}\,(aX + b))^2] = \mathrm{E}[(aX + b - a\,\mathrm{E}\, X - b)^2]$$
$$= \mathrm{E}[a^2 (X - \mathrm{E}\, X)^2] = a^2\,\mathrm{E}[(X - \mathrm{E}\, X)^2] = a^2\,\mathrm{Var}\, X$$

Čebiševljeva nejednakost: Neka je X diskretna slučajna varijabla, $\varepsilon > 0$, $\mu = E[X]$. Imamo:

$$\mathrm{Var}[X] = E[(X - \mu)^2] = \sum_{x_k \in \mathrm{Im} X} (x_k - \mu)^2 P(X = x_k)$$
$$= \sum_{|x_k - \mu| \geq \varepsilon} (x_k - \mu)^2 P(X = x_k) + \sum_{|x_k - \mu| < \varepsilon} (x_k - \mu)^2 P(X = x_k)$$
$$\geq \sum_{|x_k - \mu| \geq \varepsilon} (x_k - \mu)^2 P(X = x_k) \geq \sum_{|x_k - \mu| \geq \varepsilon} \varepsilon^2 P(X = x_k)$$
$$= \varepsilon^2 \sum_{|x_k - \mu| \geq \varepsilon} P(X = x_k) = \varepsilon^2 P(|X - \mu| \geq \varepsilon).$$

Konačno se dobije:
$$P(|X-\mu| \geq \varepsilon) \leq \frac{1}{\varepsilon^2}\mathrm{Var}[X].$$
Dobije se i:
$$P(|X-\mu| < \varepsilon) = 1 - P(|X-\mu| \geq \varepsilon) \geq 1 - \frac{1}{\varepsilon^2}\mathrm{Var}[X].$$

Pravilo 3-sigma: Neka je $\sigma(X) = \sqrt{\mathrm{Var}[X]}$ standardna devijacija slučajne varijable X. Iz Čebiševljeve nejednakosti dobijemo za $\varepsilon = 3\sigma(X)$:
$$P(|X-\mu| < 3\sigma(X)) \geq 1 - \frac{1}{9\mathrm{Var}[X]}\mathrm{Var}[X] = \frac{8}{9} = 0.89$$
što znači da se za svaku slučajnu varijablu barem 89% podataka nalazi u intervalu $(\mu - 3\sigma(X), \mu + 3\sigma(X))$.

4.3 Osnovni primjeri diskretnih slučajnih varijabli

1. Binomna slučajna varijabla
2. Poissonova slučajna varijabla
3. Hipergeometrijska slučajna varijabla

4.3.1 Binomna i Bernoullijeva slučajna varijabla

Primjer 51
1. *Izvlačimo 10 puta po jednu kuglicu uzastopce s vraćanjem iz kutije u kojoj su 2 bijele i 5 crvenih kuglica. "Uspjeh" u pojedinom izvlačenju je izvlačenje bijele kuglice. Koja je vjerojatnost da smo izvukli k bijelih kuglica,* $0 \leq k \leq 10$*?*
2. *Bacamo 100 puta simetričnu kocku. "Uspjeh" (u pojedinom bacanju) je padanje šestice. Koja je vjerojatnost da je palo k šestica,* $0 \leq k \leq 100$*?*
3. *Isto sa bacanjem novčića. "Uspjeh" je npr. padanje pisma.*
4. *Neki košarkaš pogađa koš u* 70% *slučajeva. Koja je vjerojatnost da će u* 50 *bacanja pogoditi* $k-$*puta,* $0 \leq k \leq 50$*?*

5. *Poznato je da u 80% slučajeva iz posijanog sjemena nikne biljka. Kolika je vjerojatnost da od* 10 *posijanih sjemenki nikne k biljki,* $0 \leq k \leq 10$*?*

Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnostni prostor. Kažemo da se dogodio "uspjeh " ako se dogodio neki unaprijed odabrani događaj $A \subseteq \Omega$.

Neka je X slučajna varijabla definirana sa

$$X : \Omega \to \mathbf{R}; \;\; X(\omega) \stackrel{def}{=} \begin{cases} 1; & \omega \in A \\ 0; & \omega \in A^c \end{cases}$$

Očito je distribucija slučajne varijable X dana sa

$$X = \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$$

pri čemu je $P(X = 1) = P(A) = p$, $P(X = 0) = P(A^c) = 1 - p \stackrel{oznaka}{=} q$. Tako definiranu slučajnu varijablu nazivamo **Bernoullijevom slučajnom varijablom**. Očito je:

$$E[X] = 0{\cdot}q + 1{\cdot}p = p, \;\; \mathrm{Var}[X] = E[X^2] - E^2[X] = E[X] - E^2[X] = p - p^2 = pq.$$

Pretpostavimo sada da navedeni slučajni pokus ponavljamo n puta ali tako da ishod pojedinog pokusa ne utječe na ishode preostalih pokusa (npr. vađenje kuglica iz kutije sa vraćanjem i bez vraćanja). Neka slučajna varijabla X registrira broj "uspjeha" u navedenih n nezavisnih pokusa. Odredimo distribuciju slučajne varijable X. Očito je

$$\mathrm{Im}X = \{0, 1, 2, \ldots, n\}.$$

Imamo: $P(X = 0) = (1-p)^n = q^n$, $P(X = 1) = p(1-p)^{n-1} + (1-p)p(1-p)^{n-2} + \cdots + (1-p)^{n-1}p = npq^{n-1}$. Analogno:

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

Provjera:

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

Slučajnu varijablu za koju je

$$\mathrm{Im}X = \{0, 1, 2, \ldots, n\},$$

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

nazivat ćemo **binomnom slučajnom varijablom**.

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

Primjer 52 *Poznato je da u 80% slučajeva iz posijanog sjemena nikne biljka. Posijemo 5 sjemenki.*

1. *Kolika je vjerojatnost da će proklijati barem 4 sjemenke?*
2. *Kolika je vjerojatnost da neće proklijati više od 3 sjemenke?*
3. *Kolika je vjerojatnost da će proklijati barem dvije, ali ne više od 4 sjemenke?*

Rješenje: X...broj proklijalih sjemenki među pet posijanih; "uspjeh" je kada jedna sjemenka proklije; p = vjerojatnost usjeha = 80% = 0.8. Očito je

$$X \sim B(5, 0.8).$$

Dobije se

$$X = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 0.00032 & 0.0064 & 0.0512 & 0.2048 & 0.4096 & 0.32768 \end{pmatrix}.$$

1. $P(X \geq 4) = P(X = 4) + P(X = 5) = 0.73768$
2. $P(X \leq 3) = 1 - P(X > 3) = 1 - P(X \geq 4) = 0.26272.$
3. $P(2 \leq X \leq 4) = P(X = 2) + P(X = 3) + P(X = 4) = 0.6656.$

$\square$

Odredimo očekivanje i varijancu binomne slučajne varijable $X \sim B(n, p)$ korištenjem Bernoullijevih slučajnih varijabli.

Označimo sa X_i Bernoullijevu slučajnu varijablu koja registrira rezultat $i-$tog pokusa tj. jednaka je 1 ako se u $i-$tom pokusu dogodio "uspjeh", a 0 ako se u $i-$tom pokusu nije dogodio "uspjeh", $i = 1, 2, \ldots, n$. Kazat ćemo da je zadan (konačan) niz nezavisnih jednako distribuiranih Bernoullijevih slučajnih varijabli. Očito je

$$X = X_1 + X_2 + \cdots + X_{n-1} + X_n.$$

Sada iz svojstava matematičkog očekivanja i varijance dobijemo:

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np, \ \mathrm{Var}[X] = \sum_{i=1}^{n} \mathrm{Var}[X_i] = npq.$$

Primjer 53 *Odredite $E[X]$, $\mathrm{Var}[X]$ u Primjeru 52. Izračunajte $P(|X-\mu| < 3\sigma(X))$, pri čemu je $\mu = E[X]$, $\sigma(X) = \sqrt{\mathrm{Var}[X]}$ standardna devijacija slučajne varijable X.*

Rješenje: $\mu = E[X] = 5 \cdot 0.8 = 4$, $\mathrm{Var}[X] = 5 \cdot 0.8 \cdot 0.2 = 0.8$, $\sigma(X) = \sqrt{\mathrm{Var}[X]} = \sqrt{0.8} = 0.894$.

$$P(|X-\mu| < 3\sigma(X)) = P(|X-4| < 2.683) = P(4 - 2.683 < X < 4 + 2.683)$$

$$= P(1.317 < X < 6.683) = 1 - P(X=0) - P(X=1) = 0.993.$$

$\square$

Zadatak 4 *Košarkaš gađa koš 5 puta i u svakom pokušaju pogađa s vjerojatnošću 3/4. Kolika je vjerojatnost da će košarkaš pogoditi koš:*
a) točno 3 puta
b) barem 3 puta
c) najviše 2 puta

Rješenje: Pokus je gađanje u koš; ponavljamo ga 5 puta. "Uspjeh" je pogodak u koš, "neuspjeh" je promašaj. Vjerojatnost "uspjeha" je zadana i jednaka je 3/4; vjerojatnost "neuspjeha" je tada 1/4 (=1-3/4). Definiramo slučajnu varijablu X koja broji pogotke. Ona ima binomnu distribuciju:

$$X \sim B\left(5, \ \frac{3}{4}\right)$$

Funkcija gustoće od X je:

$$p_X(k) = P(X = k) = \binom{5}{k} \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{5-k}, \quad k = 0, 1, 2, 3, 4, 5 \tag{10}$$

a) Želimo izračunati vjerojatnost da je košarkaš pogodio koš točno 3 puta. Zanima nas zapravo $P(X = 3)$. Uvrštavanjem $k = 3$ u (10) lako dobijemo rješenje:

$$P(X = 3) = \binom{5}{3} \left(\frac{3}{4}\right)^3 \left(\frac{1}{4}\right)^2 = \binom{5}{3} \frac{3^3}{4^5} = 0.26$$

b) Kolika je vjerojatnost da košarkaš pogodi koš barem 3 puta? Taj događaj izražen pomoću slučajne varijable X je $X \geq 3$. Želimo dakle izračunati:

$$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5)$$

$P(X = 3)$ smo već izračunali pod a). $P(X = 4)$ i $P(X = 5)$ računamo na sličan način - uvrštavanjem $k = 4$ odnosno $k = 5$ u (10). Slijedi:

$$P(X = 4) = \binom{5}{4} \left(\frac{3}{4}\right)^4 \left(\frac{1}{4}\right)^1 = 5 \cdot \frac{3^4}{4^5} = 0.395$$

$$P(X = 5) = \binom{5}{5} \left(\frac{3}{4}\right)^5 \left(\frac{1}{4}\right)^0 = \left(\frac{3}{4}\right)^5 = 0.24$$

$$\Longrightarrow P(X \geq 3) = 0.26 + 0.395 + 0.24 = 0.895$$

c) Kolika je vjerojatnost da košarkaš pogodi koš najviše 2 puta? Taj događaj izražen pomoću slučajne varijable X je $X \leq 2$ a to je zapravo suprotan događaj događaju kojeg smo promatrali pod b) pa vrijedi:

$$P(X \leq 2) = 1 - P(X \geq 3) = 1 - 0.895 = 0.105$$

Naravno, vjerojatnost tog događaja mogla bi se računati i direktno:

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

$\square$

Zadatak 5 *Odredite očekivanje, varijancu i devijaciju slučajne varijable* $X \sim B\left(5, \frac{3}{4}\right)$. *Konstruirajte interval* $\mathrm{E}[X] \pm 2\sigma_X$ *te izračunajte* $P(\mathrm{E}[X] - 2\sigma_X < X < \mathrm{E}[X] + 2\sigma_X)$

Rješenje:

$$\begin{aligned}
&\mathrm{E}[X] = 5 \cdot \frac{3}{4} = 3.75 \\
&\mathrm{Var}[X] = 5 \cdot \frac{3}{4} \cdot \frac{1}{4} = 0.9375 \\
&\sigma_X = \sqrt{0.9375} = 0.968
\end{aligned}$$

Traženi interval je:

$$3.75 \pm 2 \cdot 0.968 = 3.75 \pm 1.936 \quad \Rightarrow \quad 1.814 < X < 5.686$$

Nadalje,

$$\begin{aligned}
&P(1.814 < X < 5.686) = P(2 \leq X \leq 5) \\
&= P(X=2) + P(X=3) + P(X=4) + P(X=5)
\end{aligned}$$

Posljednje 3 vjerojatnosti već smo izračunali u prethodnom zadatku. Fali nam još:

$$P(X=2) = \binom{5}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^3 = 10 \cdot \frac{3^2}{4^5} = \frac{45}{512} = 0.088$$

Sada dobivamo:

$$\begin{aligned}
&P(\mathrm{E}[X] - 2\sigma_X < X < \mathrm{E}[X] + 2\sigma_X) = P(1.814 < X < 5.686) \\
&= 0.088 + 0.26 + 0.395 + 0.24 = 0.983
\end{aligned}$$

Odavde možemo zaključiti da će ova slučajna varijabla u 98.3% slučajeva odstupati od svog očekivanja za najviše 2 devijacije. □

Slabi zakon velikih brojeva daje teorijsko opravdanje intuitivno prihvatljive tvrdnje da se relativne frekvencije pojave uspjeha u uzastopnom ponavljanju pokusa približavaju vjerojatnosti uspjeha u pojedinom pokusu.

Slabi zakon velikih brojeva: Prisjetimo se Čebiševljeve nejednakosti:

$$P(|X - \mu| \geq \varepsilon) \leq \frac{1}{\varepsilon^2}\mathrm{Var}[X].$$

Neka je zadan niz slučajnih varijabli (Y_n) sa:

$$Y_n = X_1 + X_2 + \cdots + X_n$$

gdje su $X_i,\ i = 1, 2, ...$ nezavisne jednako distribuirane Bernoullijeve slučajne varijable. Primjetimo da je $Y_n \sim B(n, p)$.

Primjenimo Čebiševljevu nejednakost na slučajne varijable

$$X = \frac{Y_n}{n} \text{...relativna frekvencija "uspjeha".}$$

Kako je

$$E\left[\frac{Y_n}{n}\right] = \frac{1}{n}E[Y_n] = \frac{1}{n}np = p$$

i

$$\text{Var}\left[\frac{Y_n}{n}\right] = \frac{1}{n^2}\text{Var}[Y_n] = \frac{1}{n^2}npq = \frac{pq}{n},$$

to iz Čebiševljeve nejednakosti slijedi:

$$P\left(\left|\frac{Y_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\frac{pq}{n} = \frac{1}{n\varepsilon^2}p(1-p) \leq \frac{1}{4n\varepsilon^2}$$

odnosno

$$P\left(\left|\frac{Y_n}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{1}{4n\varepsilon^2}.$$

Primjetimo da odavde slijedi

$$\lim_{n\to\infty} P\left(\left|\frac{Y_n}{n} - p\right| < \varepsilon\right) = 1.$$

Primjer 54 *Sumnjamo da novčić nije simetričan. Koliko puta treba bacati novčić da bi sa* 95% *sigurnosti procijenili vjerojatost padanja pisma na zaokruženoj drugoj decimali?*

Rješenje:

$$X_n \ldots \text{ukupan broj pisama u } n\text{−bacanja,}$$

$$p \ldots \text{vjerojatnost pisma u pojedinom bacanju}$$

$$X_n \sim B(n, p).$$

Procjena za p je relativna frekvencija $\frac{X_n}{n}$. Iz uvjeta zadatka mora biti ispunjeno

$$P\left(\left|\frac{X_n}{n} - p\right| < 0.5 \cdot 10^{-2}\right) \geq 0.95.$$

Kako iz Čebiševljeve nejednakosti slijedi

$$P\left(\left|\frac{X_n}{n}-p\right|<0.5\cdot 10^{-2}\right)\geq 1-\frac{1}{4n0.25\cdot 10^{-4}}=1-\frac{10^4}{n}$$

dovoljno je naći n tako da je

$$1-\frac{10^4}{n}\geq 0.95$$

što lako daje $n\geq 200000$. $\square$

-odrediti mod binomne slučajne varijable

-primjeri (sa stupčastim dijagramima?) za $p\approx 0.5$, $p\approx 0$, $p\approx 1$.

4.3.2 Poissonova slučajna varijabla

Definicija 15 *Slučajna varijabla X ima* ***Poissonovu razdiobu*** *ili* ***distribuciju*** *s parametrom $\lambda>0$ ako je funkcija gustoće te slučajne varijable zadana s:*

$$p_X(k)=P(X=k)=\frac{\lambda^k}{k!}\,e^{-\lambda},\quad k=0,1,2,3,\ldots \tag{11}$$

Poissonova distribucija daje model vjerojatnosti "rijetkih" događaja (ponekad se naziva i "zakon rijetkih događaja") koji se događaju u jedinici vremena, površine, volumena i slično. Broj prometnih nesreća na određenoj dionici autoceste u jednom danu, telefonski pozivi na centrali u jednoj minuti, defekti po jedinici duljine bakrene žice, broj mjesečnih nesreća u tvornici, broj oboljelih stabala po aru šume te broj vidljivih grešaka na dijamantu su npr. varijable čije se relativne frekvencije mogu dobro aproksimirati Poissonovom distribucijom.

Neka je zadan niz binomnih slučajnih varijabli $X_n\sim B(n,p_n)$ pri čemu je $n\cdot p_n=\lambda$ (primijetite da to povlači da je $\lim_{n\to\infty}p_n=0$). Imamo

$$P(X_n=k)=\binom{n}{k}p_n^k(1-p_n)^{n-k}=\frac{n(n-1)\cdots(n-k+1)}{k!}\frac{p_n^k(1-p_n)^n}{(1-p_n)^k}$$

$$\frac{n(n-1)\cdots(n-k+1)}{k!}\frac{\left(\frac{\lambda}{n}\right)^k(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k}$$

$$= \frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{-k}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^{n},$$

što očito daje

$$\lim_{n\to\infty} P(X_n = k) = \frac{\lambda^k}{k!}e^{-\lambda}.$$

Kako je $E[X_n] = np_n$ i $\mathrm{Var}[X_n] = np_n(1-p_n)$ i kako je

$$\lim_{n\to\infty} E[X_n] = \lim_{n\to\infty} \lambda = \lambda, \quad \lim_{n\to\infty} \mathrm{Var}[X_n] = \lim_{n\to\infty} np_n(1-p_n) = \lambda$$

to zaključujemo, ako je $X \sim P(\lambda)$ to je

$$E[X] = \lambda \quad \mathrm{Var}[X] = \lambda.$$

Osnovna svojstva koja opisuju Poissonovu distribuciju:

1. Pokus se sastoji od prebrojavanja koliko puta (k) se neki događaj dogodi u jedinici vremena, jedinici površine, volumena, težine, daljine ili bilo kojoj drugoj mjerenoj jedinici.

2. Vjerojatnost da će se događaj kojeg promatramo dogoditi jednaka je za svaku mjernu jedinicu (za svaku sekundu, svaki metar, svaki karat i sl.).

3. Broj događaja koji se dogode u pojedinoj jedinici vremena, površine ili volumena nezavisan je od broja događaja koji se dogodi u bilo kojoj drugoj jedinici.

4. Prosječni ili očekivani broj događaja u jednoj jedinici jednak je parametru λ, odnosno $\mathrm{E}[X] = \lambda$.

Zadatak 6 *Dokažite da je* $\mathrm{E}[X] = \lambda$ *neposredno, samo iz distribucija Poissonove varijable.*

Rješenje: Koristeći definiciju očekivanja, funkciju gustoće Poissonove razdiobe zadanu s (11) te svojstvo

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

dobivamo:

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{k=0}^{\infty} k \cdot p_X(k) = \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k}{k!}\, e^{-\lambda} \\
&= \lambda \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}\, e^{-\lambda} = \lambda \cdot \sum_{m=0}^{\infty} \frac{\lambda^m}{m!}\, e^{-\lambda} = \lambda \cdot 1 = \lambda
\end{aligned}$$

$\square$

Zadatak 7 *Slučajna varijabla X ima Poissonovu razdiobu. Ako vrijedi*

$$P(X=1) = P(X=2),$$

izračunajte očekivanje $\mathrm{E}[X]$ *i* $P(X \geq 4)$.

Rješenje:

$$\begin{aligned}
&X \sim P(\lambda) \\
&p_X(k) = P(X=k) = \frac{\lambda^k}{k!}\, e^{-\lambda}, \quad k = 0, 1, 2, 3, \ldots \\
&P(X=1) = P(X=2) \;\Rightarrow\; \frac{\lambda^1}{1!}\, e^{-\lambda} = \frac{\lambda^2}{2!}\, e^{-\lambda} \;\Rightarrow\; \lambda = \frac{\lambda^2}{2} \;\Rightarrow\; \lambda^2 - 2\lambda = 0 \\
&\Rightarrow \lambda(\lambda - 2) = 0 \;\Rightarrow \lambda_1 = 0,\; \lambda_2 = 2
\end{aligned}$$

Budući mora biti $\lambda > 0$, jedino rješenje je $\lambda = 2$.

$$\Rightarrow X \sim P(2), \quad P(X=k) = \frac{2^k}{k!}\, e^{-2}$$

$$\begin{aligned}
&\mathrm{E}[X] = \lambda = 2 \\
&P(X \geq 4) = 1 - P(X < 4) = 1 - P(X \leq 3) \\
&\quad = 1 - P(X=0) - P(X=1) - P(X=2) - P(X=3) \\
&\quad = 1 - \frac{2^0}{0!}\, e^{-2} - \frac{2^1}{1!}\, e^{-2} - \frac{2^2}{2!}\, e^{-2} - \frac{2^3}{3!}\, e^{-2} \\
&\quad = 1 - e^{-2}\left(1 + 2 + 2 + \frac{4}{3}\right) = 1 - e^{-2} \cdot \frac{19}{3} = 0.143
\end{aligned}$$

$\square$

Zadatak 8 *Pretpostavimo da je 220 grešaka raspoređeno slučajno unutar knjige od 200 stranica. Odredite vjerojatnost da dana stranica knjige sadrži:*
a) niti jednu grešku
b) točno jednu grešku
c) barem dvije greške

Rješenje: Definirajmo slučajnu varijablu X koja broji greške na pojedinoj stranici. Ona ima Poissonovu distribuciju. Kako bi odredili njenu funkciju gustoće, potreban nam je parametar λ. Znamo da je taj parametar jednak očekivanom ili prosječnom broju događaja (= broj grešaka) koji se dogode u jednoj jedinici (= na jednoj stranici). Stoga

$$\lambda = \frac{220}{200} = 1.1$$
$$p_X(k) = P(X = k) = \frac{\lambda^k}{k!}\, e^{-\lambda} = \frac{(1.1)^k}{k!}\, e^{-1.1}, \quad k = 0, 1, 2, \ldots \tag{12}$$

Pomoću ovako definirane slučajne varijable, događaj pod a) možemo zapisati kao $X = 0$, događaj pod b) kao $X = 1$, a događaj pod c) kao $X \geq 2$. Vjerojatnosti tih događaja računamo uvrštavanjem odgovarajućih k u (12). Dobivamo:

$$a) \quad P(X = 0) = \frac{(1.1)^0}{0!}\, e^{-1.1} = e^{-1.1} = 0.333$$
$$b) \quad P(X = 1) = \frac{(1.1)^1}{1!}\, e^{-1.1} = 0.366$$
$$c) \quad P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - 0.333 - 0.366 = 0.301$$

□

Prethodni zadatak lijepo ilustrira zašto se Poissonova distribucija naziva i "zakon rijetkih događaja". Događaji da na stranici nema niti jedne greške ($X = 0$) i da je na stranici točno jedna greška ($X = 1$) - dakle "rijetki" događaji (u smislu malog broja grešaka) - imaju veću vjerojatnost nego događaj da su stranici 2 ili 3 ili 4 ili 5 ili ... ili n ili ... grešaka ($X \geq 2$).

Zadatak 9 *London je podijeljen na* $N = 576$ *područja površine* 0.25km^2. *Znamo da je na London palo* 537 *raketa (V1 i V2). Ako je pad raketa* *"slučajan" tj. da nije bilo gađanja u određene ciljeve, odredite teorijske frekvencije broja područja pogođenih sa točno* 0, 1, 2, 3 *rakete.*

Rješenje: Prosječan broj raketa po području je

$$\lambda = \frac{537}{576} = 0.9323.$$

Dijeleći područja na n "mikrovolumena" (kao kod planktona) broj palih raketa X na slučajno odabrano područje je binomna slučajna varijabla

$$X \sim B(n, \frac{0.9323}{n}) \approx Y \sim P(0.9323).$$

Neka je N_k označava broj područja sa k palih raketa. Dobije se

$$N_0 = NP(X=0) \approx NP(Y=0) = 576 \cdot e^{-0.9323} = 226.74,$$

$$N_1 = NP(X=1) \approx NP(Y=1) = 576 \cdot 0.9323e^{-0.9323} = 211.39,$$

$$N_2 = NP(X=2) \approx NP(Y=2) = 576 \cdot 0.9323^2 e^{-0.9323}/2 = 98.54.$$

□

Stabilnost binomne i Poissonove slučajne varijable:

Neka su X_1 i X_2 nezavisne slučajne varijable.

1. Ako je $X_1 \sim B(n_1, p)$, $X_2 \sim B(n_2, p)$, onda je $X_1 + X_2 \sim B(n_1 + n_2, p)$.
2. Ako je $X_1 \sim P(\lambda_1)$, $X_2 \sim P(\lambda_2)$, onda je $X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$.

Pokažimo drugu tvrdnju. Očito je

$$\mathrm{Im}(X_1 + X_2) = \{0, 1, 2, \ldots k, \ldots\}.$$

$$P(X_1 + X_2 = k) = \sum_{i=0}^{k} P(X_1 = i, X_2 = k - i) = \sum_{i=0}^{k} \frac{\lambda_1^i e^{-\lambda_1}}{i!} \frac{\lambda_2^{k-i} e^{-\lambda_2}}{(k-i)!}$$

$$= e^{-(\lambda_1+\lambda_2)} \sum_{i=0}^{k} \frac{1}{k!} \frac{k!}{i!(k-i)!} \lambda_1^i \lambda_2^{k-i} = e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1 + \lambda_2)^k}{k!}.$$

Navedena dva svojstva pokazuju da je $X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$.

4.3.3 Hipergeometrijska slučajna varijabla

Definicija 16 *Slučajna varijabla X ima* ***hipergeometrijsku razdiobu*** *ili* ***distribuciju*** *ako je funkcija gustoće te slučajne varijable zadana s:*

$$p_X(k) = P(X = k) = \frac{\binom{m}{k}\binom{n}{s-k}}{\binom{m+n}{s}}, \quad \max(0, s-n) \leq k \leq \min(m, s) \qquad (13)$$

Pišemo $X \sim H(n, m, s)$.

- Očekivanje hipergeometrijske razdiobe: $\mathrm{E}[X] = \dfrac{s \cdot m}{m+n}$
- Varijanca hipergeometrijske razdiobe: $\mathrm{Var}[X] = \dfrac{m \cdot n \cdot s \cdot (m+n-s)}{(m+n)^2(m+n-1)}$
- Osnovna svojstva koja opisuju hipergeometrijsku distribuciju:
  1. Pokus se sastoji od slučajnog izvlačenja, bez vraćanja, s elemenata iz skupa od $m+n$ elemenata, od kojih je njih m jedne vrste (izvlačenje takvog smatramo uspjehom) i n neke druge vrste (izvlačenje takvog smatramo neuspjehom).
  2. Hipergeometrijska slučajna varijabla broji broj "uspjeha" (odnosno elemenata prve vrste) k u izvlačenju ukupno n elemenata.

Primjetimo da za $\frac{m}{m+n} \approx p$ i $m+n \approx \infty$ dobijemo

$$E[X] \approx (m+n)p, \;\; \mathrm{Var}[X] \approx (m+n)p(1-p),$$

kao kod binomne razdiobe. Pogledajmo još i sljedeći limes:

$$P(X = k) = \frac{\binom{m}{k}\binom{n}{s-k}}{\binom{m+n}{s}} = \frac{\frac{m!}{k!(m-k)!}\frac{n!}{(s-k)!(n-s+k)!}}{\frac{(m+n)!}{(m+n-s)!s!}}$$

$$= \frac{s!}{k!(s-k)!} \cdot$$

$$\frac{m(m-1)\cdots(m-k+1)\cdot n(n-1)\cdots(n-s+k+1)}{(m+n)(m+n-1)\cdots(m+n-s+1)},$$

što dijeljenjem brojnika i nazivnika sa $(m+n)^s$ na limesu daje

$$P(X = k) \approx \binom{s}{k} p^k (1-p)^{s-k}, \;\; p \approx \frac{m}{m+n}.$$

Zadatak 10 *Kolika je vjerojatnost da se od 7 suglasnika i 5 samoglasnika napravi riječ koja se sastoji od 4 suglasnika i 3 samoglasnika? (riječ ne mora imati smisao)*

Rješenje: Od ukupno 7+5=12 slova želimo izabrati 4+3=7 slova. Neka je izvlačenje samoglasnika "uspjeh". Slučajna varijabla X koja broji "uspjehe" ima hipergeometrijsku razdiobu a funkcija gustoće joj je zadana s:

$$p_X(k) = P(X = k) = \frac{\binom{5}{k}\binom{7}{7-k}}{\binom{12}{7}}, \quad 0 \le k \le 5$$

Događaj da su izabrana 4 suglasnika i 3 samoglasnika pomoću slučajne varijable X možemo izraziti kao $X = 3$. Sada:

$$p_X(3) = P(X = 3) = \frac{\binom{5}{3}\binom{7}{4}}{\binom{12}{7}} = \frac{175}{396} = 0.442$$

$\square$

Zadatak 11 *Iz vaze koja sadrži 4 crvene i 6 bijelih ruža izvlačimo 3 ruže. S X označimo slučajnu varijablu koja broji izvučene crvene ruže. Odredite njen zakon razdiobe, te prosječan broj izvučenih crvenih ruža.*

Rješenje: X = broj crvenih ruža
Od ukupno 10=4+6 ruža izvlačimo 3, a izvlačenje crvene ruže smatramo uspjehom. Tada X ima hipergeometrijsku distribuciju a funkcija gustoće joj je:

$$p_X(k) = P(X = k) = \frac{\binom{4}{k}\binom{6}{3-k}}{\binom{10}{3}}, \quad 0 \le k \le 3$$

Dakle, $\mathrm{Im}X = \{0, 1, 2, 3\}$. Trebaju nam pripadne vjerojatnosti. Imamo

$$p_X(0) = P(X = 0) = \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}} = \frac{6 \cdot 5 \cdot 4}{10 \cdot 9 \cdot 8} = \frac{1}{6}$$

$$p_X(1) = P(X = 1) = \frac{\binom{4}{1}\binom{6}{2}}{\binom{10}{3}} = \frac{4 \cdot \frac{6 \cdot 5}{2}}{\frac{10 \cdot 9 \cdot 8}{3 \cdot 2}} = \frac{1}{2}$$

$$p_X(2) = P(X = 2) = \frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}} = \frac{\frac{4 \cdot 3}{2} \cdot 6}{\frac{10 \cdot 9 \cdot 8}{3 \cdot 2}} = \frac{3}{10}$$

$$p_X(3) = P(X=3) = \frac{\binom{4}{3}\binom{6}{0}}{\binom{10}{3}} = \frac{4 \cdot 1}{\frac{10 \cdot 9 \cdot 8}{3 \cdot 2}} = \frac{1}{30}$$

$$\Rightarrow X \sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1/6 & 1/2 & 3/10 & 1/30 \end{pmatrix}$$

Provjera da smo dobro računali:

$$\frac{1}{6} + \frac{1}{2} + \frac{3}{10} + \frac{1}{30} = 1$$

Izračunajmo sada $\mathrm{E}[X]$ (što je zapravo prosječan broj):

$$\mathrm{E}[X] = \sum_{k=0}^{3} a_k \cdot p_X(k) = \sum_{k=0}^{3} k \cdot p_X(k) = 0 \cdot \frac{1}{6} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{3}{10} + 3 \cdot \frac{1}{30} = \frac{6}{5} = 1.2$$

Umjesto pomoću definicije, očekivanje smo mogli izračunati i pomoću gore navedene formule:

$$\mathrm{E}[X] = \frac{s \cdot m}{m+n} = \frac{3 \cdot 4}{4+6} = \frac{12}{10} = 1.2$$

$\square$

4.4 Neprekidne slučajne varijable

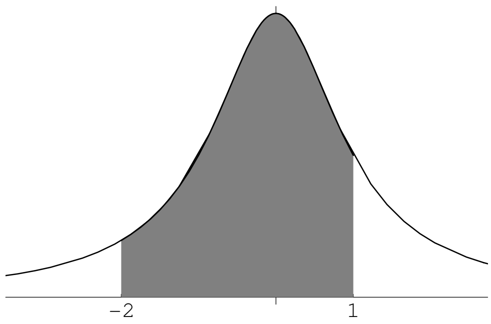
Za slučajnu varijablu X kažemo da je **neprekidna** ako vrijedi sljedeće:

(i) $\mathrm{Im}X$ je interval u $\mathbb{R}$

(ii) postoji nenegativna funkcija $f_X : \mathbb{R} \to \mathbb{R}$ tako da za svaka dva broja a, b $(a < b)$ vrijedi

$$P(a \leq X \leq b) = \int_a^b f_X(t)dt$$

Funkciju f_X zovemo **funkcija gustoće** od X. Vjerojatnost da slučajna varijabla X upadne u interval $[a, b]$ jednaka je dakle površini ispod grafa funkcije gustoće na tom intervalu. Ako je na slici prikazana funkcija gustoće od X, tada je $P(-2 \leq X \leq 1)$ jednaka sljedećoj površini:



Funkcija distribucije F_X od X definirana je s:

$$F_X(x) := P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$$

Vrijedi:

$$P(a \leq X \leq b) = F_X(b) - F_X(a)$$

Navedimo još dva svojstva neprekidne slučajne varijable:

(1) Za svaki broj $a \in \mathbb{R}$ je

$$P(X = a) = \lim_{b \to a} P(a \leq X \leq b) = \lim_{b \to a} \int_a^b f_X(t)dt = \int_a^a f_X(t)dt = 0$$

(2)

$$\int_{-\infty}^{\infty} f_X(t)dt = P(-\infty < X < \infty) = 1$$

što znači da je ukupna površina ispod grafa funkcije gustoće jednaka 1.

Matematičko očekivanje od X definirano je s:

$$\mathrm{E}\left[X\right] = \int_{-\infty}^{\infty} t \cdot f_X(t)dt,$$

a za **varijancu** vrijedi relacija kao i kod diskretnih slučajnih varijabli

$$\mathrm{Var}\left[X\right] = \mathrm{E}\left[X^2\right] - (\mathrm{E}\left[X\right])^2$$

gdje je sada

$$\mathrm{E}\left[X^2\right] = \int_{-\infty}^{\infty} t^2 \cdot f_X(t)dt.$$

Općenito, za $g : \mathbb{R} \to \mathbb{R}$ vrijedi

$$\mathrm{E}\left[g(X)\right] = \int_{-\infty}^{\infty} g(t) \cdot f_X(t) dt.$$

4.4.1 Normalna slučajna varijabla

Definicija 17 *Kažemo da neprekidna slučajna varijabla X ima* ***normalnu razdiobu*** *s parametrima μ i σ^2 ako joj je funkcija gustoće zadana s:*

$$f_X(t) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(t-\mu)^2}{2\sigma^2}} \tag{14}$$

Oznaka:

$$X \sim N(\mu, \sigma^2)$$

Vrijedi:

1. $f_X(t) > 0, \ \ \forall t \in \mathbb{R} \ \ \Rightarrow \ \mathrm{Im} X = \mathbb{R}$
2. $\mathrm{E}\left[X\right] = \displaystyle\int_{-\infty}^{\infty} t \cdot \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(t-\mu)^2}{2\sigma^2}} = \mu \qquad \left(\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}\right)$
3. $\mathrm{Var}\left[X\right] = \mathrm{E}\left[(X-\mu)^2\right] = \displaystyle\int_{-\infty}^{\infty} (t-\mu)^2 \cdot \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(t-\mu)^2}{2\sigma^2}} = \sigma^2$

Iz 2. i 3. vidimo da parametri μ i σ^2 zapravo predstavljaju očekivanje, odnosno varijancu od X.

Normalna razdioba je invarijantna na afine transformacije, tj. ako je

$$X \sim N(\mu, \sigma^2) \quad \text{i} \quad a, b \in \mathbb{R},\ a \neq 0$$

tada je

$$Y = aX + b \sim N(a\mu + b,\ a^2\sigma^2)$$

Zato svakoj normalno distribuiranoj slučajnoj varijabli $X \sim N(\mu, \sigma^2)$ možemo pridružiti **standardiziranu slučajnu varijablu**

$$X^* := \frac{X - \mathrm{E}\left[X\right]}{\sigma_X} = \frac{X-\mu}{\sigma} \sim N(0,1)$$

koja je također normalno distribuirana ali s parametrima 0 i 1.

Funkciju distribucije jedinične normalne razdiobe $N(0,1)$ označavamo s $\Phi(x)$ i vrijedi:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt, \quad x \in \mathbb{R}$$

Funkcija koju ćemo koristiti prilikom rješavanja zadataka i čije vrijednosti su tabelirane je

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{t^2}{2}} dt, \quad x > 0$$

Veza među funkcijama Φ i Φ_0:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{t^2}{2}} dt + \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{t^2}{2}} dt = \frac{1}{2} + \Phi_0(x),$$

uz dogovor

$$\Phi_0(x) = -\Phi_0(-x) \quad \text{za} \quad x < 0.$$

Ako su X i Y nezavisne slučajne varijable, pri čemu je $X \sim N(\mu_X, \sigma_X^2)$ i $Y \sim N(\mu_Y, \sigma_Y^2)$, onda je

$$X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2).$$

Pravilo $3-$sigma za normalnu razdiobu:

$$P(|X-\mu| \le k\sigma) = P(\mu - k\sigma \le X \le \mu + k\sigma) = \int_{\mu-k\sigma}^{\mu+k\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= \int_{-k}^{k} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 2 \int_{0}^{k} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 2\Phi_0(k).$$

Uvidom u tablice dobije se:

$$k = 1,\ P(|X-\mu| \le \sigma) = 0.6826,$$

$$k = 2,\ P(|X-\mu| \le 2\sigma) = 0.9544,$$

$$k = 3,\ P(|X-\mu| \le 3\sigma) = 0.9974.$$

Zadatak 12 *Neka je zadana slučajna varijabla $X \sim N(0,1)$. Odredite vjerojatnosti događaja: a) $0 \le X \le 1$, b) $-1 \le X \le 2$, c) $X \le 1$, d) $X \ge 1$*

Rješenje:

$$
\begin{aligned}
a) \quad & P(0 \le X \le 1) = \Phi(1) - \Phi(0) = \frac{1}{2} + \Phi_0(1) - \frac{1}{2} - \Phi_0(0) \\
& = \Phi_0(1) - \Phi_0(0) = 0.3413 - 0 = 0.3413 \\
b) \quad & P(-1 \le X \le 2) = \Phi(2) - \Phi(-1) = \frac{1}{2} + \Phi_0(2) - \frac{1}{2} - \Phi_0(-1) \\
& = \Phi_0(2) + \Phi_0(1) = 0.4772 + 0.3413 = 0.8185 \\
c) \quad & P(X \le 1) = \Phi(1) = \frac{1}{2} + \Phi_0(1) = 0.5 + 0.3413 = 0.8413 \\
d) \quad & P(X \ge 1) = 1 - P(X < 1) = 1 - P(X \le 1) = 1 - (1/2 + \Phi_0(1)) = 0.1587
\end{aligned}
$$

$\square$

Zadatak 13 *Neka je zadana slučajna varijabla* $X \sim N(2,4)$. *Odredite vjerojatnosti događaja:* $a)$ $0 \le X \le 4$, $\quad b)$ $X \ge 4$

Rješenje:

$$
\begin{aligned}
& X \sim N(2,4) \quad \Rightarrow \quad X^* = \frac{X-\mu}{\sigma} = \frac{X-2}{2} \sim N(0,1) \\
a) \quad & P(0 \le X \le 4) = F_X(4) - F_X(0) \\
& = P\left(\frac{0-2}{2} \le \frac{X-2}{2} \le \frac{4-2}{2}\right) = P(-1 \le X^* \le 1) \\
& = \Phi_0(1) - \Phi_0(-1) = 2\Phi_0(1) = 2 \cdot 0.3413 = 0.6826 \\
b) \quad & P(X \ge 4) = 1 - P(X < 4) = 1 - P(X \le 4) = 1 - F_X(4) \\
& = 1 - P\left(X^* \le \frac{4-2}{2}\right) = 1 - \Phi(1) = 1 - \left(\frac{1}{2} + \Phi_0(1)\right) \\
& = \frac{1}{2} - \Phi_0(1) = 0.5 - 0.3413 = 0.1587
\end{aligned}
$$

$\square$

Zadatak 14 *Slučajna varijabla* X *mjeri odstupanje aviona od sredine dozvoljenog koridora. Ona je normalno distribuirana, s očekivanjem 100m i standardnom devijacijom 200m. Ako je avion upravljen da leti sredinom koridora, nađite vjerojatnost da:*

a) avion leti kroz koridor širine 500m

b) iznad tog koridora.

Rješenje: Slučajna varijabla X mjeri odstupanje aviona od sredine koridora. Vrijedi: $X \sim N(100, 200^2)$

a) Ako želimo da avion leti sredinom koridora širine 500m, tada on od sredine tog koridora može odstupati najviše 250m prema gore ili prema dole pa imamo:

$$
\begin{aligned}
P(-250 \leq X \leq 250) &= P\left(\frac{-250-100}{200} \leq X^* \leq \frac{250-100}{200}\right) \\
&= \Phi\left(\frac{3}{4}\right) - \Phi\left(-\frac{7}{4}\right) = \Phi_0\left(\frac{3}{4}\right) + \Phi_0\left(\frac{7}{4}\right) \\
&= 0.2734 + 0.4599 = 0.7333
\end{aligned}
$$

b) Ako je avion iznad koridora, tada je $X \geq 250$.

$$
\begin{aligned}
P(X \geq 250) &= 1 - P(X < 250) = 1 - P(X \leq 250) = \\
&= 1 - \Phi\left(\frac{250-100}{200}\right) = 1 - \frac{1}{2} - \Phi_0(0.75) = 0.5 - 0.2734 = 0.2266
\end{aligned}
$$

$\square$

Teorem 4 (**Centralni granični teorem**) *Neka je* (X_k) *niz nezavisnih jednako distribuiranih slučajnih varijabli konačne varijance. Neka je*

$$Z_n = X_1 + X_2 + \cdots + X_n, \quad n \in \mathbf{N}.$$

Tada vrijedi:

$$\lim_{n\to\infty} P\left(a \leq \frac{Z_n - E[Z_n]}{\sqrt{\mathrm{Var}[Z_n]}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt.$$

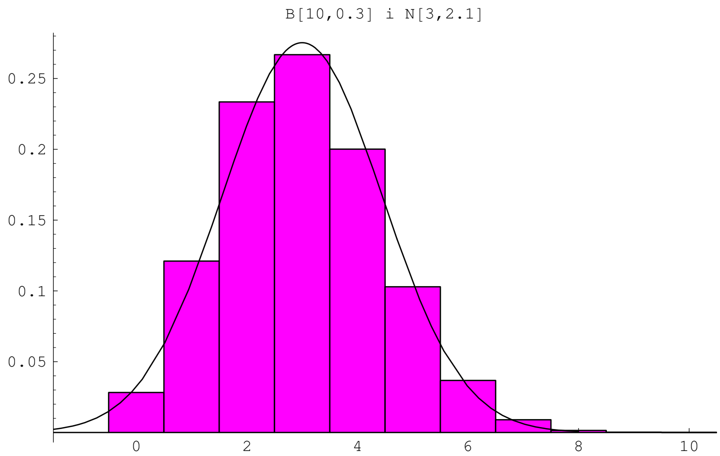
Primjetimo da su slučajne varijable

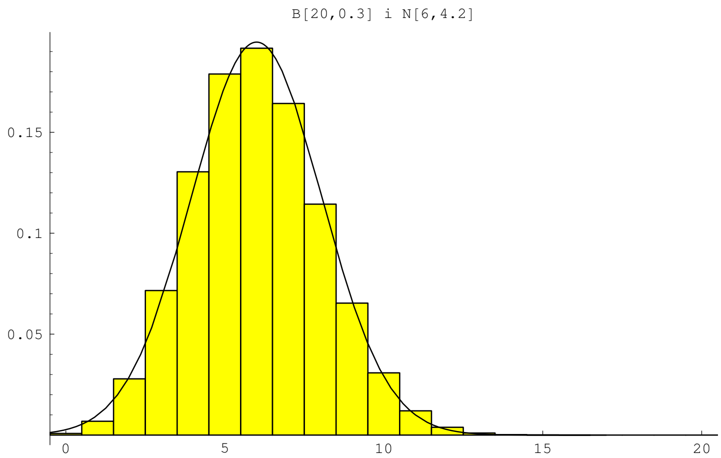
$$Z_n^* = \frac{Z_n - E[Z_n]}{\sqrt{\mathrm{Var}[Z_n]}}$$

standardizirane tj. $E[Z_n^*] = 0$, $\mathrm{Var}[Z_n^*] = 1$, te da se s desne strane jednakosti nalazi $\Phi(b) - \Phi(a)$ gdje je Φ funkcija distribucije jediniñe normalne slučajne varijable. Iz centralnog graničnog teorema slijedi da za "velike" n vrijedi aproksimacija

$$P\left(a \leq \frac{Z_n - E[Z_n]}{\sqrt{\mathrm{Var}[Z_n]}} \leq b\right) \approx \Phi(b) - \Phi(a).$$

Najčešći način upotrebe Centralnog graničnog teorema je kada su X_1, X_2, ... Bernoullijeve slučajne varijable tj. Z_n binomne slučajne varijable.





Zadatak 15 *Vjerojatnost uspjeha u svakom od* 900 *nezavisnih pokusa je* $p = 0.8$*. Izračunajte vjerojatnost da je ukupan broj uspjeha a) jednak* 750 *b) jednak* 710 *c) barem* 710 *i ne više od* 740*.*

Rješenje: Neka X registrira broj uspjeha u 900 nezavisnih ponavljanja. Očito je $X \sim B(900, 0.8)$, pa je

$$E[X] = 900 \cdot 0.8 = 720, \quad \mathrm{Var}[X] = 900 \cdot 0.8 \cdot 0.2 = 144.$$

Označimo sa X^* standardiziranu slučajnu varijablu X tj.

$$X^* = \frac{X - E[X]}{\sigma_X} = \frac{X - 720}{12}.$$

a)

$$P(X = 750) = P(749.5 < X < 750.5)$$
$$= P\left(\frac{749.5 - 720}{12} < X^* < \frac{750.5 - 720}{12}\right) = P(2.46 < X^* < 2.54)$$
$$\approx \Phi(2.54) - \Phi(2.46) = 0.014.$$

b)

$$P(X = 710) = P(709.5 < X < 710.5)$$
$$= P\left(\frac{709.5 - 720}{12} < X^* < \frac{710.5 - 720}{12}\right) = P(-0.875 < X^* < -0.792)$$
$$\approx \Phi(-0.792) - \Phi(-0.875) = \Phi_0(0.875) - \Phi_0(0.792) = 0.0254.$$

c)

$$P(710 \leq X \leq 740) = P(709.5 < X < 740.5)$$
$$= P\left(\frac{709.5 - 720}{12} < X^* < \frac{740.5 - 720}{12}\right) = P(-0.875 < X^* < 1.71)$$
$$\approx \Phi(1.71) - \Phi(-0.875) = \Phi_0(1.71) + \Phi_0(0.875) = 0.767.$$

$\square$

Primjer 55 *Sumnjamo da novčić nije simetričan. Koliko puta treba bacati novčić da bi sa 95% sigurnosti procijenili vjerojatost padanja pisma na zaokruženoj drugoj decimali?*

Rješenje:

$$X_n \ldots \text{ukupan broj pisama u } n \text{ bacanja.}$$

$$p \ldots \text{vjerojatnost padanja pisma.}$$

Očito je

$$X_n \sim B(n, p).$$

Procjena za p je relativna frekvencija

$$\frac{X_n}{n}.$$

Iz uvjeta zadatka imamo

$$P\left(\left|\frac{X_n}{n} - p\right| \leq 0.5 \cdot 10^{-2}\right) \geq 0.95.$$

Za n dovoljno veliki vrijedi:

$$P\left(\left|\frac{X_n - np}{\sqrt{npq}}\right| \leq \varepsilon\right) \approx \Phi(\varepsilon) - \Phi(-\varepsilon) = 2\Phi_0(\varepsilon).$$

Odavde je

$$2\Phi_0(\varepsilon) \approx P(|X_n - np| \leq \sqrt{npq}\varepsilon) = P\left(\left|\frac{X_n}{n} - p\right| \leq \frac{\varepsilon}{\sqrt{n}}\sqrt{p(1-p)}\right)$$

$$\leq P\left(\left|\frac{X_n}{n} - p\right| \leq \frac{\varepsilon}{2\sqrt{n}}\right).$$

Stavljajući

$$\frac{\varepsilon}{2\sqrt{n}} = 0.005$$

odnosno

$$\varepsilon = 0.01\sqrt{n}$$

n ćemo odrediti iz nejednakosti

$$0.95 \leq 2\Phi_0(0.01\sqrt{n}).$$

Uvidom u tablice dobije se

$$1.96 \leq 0.01\sqrt{n}, \ \ n \geq 196^2 = 38416.$$

$\square$