Verjetnost :

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

$\binom{n}{1} = n$ in $\binom{n}{0} = 1$

①

Distributivnost : $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

Popoln sistem dogodkov :

$A_i \neq N$, $A_i \cap A_j = N$ $i \neq j$, $\bigcup_{i=1}^{n} A_i = G$

De Morganovi pravili : $\overline{A \cup B} = \bar{A} \cap \bar{B}$

$\overline{A \cap B} = \bar{A} \cup \bar{B}$

- **Verjetnost dogodka :** $P(A) = \frac{k}{n}$ – klasična definicija

št. ugodnih izidov

št. vseh izidov

$P(A) = \frac{m(A)}{m(S)}$ – geometrijska definicija

neka mera

relativna frekvenca dogodka

$f_n(A) = \frac{k}{n} \to P(A)$ – statistična definicija

veliko št. ponovitev (Bernoullijev zakon stabilnosti)

- Verjetnost je preslikava $P : S \to \mathbb{R}$, ki ustreza naslednjim aksiomom: – aksiomatična definicija

① $P(A) \geq 0$ za vsak $A \in S$

② $P(G) = 1$

③ za A in B $(A \cap B = N)$ velja: $P(A \cup B) = P(A) + P(B)$

nezdružljivost

Posledice :

① $P(\bar{A}) = 1 - P(A)$

② $P(N) = 0$

③ $A \subseteq B \rightarrow P(A) \leq P(B)$

④ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

⑤ $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P$

$P(A \cap \bar{B}) = P(A) - P(A \cap B)$

- **Pogojna verjetnost :** $P(A|B) = \frac{P(A \cap B)}{P(B)}$ → Verjetnost, da se je zgodil A pri pogoju, da se je zgodil B.

če $P(B) \neq 0$, potem: $P(A \cap B) = P(B)\, P(A|B)$

če $P(A) \neq 0$, potem : $P(A \cap B) = P(A)\, P(B|A)$

- **Neodvisnost dogodkov**: $P(A|B) = P(A)$ oz. $P(B|A) = P(B)$

Dogodka A in B sta neodvisna natanko tedaj, ko velja produktna formula: $\underline{P(A \cap B) = P(A)P(B)}$

↳ Nezdružljiva pa sta, če: $P(A \cap B) = 0$. Ni isto!!!

- **Popolna verjetnost**: $H_1, \dots, H_n$ - popoln sistem dogodkov → $H_i \neq N$, $H_i \cap H_j = N$ $i \neq j$

$$\bigcup_{i=1}^{n} H_i = G$$

$$\underline{P(A) = \sum_{i=1}^{n} P(H_i)\, P(A|H_i)}$$

→ Uporabimo pri dvofaznen poskusu.

- **Bayesov obrazec**:

$$P(H_k|A) = \frac{\overset{\text{produktna formula}}{P(H_k)\, P(A|H_k)}}{\underbrace{\sum_{i=1}^{} P(H_i)\, P(A|H_i)}_{\text{formula o popolni verjetnosti}}}$$

Denimo, da se je dogodek A zgodil. Zanima nas, kolikšna je verjetnost, da se je pred tem zgodila hipoteza H_k?

---

diskretna (točke) / zvezna (interval)

Določena je z: zalogo vrednosti (Z_X) in porazdelitvenim zakonom.

- **Slučajna spremenljivka X**: Funkcija, ki vsakemu izidu poskusa priredi neko realna število.

↳ $X: G \to \mathbb{R}$ in $g: \mathbb{R} \to \mathbb{R}$ → potem tudi $Y = g(X): G \to \mathbb{R}$ slučajna spremenljivka!

- **Porazdelitvena funkcija**: $\underline{F_X: \mathbb{R} \to [0,1]}$ in $\underline{F_X(x) = P(X \leq x)}$

Lastnosti:
1. $0 \leq F_X(x) \leq 1$
2. $\lim\limits_{x \to -\infty} F_X(x) = 0$ in $\lim\limits_{x \to \infty} F_X(x) = 1$
3. $P(x_1 < X < x_2) = F_X(x_2) - F_X(x_1)$
4. $x_1 < x_2 \rightarrow F_X(x_1) \overset{\text{monotono naraščajoča}}{\leq} F_X(x_2)$
5. $\lim\limits_{\substack{h \to 0 \\ h > 0}} F_X(x+h) = F_X(x)$ ↘ z desne zvezna (lahko skoki)

- **Verjetnostna funkcija** (za diskretne X): $\underline{p_k = P(X = x_k),\ k = 1, 2, \dots}$ → $\underline{p_k \geq 0 \text{ in } \sum_{k=1}^{\infty} p_k \overset{\text{včasih}}{=} 1}$

Geometrijska vrsta:
$$\sum_{n=0}^{\infty} a \cdot q^n = \frac{a}{1-q}$$

↳ Verjetnostna shema: $X: \begin{pmatrix} x_1 & x_2 & \cdots & x_k & \cdots \\ p_1 & p_2 & \cdots & p_k & \cdots \end{pmatrix}$

če $x_i < x_{i+1}$, potem: $F_X(x_k) = P(X \leq x_k) = \sum_{i=1}^{k} p_i$

- **Matematično upanje**: $\underline{E(X) = \sum_{i=1}^{\infty} x_i p_i}$ (ne obstaja vedno!) - "uteženo povprečje" vrednosti (pričakovana vrednost)

Lastnosti:
1. $E(\overset{Y}{g(X)}) = \sum_{i=1}^{\infty} \overset{y_i}{g(x_i)}\, p_i$
2. $E(aX + b) = aE(X) + b$
3. $E(X + Y) = E(X) + E(Y)$
4. $E((X - E(X))^2) = E(X^2) - E(X)^2$

- **Disperzija** (varianca) ↳ razpršenost **in standardni odklon**:

$$\underline{D(X) = E((X - E(X))^2) = \sum_{i=1}^{\infty} (x_i - E(X)^2)\, p_i = E(X^2) - E(X)^2 = \sum_{i=1}^{\infty} x_i^2 p_i - \left(\sum_{i=1}^{\infty} x_i p_i\right)^2}$$

$$\underline{\sigma(X) = \sqrt{D(X)}}$$

Velja: $D(aX + b) = a^2 D(X)$

Standardizirana slučajna spremenljivka: $\underline{X^* = \dfrac{X - E(X)}{\sigma(X)}}$, velja: $E(X^*) = 0$ in $D(X^*) = 1$

- **Enakomerna diskretna porazdelitev:** $X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}$

končno mnogo enako verjetnih izidov

$$F_X(x) = \begin{cases} 0, & x \leq x_1 \\ \frac{i}{n}, & x_i < x \leq x_i+1 \\ 1, & x > x_n \end{cases}$$

$$\boxed{E(X) = \frac{n+1}{2} \quad \text{in} \quad D(X) = \frac{n^2-1}{12}}$$

možna sta le dva izida: zgodil/ne zgodil

- **Bernoullijev obrazec:** $P_n(k) = \binom{n}{k} \underbrace{p^k}_{\text{se je zgodil}} \overbrace{(1-p)^{n-k}}^{\text{se ni zgodil}}$

n – toliko ponovitev

k – A se zgodi tolikokrat

Binomski simbol:

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

$$\binom{n}{0} = 1, \quad \binom{n}{1} = n$$

možna izida poskusa

- **Bernoullijeva porazdelitev:** $X \sim \begin{pmatrix} 0 & 1 \\ \underset{\text{se ne zgodi}}{1-p} & \underset{\text{se zgodi}}{p} \end{pmatrix}$ (n = 1)

ena ponovitev poskusa

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1-p, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

$$\boxed{\begin{aligned} E(X) &= p \\ D(X) &= (1-p) \cdot p \end{aligned}}$$

kolikokrat se je nekaj zgodilo? → končna

- **Binomska porazdelitev $B(n,p)$:** $X \sim \begin{pmatrix} 0 & 1 & \dots & n \\ P_n(0) & P_n(1) & \dots & P_n(n) \end{pmatrix}$

(ponavljamo Bernoullija)

št. ponovitev A v n ponovitvah poskusa

$$\boxed{E(X) = n \cdot p \quad , \quad D(X) = n \cdot p(1-p)}$$

kdaj se nekaj prvič zgodi? → neskončna

- **Geometrijska porazdelitev:** $X \sim \begin{pmatrix} 1 & 2 & \dots & k & \dots \\ p_1 & p_2 & \dots & p_k & \dots \end{pmatrix}$, $p_k = (1-p)^{k-1} \cdot p$

št. poskusov do prve ponovitve

perioda vračanja

$$\boxed{E(X) = \frac{1}{p} \quad , \quad D(X) = \frac{1-p}{p^2}}$$

št. pojavitev A na danem intervalu

- **Poissonova porazdelitev:** $X \sim P(\lambda)$, $X \sim \begin{pmatrix} 0 & 1 & \dots & k & \dots \\ p_\lambda(0) & p_\lambda(1) & \dots & p_\lambda(k) & \dots \end{pmatrix}$, $p_\lambda(k) = \frac{e^{-\lambda}\lambda^k}{k!}$

$$\boxed{E(X) = \lambda \quad , \quad D(X) = \lambda}$$

Poissonov izrek: $\lambda \in \mathbb{R}^+$ in $n \in \mathbb{N}$ in $p = \frac{\lambda}{n}$, potem: $\lim_{n \to \infty} \binom{n}{k} p^k (1-p)^{n-k} = \frac{e^{-\lambda}\lambda^k}{k!}$

↓

aproksimiramo lahko z binomskim

povprečno št. na enoto

[vozil/min] ali [smrti/leto]

(koeficient pojavitve na eno enoto)

Velik n ali majhen p garantirata dober rezultat tega postopka!

- **Slučajna zvezna spremenljivka X:**

$$P(x_1 < X \leq x_2) = P(x_1 \leq X \leq x_2) = P(x_1 \leq X < x_2) = P(x_1 < X < x_2)$$ → upljučitav robov ne vpliva na verjetnost

$P(X=x)=0$ → ni pa nujno nemogoč (le X je neskončen)

- Gostota verjetnosti (→ zvezne slučajne spremenljivke): za poljubna $x_1, x_2 \in \mathbb{R}$ $(x_1 < x_2)$ velja: $P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} p_X(x)\,dx$

p_X - pozitivna integrabilna funkcija

izhaja iz verjetnostne funkcije za diskretne spremenljivke

$$\int_{-\infty}^{\infty} p_X(x)\,dx = 1$$

- Porazdelitvena funkcija: $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} p_X(t)\,dt$ (povsod zvezna)

1. $0 \leq F_X(x) \leq 1$
2. naraščajoča
3. $P(a < X < b) = F_X(b) - F_X(a)$

- Matematično upanje: $E(X) = \int_{-\infty}^{\infty} x \cdot p_X(x)\,dx$

nepravi integral → $E(x) = \lim_{a \to -\infty} \int_a^0 x \cdot p_x(x)\,dx + \lim_{b \to \infty} \int_0^b x \cdot p_x(x)\,dx$

zato $E(x)$ ne obstaja vedno!

- Disperzija in standardni odklon: $D(X) = E(X^2) - E(X)^2 = \int_{-\infty}^{\infty} x^2 p_X(x)\,dx - \left(\int_{-\infty}^{\infty} x\, p_X(x)\,dx\right)^2$

$D(X) = E\left((X - E(X))^2\right) = \int_{-\infty}^{\infty} (x - E(X))^2 p_X(x)\,dx$

$$\sigma(X) = \sqrt{D(X)}$$

Velja: $D(aX+b) = a^2 D(X)$

Standardizirana slučajna spremenljivka: $Z = \dfrac{X - E(X)}{\sigma(X)}$, $E(Z)=0$, $D(Z)=1$

(vse točke $Z_X = [a,b]$ so enako verjetne)

- Enakomerna zvezna porazdelitev: $p_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{sicer} \end{cases}$, $F_X(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$

$$\boxed{E(X) = \frac{a+b}{2} \quad \text{in} \quad D(X) = \frac{(b-a)^2}{12}}$$

(X je čakalna doba do 1. pojavitve s parametrom λ na $Z_X = [0, \infty]$ → nima spomina)

- Eksponentna porazdelitev (podobnost z geometrijsko porazdelitvijo): $p_X(x) = \lambda e^{-\lambda x}$, $F_X(x) = 1 - e^{-\lambda x}$

$$\boxed{E(X) = \frac{1}{\lambda} \quad \text{in} \quad D(X) = \frac{1}{\lambda^2}}$$

λ - povprečno št. pojavitev na enoto

LaHopitalovo pravilo:

$$\lim_{b \to \infty} b e^{-\lambda b} = \lim_{b \to \infty} \frac{b}{e^{\lambda b}} \overset{\text{odvod po } b}{=} \lim_{b \to \infty} \frac{1}{\lambda e^{\lambda b}} = 0$$

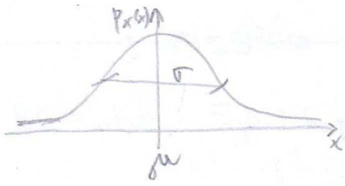
Per partes: $\int u\,dv = uv - \int v\,du$

(podobnost z binomsko porazdelitvijo)

Vsota velikega števila neodvisnih slučajnih spremenljivk limitira proti normalni porazdelitvi!

- **Normalna ali Gaussova porazdelitev** : $X \sim N(\mu, \sigma)$

↳ $Z_X = (-\infty, \infty)$



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

$$P(a < X \leq b) = \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$$

$$\boxed{E(X) = \mu \quad \text{in} \quad D(X) = \sigma^2}$$

Standardizirana normalna porazdelitev : $Z \sim N(0, 1)$

$$F_Z(z) = \phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2} dt \rightarrow \text{tabele} \quad , \quad P(a < Z \leq b) = \phi(b) - \phi(a)$$

$$\left(\phi(-z) = 1 - \phi(z)\right)$$

Kvantile, percentile : $P(X \leq x_p) = p$ - p-ta kvantila zvezne porazdelitve X

- mediana : $x_{0,5} = \mu$
- kvartile: $x_{0,25}$, $x_{0,5}$, $x_{0,75}$
- decile : $x_{0,1}$, $x_{0,2}$, ..., $x_{0,9}$
- interkvartilni razpon: $x_{0,75} - x_{0,25}$

- **Lognormalna porazdelitev**: $X = e^Y$, kjer $Y \sim N(\mu, \sigma)$, $Z_X = (0, \infty)$

Produkt velikega števila neodvisnih slučajnih spremenljivk limitira proti lognormalni porazdelitvi!

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}x} e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2} \quad , \quad F_X(x) = F_Y(\ln x) = \phi\left(\frac{\ln x - \mu}{\sigma}\right)$$

$$\boxed{E(X) = e^{\mu + \frac{\sigma^2}{2}} \quad \text{in} \quad D(X) = e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right)}$$

- **Porazdelitev hi-kvadrat** : $X \sim \chi^2(n)$, $Z_X = (0, \infty)$, n - št. prostostnih stopenj

Pri velikih n se približuje normalni porazdelitvi!

$$p_X(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad , \quad \Gamma(r) = \int_0^\infty t^{r-1} e^{-t} dt$$

$$\boxed{E(X) = n \quad \text{in} \quad D(X) = 2n}$$

- **Studentova t-porazdelitev** : $Z_T = (-\infty, \infty)$, n - št. prostostnih stopenj

$$p_T(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(\frac{n}{n+x^2}\right)^{\frac{n+1}{2}}$$

$$\boxed{\begin{array}{l} \text{za } n > 1 : E(T) = 0 \\ \text{za } n > 2 : D(T) = \frac{n}{n-2} \end{array}}$$

- **Slučajni vektor:** $X = (X_1, X_2, \dots, X_n)$ → n-terica slučajnih spremenljivk

Porazdelitvena funkcija: $F_X(x_1, x_2, \dots, x_n) = F_{(X_1, X_2, \dots, X_n)}(x_1, x_2, \dots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n)$

↳ naraščajoča in zvezna z desne

① $F_{(x,y)}(-\infty, -\infty) = 0$ in $F_{(x,y)}(+\infty, +\infty) = 1$ → $0 \le F_{(x,y)}(x,y) \le 1$

robni porazdelitvi (X,Y)

② $F_{(x,y)}(-\infty, y) = F_{(x,y)}(x, -\infty) = 0$ in $F_{(x,y)}(+\infty, y) = \boxed{F_Y(y)}$, $F_{(x,y)}(x, +\infty) = \boxed{F_X(x)}$

③ $P((X,Y) \in [a,b] \times (c,d]) = F_{(x,y)}(b,d) - F_{(x,y)}(a,d) - F_{(x,y)}(b,c) + F_{(x,y)}(a,c)$

- **Diskreten slučajni vektor:** → diskretne števno mnoge točke v $Z_{(x,y)}$

Verjetnostna funkcija slučajnega vektorja: $p_{ij} = P(X = x_i, Y = y_j)$, $i, j \in \mathbb{N}$

$p_{ij} \ge 0$ in $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{ij} = 1$

Robni porazdelitvi X in Y: $q_i = P(X = x_i) = \sum_{j=1}^{\infty} \overset{\text{seštejemo po X}}{p_{ij}}$ in $r_j = P(Y = y_j) = \overset{\text{seštejemo po y}}{\sum_{i=1}^{\infty}} p_{ij}$

Pogojne porazdelitve: $r_k \neq 0$ → $p_{i|k} = P(X = x_i \mid Y = y_k) = \frac{p_{ik}}{r_k}$

$p_{i|k} \ge 0$ in $\sum_{i=1}^{\infty} p_{i|k} = 1$

Pogojno matematično upanje: $E(X|y_k) = \sum_{i=1}^{\infty} x_i \, p_{i|k} = \frac{1}{r_k} \sum_{i=1}^{\infty} x_i \, p_{ik}$

$E(X) = \sum_{i=1}^{\infty} E(X|y_k) \, r_k$

Neodvisnost slučajnih spremenljivk:

X in Y neodvisni, če:
- $F_{(x,y)}(x,y) = F_X(x) \, F_Y(y)$
- $P(X = x, Y = y) = P(X = x) \, P(Y = y)$
- $p_{ij} = q_i r_j$
- $p_{i|k} = q_i$

neodvisni sta, če velja vsaj eden od teh pogojev

- Zvezen slučajni vektor: → $Z_{(x,y)} = Z_x \times Z_y \subseteq \mathbb{R}^2$

Gostota verjetnosti slučajnega vektorja: je taka integrabilna funkcija $p_{(x,y)}$, da velja:

$$F_{(x,y)}(x,y) = P(X \le x, Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} p_{(x,y)}(u,v)\,du\,dv$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{(x,y)}(u,v)\,du\,dv = 1 \quad \text{in} \quad P((X,Y)\in(a,b]\times(c,d]) = \int_{a}^{b}\int_{c}^{d} p_{(x,y)}(u,v)\,du\,dv$$

Robni gostoti verjetnosti: $p_X(x) = \int_{-\infty}^{\infty} p_{(x,y)}(x,y)\,dy$ in $p_Y(y) = \int_{-\infty}^{\infty} p_{(x,y)}(x,y)\,dx$

Pogojne porazdelitve: $p_{X|_Y}(x) = \frac{p_{(x,y)}(x,y)}{p_Y(y)}$

$$\int_{-\infty}^{\infty} p_{X|_Y}(x)\,dx = 1 \quad \text{in} \quad P(a \le X \le b \mid Y=y) = \int_{a}^{b} p_{X|_Y}(x)\,dx = \int_{a}^{b} \frac{p_{(x,y)}(x,y)}{p_Y(y)}\,dx$$

Pogojno matematično upanje: $\overbrace{\boxed{E(X|Y=y)}}^{\text{regresijska enačba}} = \int_{-\infty}^{\infty} x\,p_{X|_Y}(x)\,dx = \frac{1}{p_Y(y)}\int_{-\infty}^{\infty} x\,p_{(x,y)}(x,y)\,dx$

$$E(X) = \int_{-\infty}^{\infty} E(X|Y=y)\,p_Y(y)\,dy$$

Neodvisnost slučajnih spremenljivk: X in Y sta neodvisni če:

- $F_{(x,y)}(x,y) = F_X(x)F_Y(y)$
- $p_{(x,y)}(x,y) = p_X(x)\,p_Y(y)$
- $p_{X|_Y}(x) = p_X(x)$

je izpolnjen vsaj eden od teh pogojev

- Funkcije slučajnega vektorja: $Z(s) = g(X(s), Y(s))$ oz. $Z = g(X,Y)$

Diskreten vektor: $P(Z=z_k) = \sum_{i,j:\,g(x_i,y_j)=z_k} p_{ij}$ in $E(Z) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} g(x_i,y_j)\,p_{ij}$

Zvezen vektor: $F_Z(z) = \iint_{g(x,y)\le z} p_{(x,y)}(x,y)\,dx\,dy$

$$E(Z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\,p_{(x,y)}(x,y)\,dx\,dy$$

Vedno velja:

$$E(Z) = E(X) + E(Y)$$

Klancek.si

- **Kovarianca:** če sta X in Y neodvisni slučajni spremenljivki, velja:

$$E(XY)=E(X)E(Y) \quad \text{in} \quad D(X+Y)=D(X)+D(Y)$$

Če imata X in Y končni matematični upanji in končni disperziji, potem:

$$Cov(X,Y)=\sigma_{XY}=K(X,Y)=E\big((X-E(X))(Y-E(Y))\big)$$

Lastnosti:
- $Cov(X,X)=D(X)$
- $Cov(X,Y)=Cov(Y,X)$
- $Cov(a+bX, c+dY)=bd\,Cov(X,Y)$
- $Cov(X,Y)=E(XY)-E(X)E(Y)$
- $D(X+Y)=D(X)+D(Y)+2\,Cov(X,Y)$
- $|Cov(X,Y)|\leq \sigma(X)\sigma(Y)$

- **Korelacija:** $\rho(X,Y)=Corr(X,Y)=r(X,Y)=\dfrac{Cov(X,Y)}{\sigma(X)\sigma(Y)}$

Lastnosti:
- $-1\leq \rho(X,Y)\leq 1$
- če $Cov(X,Y)=0$, potem $\rho(X,Y)=0$ → X,Y nekorelirani, sicer korelirani
- $|\rho(X,Y)|=1$ → X in Y linearno korelirani

- **Matematično upanje:** $E(X)=(E(X_1),\ldots,E(X_n))$

- **Disperzija** (kovariančna matrika): ~~[illegible]~~ $K_{ij}=Cov(X_i,X_j)$

Lastnosti:
- $K_{ii}=D(X_i)$
- $K_{ji}=K_{ij}$
- vse lastne vrednosti K so nenegativne

- **Dvorazsežna normalna porazdelitev in standardna dvorazsežna normalna porazdelitev:**

Glej liste predavanj!

regresijski krivulji sta premici

(X,Y) je dvorazsežna normalna porazdelitev s parametri:

robni porazdelitvi

$$\mu_X,\ \mu_Y,\ \sigma_X,\ \sigma_Y,\ \rho \ \rightarrow\ X\sim N(\mu_X,\sigma_X) \text{ in } Y\sim N(\mu_Y,\sigma_Y)$$

Velja, da če sta X in Y neodvisni, sta tudi nekorelirani in obratno!

Standardna dvorazsežna normalna porazdelitev: $X,Y\sim N(0,1)$

- **Neenakost Čebiševa:**

Če za X obstaja E(X) in D(X) potem: $P[E(X) - \overset{k>0}{k}\sigma(X) < X < E(X) + \overset{k>0}{k}\sigma(X)] \geq 1 - \frac{1}{k^2}$

Ocene verjetnosti, da se X nahaja na teh intervalih

- **Neenakost Markova:** Za X obstaja E(X) in $Z_x \subseteq [0, \infty)$: $P[X \geq \overset{k>0}{k}E(X)] \leq \frac{1}{k}$

- **Zaporedje slučajnih spremenljivk:**

$$\overline{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$ → povprečje

$$S_n = X_1 + X_2 + \dots + X_n$$ → delna vsota

$$Z_n = \frac{S_n - E(S_n)}{\sigma(S_n)}$$ → standardizacija delne vsote

$E(S_n) = n \cdot E(X)$

$E(\overline{X}_n) = E(X)$

$D(S_n) = n \cdot D(X)$

$D(\overline{X}_n) = \frac{D(X)}{n}$

(načeloma lahko zaporedje obravnavamo kot poskus s spremenljivko S_n)

- **Zakon velikih števil:** Če so X_i medsebojno neodvisne in enako porazdeljene z $E(X_i) = \mu < \infty$ in $D(X_i) \overset{\sigma^2}{<} \infty$, potem za vsak $\varepsilon > 0$:

$$\lim_{n \to \infty} P(|\overline{X}_n - \mu| < \varepsilon) = 1$$

Če $n \to \infty$, potem se vedno približujemo $E(\overline{X}_n) = E(X) = \mu$!

- **Centralni limitni izrek:** Če so X_i medsebojno neodvisne in enako porazdeljene z $E(X_i) < \infty$ in $D(X_i) < \infty$, potem:

$$\lim_{n \to \infty} P(Z_n \leq x) = \Phi(x)$$ Gaussov integral

Če $n \to \infty$, potem se ne glede na porazdelitev slučajnih spremenljivk bližamo normalni porazdelitvi.

Binomska porazdelitev $X \sim B(n,p) \xrightarrow{\text{velik } n}$ normalna porazdelitev.

Poissonova porazdelitev $X \sim P(\lambda) \xrightarrow{\text{velik } \lambda}$ normalna porazdelitev.

Hi-kvadrat porazdelitev $X \sim \chi^2(n) \xrightarrow{\text{velik } n}$ normalna porazdelitev.

Studentova porazdelitev T z m p.s. $\xrightarrow{\text{velik } m}$ standardna normalna porazdelitev.

- **Statistika:**

statistična enota - predmet proučevanja

populacija $\mathcal{P}$ - množica vseh enot

vzorec $a_1, \dots, a_n \in \mathcal{P}$ - del populacije

statistična spremenljivka X - lastnost enote, ki nas zanima

parameter Θ - značilnost populacije

vzorčna statistika $Y = f(X_1, \dots, X_n)$ - značilnost vzorca

cenilka $\hat{\Theta}$ parametra Θ - vzorčna statistika, katere porazdelitev je odvisna le od Θ

↳ Θ ne moremo določiti na vzorcu, zato uporabljamo cenilke!

- **Vzorčenje:** Vzorcu ustreza slučajni vektor $(X_1, \dots, X_n)$.

Enostaven slučajni vzorec - X_i med seboj neodvisne in enako porazdeljene slučajne spremenljivke.

Reprezentativen - dovolj velik in nepristransko izbran.

primerjava z delno vsoto zaporedja slučajnih spremenljivk

- **Vzorčno povprečje:** $$\underline{\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}}$$

če je $(X_1, \dots, X_n)$ enostaven slučajni vzorec z $E(X_i) = \mu$ in $D(X_i) = \sigma^2$, $i = 1, \dots, n$, potem:

$$\underline{E(\bar{X}) = \mu \quad \text{in} \quad D(\bar{X}) = \frac{\sigma^2}{n}}$$

- **Porazdelitev vzorčnega povprečja:** Če je $\underline{X \sim N(\mu, \sigma)}$, potem: $\underline{\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$

Tudi če X ni $X \sim N(\mu, \sigma)$ po centralnem limitnem izreku za $n > 30$ velja:

$$\underline{\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad \text{oz.} \quad \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)}$$

standardizirano vzorčno povprečje

- **Vzorčna disperzija:** $$\underline{S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}$$ → S^2 je slučajna spremenljivka, saj je pri isti populaciji vzorec vedno različen!

če je $(X_1, \dots, X_n)$ enostaven slučajni vzorec z $E(X_i) = \mu$ in $D(X_i) = \sigma^2$, $i = 1, \dots, n$, potem:

$$\underline{E(S^2) = \sigma^2}$$

Vzorčna standardna deviacija: $\underline{S = \sqrt{S^2}}$

Porazdelitev vzorčne disperzije: če $\underline{X \sim N(\mu, \sigma)}$, potem: $\underline{\frac{(n-1) S^2}{\sigma^2} \sim \chi^2(n-1)}$

Centralni limitni izrek, $n > 30$: $\underline{S^2 \sim N\left(\sigma^2, \sqrt{\frac{2}{n-1}} \sigma^2\right)}$

- **Standardizirano vzorčno povprečje:** σ^2 ne poznamo, nadomestimo z S^2

$$X \sim N(\mu, \sigma) \rightarrow T = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}}$$ - Studentova t-porazdelitev z $(n-1)$ prost. stopnjami

- **Vzorčna korelacija:** $(X_1, Y_1), \ldots, (X_n, Y_n)$ je enostaven slučajni vzorec slučajnega vektorja (X, Y).

$$R = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

- **Ocenjevanje parametrov:**
  - intervalsko ocenjevanje → ocenimo interval, ki s predpisano vrednostjo vsebuje parameter θ (podamo razpon)
  - točkasto ocenjevanje → parameter θ ocenimo s številsko vrednostjo cenilke $\hat{\theta}$ (podamo napako)

Lastnosti cenilk:
- $\hat{\theta}$ je nepristranska cenilka za θ, če je $E(\hat{\theta}) = \theta$.
- $\hat{\theta}$ je asimptotično nepristranska cenilka za θ, če je $\lim\limits_{n\to\infty} E(\hat{\theta}) = \theta$.
- od cenilk $\hat{\theta}_1$ in $\hat{\theta}_2$ za θ je učinkovitejša $\hat{\theta}_1$, če $D(\hat{\theta}_1) < D(\hat{\theta}_2)$.
- $\sigma(\hat{\theta}) = SE(\hat{\theta})$ - standardna napaka cenilke $\hat{\theta}$.
- $\hat{\theta}$ je dosledna cenilka, če za vsak $\varepsilon > 0$: $\lim\limits_{n\to\infty} P(|\hat{\theta} - \theta| \geq \varepsilon) = 0$.

Primeri cenilk: $\overset{\text{vzorčno povprečje}}{\bar{X}}$ cenilka za μ, $\overset{\text{vzorčna disperzija}}{S^2}$ cenilka za σ^2, $\overset{\text{vzorčna standardna deviacija}}{S}$ cenilka za σ.

nepristranske, dosledne

pristranska, asimptotično nepristranska

- **Momenti višjega reda:** moment reda k glede na točko $a \in \mathbb{R}$ je: $m_k(a) = E((X-a)^k) = \sum_{i=1}^{n} (x_i - a)^k p_i$ oz $\int_{\mathbb{R}} (x-a)^k p_X(x)\,dx$

$a = 0 \rightarrow$ začetni moment $z_k = m_k(0)$ ↗ npr: $z_1 = E(X)$

$a = E(X) \rightarrow$ centralni moment $m_k = m_k(E(X))$ ↗ npr: $m_2 = D(X)$

↳ Uporabljamo za izbiro metode modela!

- **Metoda momentov (MoM):**

$z_k = E(X^k)$ - k-ti začetni moment X

$\hat{z}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ - k-ti vzorčni moment $(X_1, \ldots, X_n)$

$\hat{z}_k$ je cenilka za z_k

če se parameter izraža z momenti kot: $\theta_i = f_i(z_1, z_2, \ldots, z_m)$

potem je cenilka po MoM: $\hat{\theta}_i = f_i(\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_m)$

- **Metoda največjega verjetja (ML):**

za $(x_1, \dots, x_n)$ je porazdelitev podana s $p(x, \theta)$, ki je odvisna od $\theta = (\theta_1, \theta_2, \dots, \theta_m)$.

Cenilka po ML je vrednost $\hat{\theta}$, pri kateri $\boxed{L(\theta)}$ doseže maksimum : $L(\theta) = p(x_1, \theta) \cdot p(x_2, \theta) \cdot \ldots \cdot p(x_n, \theta)$

(funkcija verjetja

$$\overset{\shortparallel}{} \prod_{i=1}^{n} p_i = \prod_{i=1}^{n} P(Y = (x_1, \dots, x_n))$$

- **Intervalsko ocenjevanje parametrov:** → kvalitativna ocena cenilke nekega parametra

Interval zaupanja $[A, B]$, ki s predpisano verjetnostjo (stopnja zaupanja) $1-\alpha$ vsebuje θ :

$$P(\theta \in [A, B]) = 1 - \boxed{\alpha}$$ — stopnja tveganja

- **Intervali zaupanja za μ:** enostaven slučajni vzorec $(X_1, \dots, X_n)$ z $E(x_i) = \mu$ in $D(x_i) = \sigma^2$.

$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$ – nepristranska cenilka za μ

Interval zaupanja (simetričen):

$$P(-z(\alpha/2) \le Z \le z(\alpha/2)) = 1 - \alpha$$

za velike n : $\boxed{Z} = \frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$ → z-test

- **σ^2 poznamo:**
  1. izberemo α,
  2. določimo $z(\alpha/2) = z_{1-\alpha/2}$: $\phi(z(\alpha/2)) = 1 - \alpha/2$
  3. izračunamo $\bar{X}$.

Interval zaupanja: $\left[\bar{X} - z(\alpha/2)\boxed{\sqrt{\frac{\sigma^2}{n}}}, \bar{X} + z(\alpha/2)\boxed{\sqrt{\frac{\sigma^2}{n}}}\right]$ ← $SE(\bar{x})$

s stopnjo zaupanja $(1-\alpha)$

- **σ^2 ne poznamo:** nepoznan σ^2 nadomestimo s cenilko : $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

za $n > 30$: enako kot prej, le da z S^2 namesto σ^2.

za $n < 30$: Studentova porazdelitev z $(n-1)$ prostostnimi stopnjami : $\boxed{T} = \frac{\bar{x} - \mu}{\sqrt{s^2/\mu}}$ ← T-test

$$t(\alpha/2) = t_{1-\alpha/2} \Leftrightarrow P(T > t(\alpha/2)) = \alpha/2$$

Interval zaupanja: $\left[\bar{X} - t(\alpha/2)\sqrt{\frac{s^2}{n}}, \bar{X} + t(\alpha/2)\sqrt{\frac{s^2}{n}}\right]$

s stopnjo zaupanja $(1-\alpha)$

- **Enostranski interval zaupanja:** imamo samo eno mejo $-z(\alpha)$ oz. $z(\alpha)$, druga je $-\infty$ oz. ∞.

Spodnja meja zaupanja za μ s stopnjo zaupanja $(1-\alpha)$: $\bar{X} - z(\alpha)\sqrt{\frac{\sigma^2}{n}} \le \mu$

Zgornja meja zaupanja za μ s stopnjo zaupanja $(1-\alpha)$: $\bar{X} + z(\alpha)\sqrt{\frac{\sigma^2}{n}} \ \mu$

$\phi(z(\alpha)) = 1 - \alpha$

- **Interval zaupanja za σ:** če je $[A,B]$ i.z. za σ^2 s stopnjo z. $(1-\alpha)$, potem je $[\sqrt{A}, \sqrt{B}]$ i.z. za σ z enako stopnjo z. $(1-\alpha)$!

- **μ poznamo:**

1. vzamemo cenilko: $\hat{D} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$
2. če je $X_i \sim N(\mu,\sigma)$, potem: $X^2 = \frac{n\hat{D}}{\sigma^2} \sim \chi^2(n)$
3. porazdelitev ni simetrična, zato iščemo ustrezna a in b:

$$P(X^2>b) = P(X^2<a) = \alpha/2 \rightarrow a = \chi^2_{\alpha/2} \text{ in } b = \chi^2_{1-\alpha/2}$$

Interval zaupanja: $\left[\frac{n\hat{D}}{b}, \frac{n\hat{D}}{a}\right]$

s stopnjo zaupanja $(1-\alpha)$

- **μ ne poznamo:**

1. cenilki za μ in σ^2 → $\bar{X}$ in S^2.
2. če $X_i \sim N(\mu,\sigma)$, potem: $X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
3. iščemo: $P(X^2>b) = P(X^2<a) = \alpha/2 \rightarrow a = \chi^2_{\alpha/2}$ in $b = \chi^2_{1-\alpha/2}$

Interval zaupanja: $\left[\frac{(n-1)S^2}{b}, \frac{(n-1)S^2}{a}\right]$

s stopnjo zaupanja $(1-\alpha)$

- **Normalna aproksimacija binomske porazdelitve:**

Iščemo verjetnost p, da slučajen element populacije ustreza danemu kriteriju!

n - velikost vzorca

X - št. elementov vzorca, ki ustrezajo kriteriju

$X \sim B(n,p)$, $E(X) = np$ in $D(X) = np(1-p)$

če je $n \gg p$, potem: $Z = \frac{X-np}{\sqrt{np(1-p)}} \sim N(0,1)$

V praksi za $np > 5$ in $n(1-p) > 5$ ta aproksimacija ustreza.

- **Interval zaupanja za delež populacije:**

cenilka za p je: $\hat{p} = \frac{X}{n}$

1. Določimo $z(\alpha/2)$ z: $\Phi(z(\alpha/2)) = 1-\alpha/2$.
2. Neznani p v $SE(p)$ nadomestimo s $\hat{p}$. → velik vzorec!

$$Z = \frac{\hat{p}-p}{\underbrace{\sqrt{\frac{p(1-p)}{n}}}_{SE(p)}} \sim N(0,1)$$

Interval zaupanja: $\left[\hat{p} - z(\alpha/2)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z(\alpha/2)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$

s stopnjo zaupanja $(1-\alpha)$

$(\varepsilon = |p-\hat{p}| > |z(\alpha/2)\, SE(\hat{p})|)$

Klancek.si

- **Statistična domneva ali hipoteza :** Izjava o porazdelitvi neke slučajne spremenljivke.

Govori o lastnosti populacije!

H_0 - ničelna domneva → to preizkušamo

H_A - alternativne domneve

- **Testna statistika:** $T: (X_1, \dots, X_n) \to \mathbb{R}$ z zalogo vrednosti: $Z(T) \subseteq \mathbb{R}$

$W_0 \subseteq Z(T)$ - kritično območje testa (H_0 zavrnemo)

$Z(T) \setminus W_0$ - sprejemljivo območje testa (H_0 ne zavrnemo)

- **Lastnosti testov :**

Napaka I. vrste - zavrnemo pravilno domnevo H_0. Maksimalna verjetnost te napake je α oz. stopnja značilnosti / tveganje testa.

Napaka II. vrste - ne zavrnemo H_0, čeprav je napačna. Verjetnost te napake β je odvisna od parametrov, velikosti vzorca...

$\alpha < p$ → H_0 ne zavrnemo
$\alpha > p$ → H_0 zavrnemo

Moč testa - verjetnost zavrnitve H_0, ki je napačna je $1-\beta$.

p-vrednost - najmanjša stopnja značilnosti α, pri kateri H_0 še zavrnemo.

Splošni postopek :

1. Postavimo H_0 in H_A
2. Izberemo stopnjo značilnosti α.
3. Določimo ustrezno testno statistiko T.
4. Poiščemo kritično območje W_0, da bo: $P(T \in W_0 | H_0) = \alpha$.
5. Izračunamo vrednost testne statistike T na vzorcu.
6. Za $T \in W_0$ domnevo H_0 s stopnjo značilnosti α zavrnemo oz. obratno ne zavrnemo.

- **Preskušanje μ normalne porazdelitve:** $(X_1, \dots, X_n)$; $X \sim N(\mu, \sigma)$

$H_0: \mu = \mu_0$

$H_A: \mu \neq \mu_0$

Testna statistika: $Z = \dfrac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} \sim N(0,1)$

$W_0 = (-\infty, -z(\alpha/2)) \cup (z(\alpha/2), \infty)$ ← $\Phi(z(\alpha/2)) = 1 - \alpha/2$

Enostranska domneva: $H_A: \mu < \mu_0 \to W_0 = (-\infty, -z(\alpha))$

$H_A: \mu > \mu_0 \to W_0 = (z(\alpha), \infty)$

Če σ^2 ne poznamo, ga nadomestimo s S^2 → studentova porazdelitev z $(n-1)$ prost. stopnjami.

Če preizkušamo druge parametre, so zadeve zelo sorodne intervalom zaupanja!

• Preskušanje skladnosti (χ^2-test):

Ho je domneva o modelu porazdelitve za dano populacijo.

1. Slučajni vzorec n razdelimo na k razredov $S_1, \dots, S_k$.

o_i - frekvenca razreda S_i

$e_i = n \cdot P(X \in S_i | H_0)$ - domnevna frekvenca S_i

2. Testna statistika: $X_0^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$

3. Če je Ho pravilna in je m (nepoznano) št. parametrov domnevane porazdelitve, za velike n ($e_i > 5$) velja:

$$X_0^2 \sim \chi^2(k-m-1)$$

4. $W_0 = (c_\alpha, \infty)$: $P(\chi^2(k-m-1) > c_\alpha) = \alpha$ oz. $c_\alpha = \chi^2_{1-\alpha}$

5. Domnevo zavrnemo, če : $X_0^2 > c_\alpha$.

• Preskušanje neodvisnosti (χ^2-test):

Ho je domneva, da sta lastnosti populacije statistično neodvisni.

1. Vzorec velikosti n razdelimo v r oz. s razredov.

o_{ij} - frekvenca i-tega razreda prve in j-tega razreda druge lastnosti

e_{ij} - domnevne frekvence razredov

2. Testna statistika: $X_0^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \sim \chi^2((r-1)(s-1))$

predpostavimo, da Ho drži

Če sta neodvisni, potem: $e_{ij} = P(L_1 \in r_i \text{ in } L_2 \in s_j) = P[L_1 \in r_i] \cdot P[L_2 \in s_j]$

3. Določimo c_α ($= \chi^2_{1-\alpha}$) in domnevo zavrnemo, če : $X_0^2 > c_\alpha$

• Preskušanje σ^2 normalne porazdelitve:

$(X_1, \dots, X_n) \sim N(\mu, \sigma)$

$H_0: \sigma^2 = \sigma_0^2$

$H_A: \sigma_A^2 \neq \sigma_0^2$

Poznamo μ: $D^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 \rightarrow D^2 \sim \chi^2(n)$

Ne poznamo μ: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \rightarrow S^2 \sim \chi^2(n-1)$

($\bar{X}$ namesto μ)

$$X_0^2 = \frac{(n-1) \cdot S^2}{\sigma^2} \sim \chi^2(n-1) \quad \text{oz.} \quad X_0^2 = \frac{n \cdot D^2}{\sigma^2} \sim \chi^2(n)$$

$$W_0 = (\chi^2_{1-\alpha}(n \text{ oz. } (n-1)), \infty)$$