Odvodi	
funkcija	odvod
c	0
x^n	nx^{n-1}
a^x	$a^x \ln a$
$\frac{a^x}{\ln a}$	a^x
x^x	$x^x(1+\ln x)$
$\ln(x)$	$\frac{1}{x}$
$\log_a(x)$	$\frac{1}{x \ln(a)}$
$\sin(x)$	cos(x)
$\cos(x)$	-sin(x)
tan(x)	$\frac{1}{\cos^2(x)}$
$\cot(x)$	$-\frac{1}{\sin^2(x)}$
$\arcsin(x)$	$ \frac{1}{\sqrt{1-x^2}} \\ -\frac{1}{\sqrt{1-x^2}} \\ \underline{1} $
$\arccos(x)$	$-\frac{1}{\sqrt{1-x^2}}$
$\arctan(x)$	$\frac{\frac{1}{1+x^2}}{1}$
$\operatorname{arccot}(x)$	$-\frac{1}{1+x^2}$
$\operatorname{sh}(x) = \frac{e^x - e^{-x}}{2}$	ch(x)
$ch(x) = \frac{e^x + e^{-x}}{2}$	sh(x)
$sh(x) = \frac{e^x - e^{-x}}{2}$ $ch(x) = \frac{e^x + e^{-x}}{2}$ $th(x) = \frac{sh(x)}{ch(x)}$	$\frac{1}{\operatorname{ch}^2(x)}$
$cth(x) = \frac{1}{th(x)}$	1
$arsh(x) = \ln(x + \sqrt{x^2 + 1})$	$\frac{-\frac{1}{\sinh^2(x)}}{\frac{1}{\sqrt{1+x^2}}}$ $\frac{1}{\sqrt{1-x^2}}$
$\operatorname{arch}(x) = \ln(x + \sqrt{x^2 - 1})$	$\frac{1}{\sqrt{1-x^2}}$
$\operatorname{arth}(x) = \frac{1}{2} \ln \frac{1+x}{1-x}$	$\frac{1}{(1+x)(1-x)}$

Osnove kombinatorike Rodovne funkcije

$$\sum_{n=0}^{\infty} a^n = \frac{1}{n} \sum_{n=0}^{\infty} a^n$$

$$\sum_{n=0}^{\infty} q^n = \frac{1}{1-q} \quad \sum_{n=0}^{b} q^n = \frac{1-q^{b+1}}{1-q}$$
$$\sum_{n=a}^{\infty} q^n = \frac{q^a}{1-q} \quad \sum_{n=a}^{b} q^n = \frac{q^a-q^{b+1}}{1-q}$$

$$a^{n} - b^{n} = (a - b)(a^{n-1} + a^{n-2}b + \dots + ab^{n-2} + b^{n-1})$$

$$\frac{a_0+\ldots+a_{k-1}x^{k-1}}{1-x^k}=a_0+\ldots+a_{k-1}x^{k-1}+a_0^k+\ldots+a_{k-1}x^{2k-1}+\ldots$$

$$Verjetnost na (Ω,\mathcal{F}) je preslikava $P:\mathcal{F}\to\mathbb{R}$ z lastnostmi:$$

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k$$

$$\frac{1}{(1-x)^n} = \sum_{k=0}^n \binom{n+k-1}{k} x^k$$

$$B_{\lambda}(x) = \sum_{n} {\lambda \choose n} x^{n} = (1+x)^{\lambda}; \qquad {\lambda \choose n} = \frac{\lambda^{n}}{n!}$$

Izbori

Imamo n oštevilčenih kroglic. Na koliko načinov lahko izberemo k kroglic?

	s pon.	brez pon.
variacije vrstni red je pomemben	n^k	$n^{\underline{k}}$
kombinacije vrstni red ni pomemben	$\binom{n+k-1}{k}$	$\binom{n}{k}$

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

Pravila za računanje z dogodki

idempotent

$$A \cup A = A = A \cap A$$
 komutativnost
$$A \cup B = B \cup A$$

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

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

$$(A \cap B) \cup C = (A \cup C) \cap (A \cup C)$$
 De Morgan
$$(\bigcup_{i \in I} A_i)^{\complement} = \bigcap_{i \in I} A_i^{\complement}$$

$$(\bigcap_{i \in I} A_i)^{\complement} = \bigcup_{i \in I} A_i^{\complement}$$

$$A \cup \Omega = \Omega \quad A \cap \Omega = A$$

$$A \cup \emptyset = A \quad A \cap \emptyset = \emptyset$$

$$A \cup A^{\complement} = \Omega \quad A \cap A^{\complement} = \emptyset$$

Neprazna družina dogodkov \mathcal{F} v Ω je σ -algebra, če velja

• zaprtost za komplemente:

$$A \in \mathcal{F} \implies A^{\complement} \in \mathcal{F}$$

• zaprtost za števne unije:

$$A_1, A_2, \dots \in \mathcal{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$$

 $\check{C}e$ zahtevamo zaprtost le za končne unije, je ${\mathcal F}$ le algebra. Ker je po De Morganovem zakonu $\left(\bigcup_{i\in I} A_i^{\complement}\right)^{\complement} = \bigcap_{i\in I} A_i$ imamo zaprtost tudi za preseke.

Ker je $A \setminus B = A \cap B^{\complement}$ je algebra zaprta tudi za razlike. Najmanjša algebra je **trivialna**: $\{\emptyset, \Omega\}$.

Največja algebra je: $\mathcal{P}(\Omega)$.

Najmanjša algebra, ki vsebuje E je $\{\emptyset, E, E^{\complement}, \Omega\}$.

Dogodka A in B sta **nezdružljiva** (disjunktna), če je $A \cup B = \emptyset$.

Zaporedje $\{A_i\}_i \in \mathcal{F}$ (končno ali števno mnogo) je **popoln** sistem dogodkov, če

$$\bigcup_{i} A_{i} = \Omega \qquad \qquad A_{i} \cup A_{j} = \emptyset, \, \forall i, j : i \neq j$$

- P(A) > 0 za $\forall A \in \mathcal{F}$
- \bullet $P(\Omega) = 1$
- Za paroma nezdružljive dogodke $\{A_i\}_{i=1}^{\infty}$ velja *števna*

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Lastnosti P:

- $P(\emptyset) = 0$
- P je končno aditivna.
- P je monotona: $A \subseteq B \implies P(A) < P(B)$
- $P(A^{\complement}) = 1 P(A)$
- P je zvezna:

$$A_1 \subseteq A_2 \subseteq \cdots \implies P(\bigcup_{i=1}^{\infty}) = \lim_{i \to \infty} P(A_i)$$

$$B_1 \supseteq B_2 \supseteq \cdots \implies P(\bigcap_{i=1}^{\infty}) = \lim_{i \to \infty} P(B_i)$$

$$P(A^{\complement}) = 1 - P(A)$$

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

Če lahko prostor izidov razbijemo na paroma nezdružljive enako verjetne dogodke, jih lahko obravnavamo kot izide: če je dogodek A unija k od n takih dogodkov, je P(A) = k/n.

Načelo vključitev in izključitev

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

$$P(\bigcup_{i=1}^n A_i) = \sum_{\emptyset \neq S \subseteq [n]} (-1)^{|S|+1} P(\bigcap_{i \in S} A_i)$$

Pogojna verjetnost

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Izrek o polni verjetnosti

Če H_1, H_2, H_3, \ldots tvorijo **popoln sistem dogodkov** (tj. vedno se zgodi natanko eden izmed njih), velja:

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + \dots$$

Dogodkom H_i često pravimo hipoteze in jih je lahko končno ali pa števno neskončno

Bayesova formula

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{P(A)}$$

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{P(H_i)P(A|H_i) + P(H_i)P(A|H_i) + \dots}$$

Brezpogojnim verjetnostim $P(H_i)$ pravimo **apriorne**, pogojnim verjetnostim $P(H_i|A)$ pa aposteriorne verjetnosti hipotez.

Neodvisnot dogodkov

Dogodka A in B sta **neodvisna**, če velja:

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

Če je P(B) > 0, je to ekvivalentno pogoju, da je P(A|B) =P(A). Če je 0 < P(B) < 1, pa je to ekvivalentno tudi pogoju, da je $P(A|B) = P(A|B^{\hat{c}})$. Dogodki A1, A2, A3, ...so neodvisni, če za poljubne različne indekse i_1, i_2, \ldots, i_k

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2})\dots P(A_{i_k})$$

Slučajne spremenljivke

Slučajna spremenljivka je funkcija $X:\Omega\to\mathbb{R}$ z lastnostijo, da je $\forall x \in \mathbb{R}$ množica $\{\omega \in \Omega : X(\omega) < x\} \equiv$ $X^{-1}((-\infty, x]) \equiv (X \le x)$ dogodek.

Diskretne porazdelitve

Diskretna enakomerna porazdelitev na n točkah:

$$X \sim \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix} = \text{Unif}\{a_1, \dots, a_n\}$$

Binomska porazdelitev

 $Bin(n, p), n \in \mathbb{N}, p \in (0, 1)$

Naj bo X št. uspelih (z verjetnostjo p) poskusov v zaporedju n poskusov. $X \sim \text{Bin}(n, p)$:

$$p_k = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Bernulijeva porazdelitev $Ber(p) \sim Bin(1, p)$ Geometrijske porazdelitev

 $Geo(p), p \in (0,1)$

(X = k) je dogodek, da se A zgodi prvič v k-ti ponovitvi.

$$P(X = k) = (1 - p)^{k-1}p$$

Pascalova / negativna binomska porazdelitev

 $Pas(m,p) = NB(m,p), m \in \mathbb{N}, p \in (0,1)$

(X = k) je dogodek, da se dogodek A zgodi m-tič v k-ti

Oziroma X je število poskusov do vključno m-tega uspelega, pri katerig vsak uspe z verjetnostjo $p. X \sim Pas(m, p)$:

$$p_k = P(X = k) = {k-1 \choose n-1} p^k (1-p)^{k-n}$$

Hipergeometrijska porazdelitev

Iz posode, v kateri je n kroglic, od tega r rdečih, na slepo in brez vračanja izvlečemo s kroglic. Če z X označimo število rdečih med izvlečenimi, ima ta slučajna spremenljivka hipergeometrijsko porazdelitev: $X \sim \text{Hip}(s, r, n) = \text{Hip}(r, s, n)$.

$$P(X=k) = \frac{\binom{r}{k}\binom{n-r}{s-k}}{\binom{n}{s}} = \frac{\binom{s}{k}\binom{n-s}{r-k}}{\binom{n}{r}}$$

Aproksimacija binomske porazdelitve

Poissonova porazdelitev $Poi(\lambda), \lambda > 0$

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

Če imamo veliko ponovitev $(n \to \infty)$ z malo verjetnostjo $(p \to 0)$, je $Bin(n, p) \approx Poi(np)$

Laplaceova lokalna formula: Če je $p, 1-p \gg \frac{1}{n}$, lahko $X \sim \text{Bin}(n, p)$ aproksimiramo

$$P(X=k) \approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(k-np)^2}{2\sigma^2}} \quad \sigma = \sqrt{np(1-p)}$$

Laplaceova integralska formula:

$$P(a \leq X \leq b) \approx \Phi\left(\frac{b-np}{\sigma}\right) - \Phi\left(\frac{a-np}{\sigma}\right)$$

Za majhno relativno napako zahtevamo še:

- $|a-np| \ll \sigma^{4/3}$ ali $|b-np| \ll \sigma^{4/3}$
- $a, b \in \mathbb{Z} + \frac{1}{2}$ ali $b a \gg 1$

Kumulativna porazdelitvena funkcija

$$F_X(x) = P(X \le x)$$

Zvezno porazdeljene slučajne spremenljivke

Realna slučajna spremenljivka X je porazdeljena zvezno, če obstaja taka integrabilna funkcija $p_X:\mathbb{R}\to[0,\infty),$ da za Diskretno porazdeljena sl. sprem. poliubna $a \le b$ velia:

$$P(a \le X \le b) = \int_{a}^{b} p_X(t) dx$$

Funkciji p_X pravimo porazdelitvena gostota Komulativna funkcija slučajne spremenljivke X

$$F_X(x) = \int_{-\infty}^x p_X(t)dt$$

Če je F_X zvezna in odsekoma zvezno odvedljiva, je X porazdeljena zvezno in za vse razen za končno mnogo točk \boldsymbol{x} velja $p_x(x) = F'_X(x)$.

Slučajni vektorji

Slučajni vektor je n-terica slučajnih spremenljivk X = $(X_1,\ldots,X_n):\Omega->R$

Porazdelitvena funkcija slučajnega vektorja

$$F_{(X_1,...,X_n)}(x_1,...,x_n) = P(X_1 \le x_1,...,X_n \le x_n)$$

Neodvisnot slučajnih spremenljivk

Slučajne spremenljivke so neodvisne, če je

$$F_{(X_1,...,X_n)}(x_1,...,x_n) = F_{X_1}(x_1)...F_{X_n}(x_n)$$

Torej so dogodki $(X_1 \leq x_1), \ldots, (X_n \leq x_n)$ neodvisni. Naj bo (X,Y) diskretno porazdeljen sluč. vektor:

$$p_{ij} = P(X = x_i, Y = y_j) \quad p_i = P(X = x_i) \quad P(Y = y_j)$$

$$X, Y$$
neodvisni $\iff p_{ij} = p_i q_i$

Naj bo (X,Y) zvezno porazdeljen sluč. vektor z gostoto – Indikator dogodka A je sl. sprem.: $p_{(X,Y)}(x,y)$, potem velja:

$$X, Y$$
neodvisni $\iff \exists p_X, p_Y : p_{(X,Y)}(x,y) = p_X(x)p_Y(y)$

Funkcije slučajnih spremenljivk

Naj bosta $A, B \subseteq \mathbb{R}$ odprti množici in $h: A \to B$ taka bijekcija, da je funkcija $h^{-1}: B \to A$ zvezno odvedljiva. Nadalje naj bo X zvezno porazdeljena slučajna spremenljivka z gostoto p_X , ki je izven množice A enaka nič. Tedaj je slučajna spremenljivka Y := h(X) porazdeljena zvezno z gostoto:

$$p_Y(y) = \begin{cases} p_X \left(h^{-1}(y) \right) \left| (h^{-1})'(y) \right| & y \in B \\ 0 & sicer \end{cases}$$

Naj boXzvezno porazdeljena slučajna spremenljivka z gostoto p_X , skoncentrirana na odprti množici A. Naj bo $h:A\to\mathbb{R}$ zvezno odvedljiva in $h'(x)\neq 0$ za $\forall x\in A$. Tedaj je slučajna spremenljivka Y = h(X) porazdeljena zvezno z

$$p_Y = \sum_{\substack{x \in A \\ h(x) = y}} \frac{p_X(x)}{|h'(x)|}$$

Naj bosta $A, B \in \mathbb{R}^n$ odprti množici; $h : A \to B$ taka bijekcija, da je h^{-1} parcialno zvezno odvedlijva; X sl. vek. porazdeljen zvezno z gostoto p_X . Tedaj je Y = h(X) po-

$$p_Y(y) = \begin{cases} p_X(h^{-1}(y)|\det J(h^{-1}(y))|) & y \in B\\ 0 & \text{sicer} \end{cases}$$

$$J_g = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \dots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial x_1} & \dots & \frac{\partial h_n}{\partial x_n} \end{bmatrix}$$

Matematično upanje

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots \\ p_1 & p_2 & \dots \end{pmatrix}$$

$$E(X) = \sum_{k=0}^{\infty} x_k p_k \quad \text{\'e} \quad \sum_{k=0}^{\infty} |x_k| p_k < \infty$$

$$E(h(X)) = \sum_{k=0}^{\infty} h(x_k) p_k$$

$$E(h(X,Y)) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} h(x_i, y_j) P(X = x_i, Y = y_j)$$

Zvezno porazdeljena sl. sprem. X z gostoto p_X

$$E(X) = \int_{-\infty}^{\infty} x p_X(x) dx$$
 če $\int_{-\infty}^{\infty} |x| p_X(x) dx < \infty$

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)p_X(x)dx$$

$$E(h(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y)p_{X,Y}(x,y)dxdy$$

Za neko zvezno funkcijo h.

Lastnosti

Če ima |X| mat. up., ga ima tudi X in velja

$$|E(X)| \le E(|X|)$$

Če obstaja $E(X^2)$ in $E(Y^2)$, obstaja tudi E(XY) in velja

$$|E(XY)| \le E(|XY|) \le \sqrt{E(X^2)E(Y^2)}$$

Za poljubne sl. sprem X_1, \ldots, X_n velja:

$$E(a_1X_1 + \dots a_nX_n) = a_1E(X_1) + \dots + a_nE(X_n)$$

Indikator dogodka

$$1_A = \begin{cases} 1 & \text{Ase zgodi} \\ 0 & \text{Ase ne zgodi} \end{cases} \qquad E(1_A) = P(A)$$

Disperzija (varianca)

$$D(X) = E((X - E(X))^{2}) = E(X^{2}) - (E(X))^{2}$$

Lastnosti

- $D(X) \ge 0$
- $D(X) = 0 \iff P(X = E(X)) = 1$
- $D(aX) = a^2D(X)$

Standardna diviacija/odklon:

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

zanjo velja $\sigma(aX) = |a|\sigma(X)$.

Nekoreliranost

Sl. sprem. X in Y sta nekorelirani, če velja:

$$E(XY) = E(X)E(Y)$$

X, Y neodvisni $\implies X, Y$ nekorelirani

Če imata X in Y, je nekoreliranost ekvivalentna zvezi:

$$D(X+Y) = D(X) + D(Y)$$

Kovarianca

$$K(X,Y) = E((X - E(X))(Y - E(Y)))$$

= $E(XY) - E(X)E(Y)$

- K(X,X) = D(X)
- $K(X,Y) = 0 \iff X, Y$ nekorelirani
- K(aX, bY, Z) = aK(X, Z) + bK(Y, Z)
- K(X,Y) = K(Y,X)
- K(aX + b, cY + d) = acK(X, Y)
- $|K(X,Y)| \le \sqrt{D(X)D(Y)}$
- D(X + Y) = D(X) + D(Y) + 2K(X, Y)
- $D(X_1 + \dots + X_n) = D(X_1) + \dots + D(X_n) + 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} K(X_i, X_j)$

Standardizacija

$$X_S = \frac{X - E(X)}{\sigma(X)}$$

Korelacijski koeficient

$$r(X,Y) = \frac{K(X,Y)}{\sigma(X)\sigma(Y)} = E(X_S, Y_S)$$

Lastnosti:

- $r(X,Y) = 0 \iff X, Y$ nekorelirani
- $-1 \le r(X, Y) \le 1$
- $r(X,Y) = 1 \iff P(X_S = Y_S) = 1$
- $r(X,Y) = -1 \iff P(X_S = -Y_S) = 1$
- r(aX + b, cY + d) = r(X, Y)

Pogojne porazdelitve

Pogojna porazdelitev sl. sprem. X glede na dogodek B:

$$X|B \sim \begin{pmatrix} a_1 & a_2 & \dots \\ P(X=a_1|B) & P(X=a_2|B) & \dots \end{pmatrix}$$

Pogojna porazdelitvena funkcija

sl. sprem. X glede na dogodek B:

$$F_{X|B}(x) = F_X(x|B) = P(X \le x|B) = \frac{P((X \le x) \cap B)}{P(B)}$$

Če je pogojna porazdelitev zvezna, objstaja tudi **pogojna** porazdelitvena gostota:

$$p_{X|B}(x) = F'_{X|B}(x)$$

Pogojna gostota

$$p_X(x|Y=y) \equiv p_X(x|y) = \frac{p_{(X,Y)}(x,y)}{p_Y(y)}$$

Pogojno matematično upanje

$$E(h(X)|B) = \sum h(x)P(X = x|B)$$

$$\begin{split} E(X|Y=y) &= \int_{-\infty}^{\infty} x p_{(X|Y)(x|y)} dx \\ &= \frac{1}{p_Y(y)} \int_{-\infty}^{\infty} x p_{(X,Y)(x,y)} dx \end{split}$$

$$E(h(X,Y)|Y=y) = E(h(X,y)|Y=y)$$
$$E(h(X,Y)|Y) = \sum_{x} h(x,Y)P(X=x|Y)$$

$$E(h(X,Y)|Y) = \int_{-\infty}^{\infty} h(x,Y) P_{X|Y}(x|Y) dx$$

Za vsako slučajno spremenlivko X in dogodek B veleja

$$E(X|B) = \frac{E(XZ)}{P(B)} = \frac{E(XZ)}{E(Z)}$$

kjer je sl. sprem. Z indikator dogodka B.

Za vsako sl. sprem. X z mat. up. in popoln sistem dogodkov H_1, H_2, \ldots velja izrek o polni pričakovani vrednosti

$$E(X) = P(H_1)E(X|H_1) + P(H_2)E(X|H_2) + \dots$$

Regresijska funkcija

$$\varphi(y) = E(X|Y=y)$$

Za vsako sl. sprem. X z mat. up. in diskretno sl. sprem. Y

$$E(Xg(Y)|Y) = E(X|Y)g(Y)$$

$$E(Xg(Y)) = E(E(X|Y)g(Y))$$

$$E(X) = E(E(X|Y))$$

Za vsak dododek A in vsako sl. sprem Y velja:

$$E(P(A|Y)) = P(A)$$

Momenti

Moment reda k glede na točko a je

$$m_k(a) = E((X-a)^k)$$
 če obstaja

- Začetni moment $z_k := m_k(0) = E(X^k)$
- Centralni moment $m_k := m_k(E(X)) = E((X X))$
- Faktorski moment reda r: $E(X(X-1)...(X-r+X_n. Potem je:$ $1)) = G_X^{(r)}(1)$

$$z_1 = E(X) \qquad m_2 = D(X)$$

Če obstaja $m_n(a)$, obstaja tudi $m_k(a)$ za $\forall k < n$. Če obstaja z_n , obstaja tudi $m_n(a)$ za $\forall a \in \mathbb{R}$ Centralne momente lahko izračunamo iz začetnih

$$m_n = \sum_{k=0}^{n} {n \choose k} (-1)^{n-k} z_1^{n-k} z_k$$

$$A(X) = E(X_S^3) = E\left(\frac{X - E(X)}{\sigma(X)}\right)^3 = \frac{m_3}{m_2^3}$$

 $\forall \lambda > 0 : A(\lambda X) = A(X)$

Sploščenost (kurtozis)

$$K(X) = E\left[\left(\frac{X - E(X)}{\sqrt{D(X)}}\right)^4\right] = \frac{m_4}{m_2^2}$$

Presežna sploščenost

$$K^*(X) = K(X) - 3$$

Vrstilne karakteristike

Kvantil reda p

je vsaka vrednost x_p , za katero velja:

$$P(X \le x_p) \ge p$$
 in $P(X \ge x_p) = 1 - p$

oz. $F(x_p-) \le p \le F(x_p)$

- Mediana: x_1
- Kvartili: $x_{\frac{1}{4}}, x_{\frac{2}{4}}, x_{\frac{3}{4}}$
- (Per)centili: $x_{\frac{1}{100}}, \dots, x_{\frac{99}{100}}$

Semi interkvartilni razmik

$$s = \frac{1}{2} \left(x_{\frac{3}{4}} - x_{\frac{1}{4}} \right)$$

Rodovne funkcije

Naj bo X sl. sprem. z zalogo vrednosti $\mathbb{N} \cup \{0\}$

$$p_k = P(X = k)$$
 $k = 0, 1, 2, ...$

Rodovna funkcija sl. sprem. X:

 $\forall s \in [-1, 1] : G_X(s) = G_Y(s)$

$$G_X(s) = p_0 + p_1 s + p_2 s^2 + \dots = \sum_{k=0}^{\infty} p_k s^k$$

Obstaja za vse $|s| \leq 1.$

$$P(X = k) = \frac{G_X^{(k)(0)}}{k!}$$

$$G_X(0) = p_0$$
 $G_X(1) = 1$ $G_X(s) = E(s^X)$

$$\iff P(X=k) = P(Y=k) \ \forall k = 0, 1, 2, \dots$$

$$\lim_{s \uparrow 1} G_X'(s) = \lim_{s \uparrow 1} \sum_{k=1}^{\infty} k p_k s^{k-1} = \sum_{k=1}^{\infty} \lim_{s \uparrow 1} k p_k s^{k-1} E(X)$$

Naj bo X sl. sprem. z rodovno funkcijo G_X , potem je:

$$G_X^{(n)}(1-) = E(X)(X-1)(X-2)\dots(X-n+1)$$

Naj bosta X_1, \ldots, X_n nedovisne sl. sprem. z rodovnimi funkcijami $G_{X_1}, \ldots G_{X_n}$:

$$G_{X_1+\cdots+X_n} = G(X_1)\cdots G(X_n) \quad \forall s \in [-1,1]$$

Naj bodo $\forall n \in \mathbb{N}$ sl. sprem N, X_1, \dots, X_n ne
odvisne. Naj ima N rodovno funkcijo G_N in X_i rodovno funkcijo G_X $(X_1,\ldots,X_n \text{ so enako porazdeljene})$. Naj bo $S=X_1+\cdots+X_n$

$$G_S = G_N(G_X(s)) \quad \forall s \in [-1, 1]$$

Velja tudi E(S) = E(N)E(X).

$$G_{2X}(s) = G_X(s^2)$$

Znane rodovne funkcije

$$\sum_{n=0}^{\infty} q^n = \frac{1}{1-q} \quad \sum_{n=0}^{b} q^n = \frac{1-q^{b+1}}{1-q}$$
$$\sum_{n=0}^{\infty} q^n = \frac{q^a}{1-q} \quad \sum_{n=0}^{b} q^n = \frac{q^a-q^{b+1}}{1-q}$$

$$a^{n} - b^{n} = (a - b)(a^{n-1} + a^{n-2}b + \dots + ab^{n-2} + b^{n-1})$$

$$\frac{a_{0} + \dots + a_{k-1}x^{k-1}}{1 - x^{k}} = a_{0} + \dots + a_{k-1}x^{k-1} + a_{0}^{k} + \dots + a_{k-1}x^{2k-1} + \dots$$

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k$$

$$\frac{1}{(1-x)^n} = \sum_{k=0}^n \binom{n+k-1}{k} x^k$$

$$B_{\lambda}(x) = \sum_{n} {\lambda \choose n} x^{n} = (1+x)^{\lambda}; \qquad {\lambda \choose n} = \frac{\lambda^{n}}{n!}$$

Momentno rodovna funkcija

$$M_X(t) = E(e^{tX})$$
 $\forall t \in R$ če obstaja
$$= 1 + z_1 t + \frac{z_2}{2!} t^2 + \frac{z_3}{3!} t^3 + \dots$$

V primeru, ko ima X zalogo vrednosti v $\mathbb{N} \cup \{0\},$ je

$$M_X(t) = E(e^{tX}) = G_X(e^t)$$

Za zvezno porazdeljeno sl. sprem. X velja:

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} p_X(x) dx$$

Naj pri nekem $\delta > 0$ $M_X(t)$ obstaja za vse $t \in (-\delta, \delta)$. Potem je porazdelitev za X natanko določana z M_X in vsi začetni momenti obstajajo:

$$z_k = E(X^k) = M_Y^{(k)}(0) \quad \forall k \in \mathbb{N}$$

$$M_X(t) = \sum_{k=0}^{\infty} \frac{z_k}{k!} t^k \quad \forall t \in (-\delta, \delta)$$

Trditev:

$$M_{aX+b}(t) = e^{bt} M_X(at)$$

Če sta X in Y neodvisni, ie:

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

Izreka o velikih številih

Zaporedje sl. sprem. $\{X_n\}_{n\in\mathbb{N}}$ verjetnostno konvergira proti sl. sprem. X, če

$$\forall \varepsilon > 0 : \lim_{n \to \infty} P(|X_n - X| \ge \varepsilon) = 0$$

$$\forall \varepsilon > 0 : \lim_{n \to \infty} P(|X_n - X| \le \varepsilon) = 1$$

Zaporedje sl. sprem. $\{X_n\}_{n\in\mathbb{N}}$ skoraj gotovo (s.g.) konvergira proti sl. sprem. X, če

$$P(\lim_{n \to \infty} 1) = 1$$

Če $X_n \xrightarrow[n \to \infty]{s.g.} X$, potem

$$\forall \varepsilon > 0$$
: $\lim P(|X_n - X| < \varepsilon, \forall n \ge m) = 1$

Če $X_n \xrightarrow[n \to \infty]{s.g.} X$, potem $X_n \xrightarrow[n \to \infty]{ver.} X$ Naj bo X_1, X_2, X_n, \dots zaporedje sl. sprem. z mat. up. in naj bo $S = X_1 + \dots + X_n$, $Y_n = \frac{S_n - E(S_n)}{n}$. Potem je

 Za $\{X_n\}_{n\in\mathbb{N}}$ velja šibki zakon o velikih številih, če zap $\{Y_n\}_{n\in\mathbb{N}}$ konvergira proti 0 verjetnostno, tj.

$$\forall \varepsilon > 0 : \lim_{n \to \infty} P\left(\frac{|S_n - E(S_n)|}{n} < \varepsilon\right) = 1$$

Za $\{X_n\}_{n\in\mathbb{N}}$ velja **krepki zakon o velikih številih**, če zap. $\{Y_n\}_{n\in\mathbb{N}}$ konvergira proti 0 skoraj gotovo, tj.

$$P\left(\lim_{n\to\infty}\frac{S_n - E(S_n)}{n} = 0\right) = 1$$

Neenakost Markova

Če je X sl. sprem. z mat. up., potem je

$$P(|X| \ge a) \le \frac{E(|X|)}{a} \quad \forall a > 0$$

Neenakost Čebiševa

 \check{C} e ima X disperzijo, je

$$P(|X - E(X)| \ge a\sigma(X)) \le \frac{1}{a^2} \quad \forall a > 0$$

oziroma za $\varepsilon := a\sigma(X)$

 X_1, X_2, \ldots velja pogoj

$$P(|X - E(X)| \ge \varepsilon) \le \frac{D(X)}{\varepsilon^2}$$

Izrek Markova: Če za sl. sprem. $\{X_n\}_{n\in\mathbb{N}}$ velja $\frac{D(S_n)}{n^2}\xrightarrow[n\to\infty]{}0,$ potem velja ŠZVŠ.

 $\mathit{Izrek}\ \check{\mathit{Cibiševa}}$: Če so sl. sprem. X_1, X_2, \ldots paroma nekorelirane in je sup_{$n\in\mathbb{N}$} $D(X_n)<\infty$, potem velja ŠZVŠ. Izrek Kolmogorova Naj za neodvisne slučajne spremenljivke

$$\sum_{n=1}^{\infty} \frac{D(X_n)}{n^2} < \infty$$

potem za $\{X_n\}_{n\in\mathbb{N}}$ velja KZVŠ. Zgornji pogoj velja, če je $\sup_n D(X_n) < \infty$

Naj bodo X_1, X_2, \ldots neodvisne in enako porazdeljene sl. sprem. z disperzijo. Potem velja KZVŠ.

Centralni limitni izrek

Naj bo X_1, X_2, X_n, \ldots zaporedje sl. sprem. z mat. up. in naj bo $S = X_1 + \cdots + X_n, Z_n = \frac{S_n - E(S_n)}{n}$. Potem je $E(Z_n) = 0$ in $D(Z_n) = 1$.

Za $\{X_n\}_{n\in\mathbb{N}}$ veljaj **centralni limitni zakon**, če

$$F_{Z_n}(x) \xrightarrow[n \to \infty]{} F_{N(0,1)} \quad \forall x \in \mathbb{R}$$

Če so X_1, X_2, \ldots neodvisne in enako porazdeljene, velja centralni limitni zakon. Za velike n je $S_n \sim N(E(S_n), \sigma(S_n))$

$$P\left(\frac{S_n - E(S_n)}{\sigma(S_n)} \le x\right) \xrightarrow[n \to \infty]{} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{\frac{t^2}{2}} dt$$

$$P(a \le S_n \le b) \approx \Phi\left(\frac{b - E(S_n)}{\sigma(S_n)}\right) - \Phi\left(\frac{a - E(S_n)}{\sigma(S_n)}\right)$$

Izrek o zveznosti rodovne funkcije

Naj za zaporedje $\{Z_n\}_{n\in\mathbb{N}}$ sl. sprem. velja

$$M_{Z_n}(t) \to M_{N(0,1)}(t) = e^{\frac{t^2}{2}} \quad \forall t \in (-\delta, \delta)$$

potem

$$F_{Z_n}(x) \to F_{N(0,1)}(x) \qquad \forall x \in \mathbb{R}$$

Funkcija Γ

• $\Gamma(1) = 1$

• $\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} dx$, $\forall s > 0$

- $\Gamma(s+1) = s\Gamma(s)$ • $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
- $\Gamma(n+1) = n!$ $n \in \mathbb{N}$
- $\Gamma(x)\Gamma(x+1) = \frac{\pi}{\sin(\pi x)}$

Funkcija B

• $B(p,q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx$,

$$\bullet \ B(p,q) = \int_0^\infty \frac{u^{p-1}}{(1+u)^{p+q}} du$$

•
$$B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

•
$$\frac{1}{2}B(p,q) = \int_0^{\frac{\pi}{2}} (\sin x)^{2p-1} (\cos x)^{2q-1}$$

• simetričnost:
$$B(p,q) = B(q,p)$$

Osnovni pojmi statistike

Na verjetnostnem prostoru (Ω, \mathcal{F}) imamo sl. sprem X. Vzorec velikosti n je sl. vektor (X_1, \ldots, X_n) , kjer so X_i paroma neodvisni in porazdeljeni kot X.

 (x_1,\ldots,x_n) . To so konkretni podatki, ki jih analiziramo. Ocene za μ (mat. up. sl. sprem X)

- vzorčno povprečje: $\overline{x} = \frac{x_1 + \dots + x_n}{n}$
- vzorčni modus: najpogostejša vrednost
- vzorčna mediana: srednja vrednost v po velikosti urejenem vzorcu

Ocene za σ (standardna diviacija X)

- vzorčni razmak: $\max(x) \min(x)$
- vzorčna disperzija: $s_0^2 = \frac{1}{n} \sum_{i=1}^n (\overline{x} x_i)^2$
- vzorčna diviacija: $s_0 = \sqrt{s_0^2}$
- popravljena vzorčna disperzija: $s^{2} = \frac{1}{n-1} \sum_{i=1}^{n} (\overline{x} - x_{i})^{2} = \frac{n}{n-1} s_{0}^{2}$
- popravljena vzorčna diviacija: $s = \sqrt{s^2}$

Vzorčne statistike in cenilke

Naj bo (X_1, \ldots, X_n) vzorec velikosti n.

Vzorčna statistika je simetrična funkcija vzorca:

$$Y = Y_n = g(X_1, \dots, X_n)$$

Praviloma vzorčna statistika ocenjuje nek parameter ζ . Tedaj je Y **cenikla** parametra.

Y je nepristranska cenilka, če $E(Y) = \zeta$, sicer pristarn**skost** merimo kot $B(Y) = E(Y) - \zeta$.

Y je **dosledna** cenilka, če $Y_n \xrightarrow[n \to \infty]{ver.} \zeta$ oziroma $\forall \varepsilon > 0$: $\lim_{n\to\infty} P(|Y_n - \zeta| < \varepsilon) = 1.$

Standardna napaka vzorčne statistike SE(Y) je standardna diviacija $\sigma(Y)$.

Srednja kvadratična napaka: $MSE(Y) = E((Y - \zeta)^2) =$ $D(Y) + B(Y)^2$.

Naj bo Y_n cenilka za ζ . Če je $E(Y_n) \xrightarrow[n \to \infty]{} \zeta$ in $D(Y_n) \xrightarrow[n \to \infty]{} 0$, potem je Y_n dosledna cenilka za ζ .

Vzorčna statistika χ^2

$$\chi^{2} = \frac{1}{\sigma^{2}} \sum_{i=1}^{n} (X_{i} - \overline{X})^{2} = \frac{n}{\sigma^{2}} S_{0}^{2} = \frac{n-1}{\sigma^{2}} S^{2}$$

$$\chi^{2} \sim \chi^{2} (n-1)$$

Studentova porazdelitev

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

Metode za pridobivanje cenilk

Momentna metoda

Naj bo (X_1,\ldots,X_n) vzorec velikosti n in $k\in\mathbb{N}$. **Vzorčni** k-ti moment

$$Z_k = \frac{1}{n} \left(X_1^k + X_n^k \right)$$

je nepristranska dosledna cenikla za $z_k = E(X^k)$

Naj bo X zvezno porazdeljena z gostoto $p(x;\xi_1,\ldots,\xi_n)$ in naj obstajajo začetni momenti $z_k = E(X^k) =$ $\int_{-\infty}^{\infty} x^k p(x; \xi_1, \dots, \xi_n) dx \text{ za } k = 1, \dots, m. \text{ Denimo, da se iz}$ teh enačb da izraziti parametre $\xi_k = \varphi_k(z_1, \dots, z_m)$. Potem je $C_k = \varphi(Z_1, \ldots, Z_m)$ cenilka za parameter ξ_k .

Metoda maksimalne zaneslijvosti (verjetja)

Naj bo gostota odvisna od parametra ξ : $p(x,\xi)$. Funkcija Test T zaneslijivosti:

$$L(x_1,\ldots,x_n,\xi)=p(x_1,\xi)\ldots p(x_n,\xi)$$

Pri danih x_1, \ldots, x_n izberemo ξ , da je dosežen maksimum

Ta vrednost je odvisna le od x_1,\ldots,x_n , torej $\xi_{\max}=$ $\varphi(x_1,\ldots,x_n)$. Tako dobimo cenilko za ξ :

$$C = \varphi(X_1, \dots, X_n)$$

Intervalsko ocenjevanje parametrov

Naj bo gostota sl. spremXodvisna od parametra ξ in naj bo (x_1,\ldots,x_n) vzorec.

Interval [A, B] (ki je odvisen le od vzorca) je inteval zau-Vrednost tega sl. vektorja pri enem naboru meritev je panja za parameter ξ pri **stopnji tveganja** $\alpha \in [0,1]$, če je

$$P(\xi \in [A, B]) = 1 - \alpha$$
 $P(\xi \notin [A, B]) = \alpha$

Številu $1-\alpha$ rečamo **stopnja zaupanja** A in B pa sta $H_1^{\pm}:\mu_X\neq\mu_Y$ $K_{\alpha}=(-\infty,-t_{\alpha/2}(n-1)]\cup[z_{\alpha/2}(n-1),\infty)$ Potem je vzorčni statistiki.

Waldov interval zaupanja

n neodvisnih poskusov vsak uspe z verjetnostjo p. Opazimo, da uspe S poskusov.

$$\hat{p} = \frac{S}{n}$$
 $c = z_{\alpha/2} = \Phi^{-1}\left(\frac{1-\alpha}{2}\right)$

Waldov interval zaupanja za p:

$$\hat{p} - c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Ta interval zaupanja ni preveč natančen. Računamo le na toliko mest kot jih ima n in enice ne štejemo za mesto.

Ocenjevanje μ in σ

Opazimo $X_1, \ldots, X_n \sim N(\mu, \sigma)$

• Zanima nas μ , σ je znan.

$$\frac{\overline{X} - \mu}{\sigma} \sqrt{n} \sim N(0, 1) \qquad \Delta = \frac{c\sigma}{\sqrt{n}}$$

$$c = z_{\alpha/2} = \Phi^{-1} \left(\frac{1 - \alpha}{2}\right) \quad \overline{X} - \Delta < \mu < \overline{X} + \Delta$$

• Zanima nas μ , σ ni znan. Za σ vzamemo cenilko s.

$$\begin{split} \overline{X} - \mu \\ s & \sqrt{n} \sim \text{Student}(n-1) & \Delta = \frac{c\sigma}{\sqrt{n}} & \text{\'e H_0 velja, je $T \sim$} \\ c &= t_{\alpha/2}(n-1) & \overline{X} - \Delta < \mu < \overline{X} + \Delta & \text{Test χ^2 (Pearson)} \end{split}$$

• Zanima nas σ , μ ni znan.

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

$$c_1 = \chi_{1-\alpha/2}^2(n-1)$$
 $c_2 = \chi_{\alpha/2}^2(n-1)$ $s\sqrt{\frac{n-1}{c_2}} < \sigma < s\sqrt{\frac{n-1}{c_1}}$

Preizkušanjue hipotez

 Hipoteza je **enostavna**, če natančno določa porazdelitev, $\,$ Pri velikemnima statistka sicer pa je sestavljena.

Izberemo stopnjo značilnosti α , to je verjstnost, da zavrnemo pravilno hipotezo. Testi značilnosti nam povejo ali pri dani α in vzorčni vrednosti zavrnemo hipotezo ali ne.

 $X \sim N(\mu, \sigma)$, σ poznamo, μ_0 dano število

$$H_0(\mu = \mu_0) : H_1(\mu \neq \mu_0)$$

Testna statistika, ki odloča o zavrnitvi hipoteze:

$$Z = \frac{\overline{X} - \mu_0}{5} \sqrt{n}$$

Če velja H_0 , je $Z \sim N(0,1)$.

Iz tabele razberemo $z_{\alpha/2} > 0$, da je $P(Z > z_{\alpha/2}) = \alpha/2$. Hipotezo H_0 zavrnemo, če vzorčna vrednost za Z leži na kritičnem območju

$$K_{\alpha} = (-\infty, -z_{\alpha/2}] \cup [z_{\alpha/2}, \infty)$$

 $X \sim N(\mu, \sigma)$, σ ni znan, μ_0 dano število

$$H_0(\mu = \mu_0) : H_1(\mu \neq \mu_0)$$

Testna statistika, kjer je s popravljena vzorčna disperzija:

$$T = \frac{\overline{X} - \mu_0}{s} \sqrt{n}$$

Če velja H_0 , je $T \sim Student(n-1)$.

Vzorci po parih

Imamo neodvisne pare $(X_1,Y_1),\ldots,(X_n,Y_n)$ z (X_i,Y_i) ~ $N(\mu_X, \mu_y, \sigma_X, \sigma_Y, \rho)$. Preverjamo $H_0: \mu_X = \mu_Y$. Pare pretvorimo v razlike: $R_i = X_i - Y_i$ in na njih naredimo T test za $\mu_0 = 0$.

$$\begin{split} H_1^X : & \mu_X > \mu_Y \quad K_\alpha = [t_\alpha(n-1), \infty) \\ H_1^Y : & \mu_X < \mu_Y \quad K_\alpha = (-\infty, -t_\alpha(n-1)] \\ H^{\pm} : & \mu_X \neq \mu_Y \quad K_\alpha = (-\infty, -t_\alpha(n-1)] \\ \end{bmatrix}$$

P - vrednost

je najmanjša stopnja značilnosti α pri kateri še lahko zavrnemo hipotezo (pri danih vzorčnih podatkih).

Studentov primerjalni test

Imamo 2 neodvisna vzorca velikosti m in n. Prvi je vzet iz populacije na kateri imamo sl. sprem $X \sim N(\mu_X, \sigma)$, drugi pa je vzet iz populacije na kateri imamo sl. sprem $Y \sim N(\mu_Y, \sigma)$.

Paramatrov μ_X,μ_Y,σ ne poznamo.

 Naj bosta S_X^2 in S_Y^2 popravljeni vzorčni disperziji za (X_1,\ldots,X_m) in (Y_1,\ldots,Y_n) .

Skupna vzorčna varianca je

$$S^{2} = \frac{(m-1)S_{X}^{2} + (n-1)S_{X}^{2}}{m+n-2}$$

Testiramo hipotezo $H_0(\mu_X = \mu_Y) : H_1(\mu_X \neq \mu_Y)$. Testna statistika:

$$T = \frac{\overline{X} - \overline{Y}}{S} \sqrt{\frac{mn}{m+n}}$$

Če H_0 velja, je $T \sim \text{Student}(m+n-2)$.

Naj ima sl. sprem \boldsymbol{X} porazdelitveno funkcijo $\boldsymbol{F},$ ki ni znana. Preizkušamo domnevo o tipu porazdelitvenega zakona:

$$H_0(F = F_0) : H_1(F \neq F_0)$$

kjer je F_0 dana porazdelitvena funkcija.

Zalogo vrednosti sl. sprem X razdelimo na r razredov: S_1, \ldots, S_r , da je $p_k = P(X \in S_k | H_0) > 0$ za $\forall k = 1, \ldots, r$. Naj bo (X_1, \ldots, X_n) slučajni vzorec za sl. sprem X in N_k število vrednosti vzorca, ki padejo v S_k .

$$N_k \sim \text{Bin}(n, p_k), \qquad k = 1, \dots, r$$

$$\chi^2 = \sum_{r=1}^{r} \frac{(\overbrace{N_k}^{\text{opažena f.}} - \overbrace{np_k}^{\text{pričakovana f.}})^2}{nn_{l.}}$$

približno porazdelitev $\chi^2(r-1)$

Iz tabele razberemo $c_{\alpha} > 0$, da je $P(\chi^2 > c_{\alpha}) = \alpha$. Kritično območje: $K_{\alpha} = [c_{\alpha}, \infty)$.

Trditev: Če so v testu χ^2 verjetnosti p_k odvisne od parame $tra \theta$, potem ima statistika

$$\chi^{2} = \sum_{k=1}^{r} \frac{(N_{k} - np_{k}(\hat{\theta}))^{2}}{np_{k}(\hat{\theta})}$$

porazdelitev približno $\chi^2(r-2),$ kjer je $\hat{\theta}$ cenilka za θ po metodi maksimalne zaneslijivosti.

Linearna regresija

Linearni regresijski model

$$Y = a + bx + U$$

Pri fiksnem x predpostavimo, da Y = a + bx + U, kjer sta a,b konstanti in $U \sim N(0,\sigma)$. Z drugimi besedami $Y \sim N(a + bx, \sigma)$. y = a + bx je regresijska premica. Za različne vrednosti x_1, \ldots, x_n dobimo slučajni vektor (Y_1,\ldots,Y_n) , kjer je $Y_k \sim N(a+bx_k,\sigma)$. Radi bi ocenili vrednost a in b.

$$\hat{b} = \frac{\sum_{k=1}^{n} (x_k - \overline{x})(Y_k - \overline{Y})}{\sum_{k=1}^{n} (x_k - \overline{x})^2} \qquad \hat{a} = \overline{Y} - \hat{b}\overline{x}$$

vperjemo vsote:
$$S_{xx} = \sum_{k=1}^{n} x_k^2 \quad S_{xY} = \sum_{k=1}^{n} x_k Y_k \quad S_x = \sum_{k=1}^{n} x_k \quad S_Y = \sum_{k=1}^{n} Y_k \quad P(S^+ \le k_\alpha) = \sum_{k=0}^{k_\alpha} p_k \le \frac{\alpha}{2} \quad \text{in} \quad P(S^+ \le k_\alpha + 1) = \sum_{k=0}^{k_\alpha + 1} p_k > \frac{\alpha}{2}$$

$$\hat{b} = \frac{nS_{xY} - S_x S_Y}{nS_{xx} - S_x^2}$$
 $\hat{a} = \frac{1}{n}S_Y - \hat{b}\frac{1}{n}S_x$

Prilagoditveni test

To je poseben primer Pearsonevega testa χ^2 H_0 : dogodka A in B sta neodvisna.

$$p = P(A)$$
 $q = P(B)$

Če sta p in q znana, uporabimo test χ^2 z r=4

$$\begin{split} \chi^2 &= \frac{(N_{A \cap B} - npq)^2}{npq} + \frac{(N_{A \cap B} \mathfrak{C} - np(1-q))^2}{np(1-q)} + \\ &+ \frac{(N_{A} \mathfrak{C}_{\cap B} - n(1-p)q)^2}{n(1-p)q} + \frac{(N_{A} \mathfrak{C}_{\cap B} \mathfrak{C} - n(1-p)(1-q))^2}{n(1-p)(1-q)} \end{split}$$

Iz tabele razberemo c_{α} , da je $P(\chi^2 > c_{\alpha}) = \alpha$. Kritično območje: $K_{\alpha} = [-c_{\alpha}, c_{\alpha}]$

Če pap in qnista znana, ju ocenimo iz podatkov:

$$\hat{p} = \frac{N_{A \cap B} + N_{A \cap B} \mathbf{c}}{n} \qquad \quad \hat{q} = \frac{N_{A \cap B} + N_{A} \mathbf{c}_{\cap B}}{n}$$

Testna statistkia χ^2 je potem porazdeljena $\chi^2(1)$.

Prilagoditveni test lahko posplošimo na večje kontingenčne tabele. Denimo, da prva karakteristika določa r kategorij (A_1, \ldots, A_r) , druga pa s kategorij (B_1, \ldots, B_s) .

Naj bo $p_i = P(A_i)$ in $q_i(B_i)$.

Naj bo X_{ij} opažena frekvenca kategorije A_i in B_j . Pričakovana frekvenca za X_{ij} je $n\hat{p}_i\hat{q}_j$.

$$\hat{p}_i = \frac{1}{n} \sum_{j=1}^s X_{ij}$$
 ... cenika za p_i

$$\hat{q}_j = \frac{1}{n} \sum_{i=1}^r X_{ij}$$
 ... cenika za q_j

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(X_{ij} - n\hat{p}_i\hat{q}_j)^2}{n\hat{p}_i\hat{q}_j}$$

$$\chi \sim \chi \; ((r-1)(s-1))$$

Pri dani stopnji zaupanja α iz tabele razberemo c_{α} , da

 $P(\chi^2 > c_\alpha) = \alpha$. Kritično območje: $K_\alpha = [c_\alpha, \infty)$ $\check{C}e \ je \ n\hat{p}\hat{q} < 5 \ za \ kak \ i,j \ moramo \ združiti \ nekatere \ razrede.$

Neparametrični testi

Test z znaki

analoa testa T

Testiramo hipotezo: $H_0(F_X = F_Y)$

Definirajmo razliko $D_i = X_i - Y_i$ za i = 1, ..., n. Privza- $H_1^{\pm}(F_X \neq F_Y) : K_{\alpha} = (-\infty, z_{\alpha/2}] \cup [z_{\alpha/2}, \infty)$ memo, da so vzorčne vrednosti $D_i \neq 0$, sicer jih izpustimo in zmanjšamo n.

Če velja
$$H_0$$
, je $P(D_i > 0) = \frac{1}{2} = P(D_i < 0)$.

 $\binom{n}{k} 2^{-k}$.

Pri dani stopnji značilnosti α je kritično območje:

$$K_{\alpha} = \{k \in \{0, \dots, n\} : k \le k_{\alpha} \text{ ali } k \ge n - k_{\alpha}\}$$

kjer je k_{α} doočena z zahtevama:

$$P(S^{+} \le k_{\alpha}) = \sum_{k=0}^{k_{\alpha}} p_{k} \le \frac{\alpha}{2}$$
 in $P(S^{+} \le k_{\alpha} + 1) = \sum_{k=0}^{k_{\alpha} + 1} p_{k} > \frac{\alpha}{2}$

Za velike n je $S^+ \approx N(\frac{n}{2}, \sqrt{n\frac{1}{2}\frac{1}{2}})$.

Dodatek iz vaj

Ničelna domneva $H_0: P(X > Y) = P(Y > X)$. Alter-

•
$$H_1^X: P(X > Y) > P(Y > X)$$

$$p(k, l) = P(S' > k) = P(S' < l)$$

•
$$H_1^Y: P(Y > X) > P(X > Y)$$

$$p(k,l) = P(S' \ge l) = P(S' \le k)$$

• H_1^{\pm} : velia bodisi H_1^X bodisi H_1^Y

$$p(k, l) = 2\min(P(S' \ge k), P(S' \ge k))$$

= $2\min(P(S' \le k), P(S' \le k))$

 S_X št. meritev kjer je $X > Y, S_Y$ št. meritev kjer je X < Y, $\hat{n} = S_X + S_Y, S' \sim \text{Bin}(\hat{n}, 1/2)$

Če je $\hat{n} > 10$ lahko H_0 zavrnemo: • proti H_1^X , če $\frac{S_x - S_Y - 1}{\sqrt{s}} \ge z_\alpha$

 H_0 zavrnemo če je $p(S_X, S_Y) \leq \alpha$.

• proti
$$H_1^Y$$
, če $\frac{S_x - S_Y - 1}{z} < -z_\alpha$

• proti
$$H_1^{\pm}$$
, če $\frac{|S_x - S_Y| - 1}{\sqrt{s}} < -z_{\alpha/2}$

Inverzijski test

Analog primerjalnega T testa

Naj bosta X in Y sl. sprem s porazdelitvenima funkcijama F_X in F_Y . Vzorca (X_1, \ldots, X_m) in (Y_1, \ldots, Y_n) sta neodvisna. Predpostavimo, da $m \leq n$. Vzorčne vrednosti x_1, \ldots, x_m in y_1, \ldots, y_n razvrstimo po velikosti: $z_1 \leq z_2 \leq$

$$R(x) = \operatorname{rang}(x) = \frac{\#\{i; z_i < x\} + \#\{i; z_i \le x\} + 1}{2}$$

Če je več zaporednih vrednosti enakih, za rang vzamemo povprečno mesto.

$$V = R_1 + \dots + R_m$$

Če velja $H_0(F_X = F_Y)$ in $m+n \ge 20$ in m, n, je V približno normalno porazdeliena:

$$V \sim N\left(\frac{(m+n+1)m}{2}, \sqrt{\frac{mn(m+n+1)}{12}}\right)$$

$$Z = \sqrt{\frac{3}{mn(m+n+1)}}(2V - m(m+n+1)) \sim N(0,1)$$

 $H_1^X(X \text{ je stoh. strog. več od } Y(F_X < F_Y)) : K_\alpha = [z_\alpha, \infty)$ $H_1^Y(Y \text{ je stoh. strog. več od } X(F_X > F_Y)) : K_\alpha = (-\infty, z_\alpha]$

Inverzija med X_i in Y_i se pojavi, če ima Y_i manjši rang

Naj bo U št. inverzij v zaporedju z_1, \ldots, z_{n+m} . Potem je Tedaj je $S^+ \sim \text{Bin}(n, \frac{1}{2})$, torej je $p_k = P(S^+ = k) = V = U + \frac{m(m+1)}{2}$, ker vsaka inverzija poveča vsoto rangov

je
$$nS_{xY} - S_xS_Y$$
 $\hat{n} = 1$ \hat{n} \hat{n} \hat{n} \hat{n}

Testiranje neodvisnosti

$$p = P(A) q = P(B)$$

$$\frac{\text{kategorije}}{\text{verjetnost}} \begin{vmatrix} A \cap B & A \cap B^{\complement} & A^{\complement} \cap B & A^{\complement} \cap B^{\complement} \\ pq & p(1-q) & (1-p)q & (1-p)(1-q) \end{vmatrix}$$

$$\chi^{2} = \frac{(N_{A \cap B} - npq)^{2}}{npq} + \frac{(N_{A \cap B} \mathbf{C} - np(1-q))^{2}}{np(1-q)} + \frac{(N_{A} \mathbf{C}_{\cap B} - n(1-p)q)^{2}}{n(1-p)q} + \frac{(N_{A} \mathbf{C}_{\cap B} \mathbf{C} - n(1-p)(1-q))^{2}}{n(1-p)(1-q)}$$

$$N_{A \cap B} + N_{A \cap B}$$
 $N_{A \cap B} + N_{A \cap B}$

 H_0 : A_i in B_i sta neodvisna.

$$\hat{p}_i = rac{1}{s} \sum_{i=1}^s X_{ij} \ldots$$
 cenika za p_i

$$\sum_{s} \sum_{s}^{r} (X_{ij} - n\hat{p}_i\hat{q}_i)^2$$

$$\chi^2 \sim \chi^2((r-1)(s-1))$$

Na populaciji imamo sl. sprem. X s porazdelitveno fun. F_X in sl. sprem. Y s porazdelitveno fun. F_Y .

Obravnavamo sl. vektorja (X_1, \ldots, X_n) in (Y_1, \ldots, Y_n) , kjer X_i in Y_i dobimo na istem elementu populacije.

Naj bo
$$S^+$$
 število pozitivnih D_i , S^- pa število negativnih.