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	$x^x(1+\ln x)$
$\ln(x)$	$\frac{1}{x}$
$\log_a(x)$	$\frac{1}{x\ln(a)}$
$\sin(x)$	$cos(x)$
$\cos(x)$	$-sin(x)$
$\tan(x)$	$\frac{1}{cos^2(x)}$
$\cot(x)$	$-\frac{1}{sin^2(x)}$
$\arcsin(x)$	$\frac{1}{\sqrt{1-x^2}}$
$\arccos(x)$	$-\frac{1}{\sqrt{1-x^2}}$
$\arctan(x)$	$\frac{1}{1+x^2}$
$\mathrm{arccot}(x)$	$-\frac{1}{1+x^2}$
$\mathrm{sh}(x) = \frac{e^x-e^{-x}}{2}$	$\mathrm{ch}(x)$
$\mathrm{ch}(x) = \frac{e^x+e^{-x}}{2}$	$\mathrm{sh}(x)$
$\mathrm{th}(x) = \frac{\mathrm{sh}(x)}{\mathrm{ch}(x)}$	$\frac{1}{\mathrm{ch}^2(x)}$
$\mathrm{cth}(x) = \frac{1}{\mathrm{th}(x)}$	$-\frac{1}{\mathrm{sh}^2(x)}$
$\mathrm{arsh}(x) = \ln(x+\sqrt{x^2+1})$	$\frac{1}{\sqrt{1+x^2}}$
$\mathrm{arch}(x) = \ln(x+\sqrt{x^2-1})$	$\frac{1}{\sqrt{1-x^2}}$
$\mathrm{arth}(x) = \frac{1}{2}\ln\frac{1+x}{1-x}$	$\frac{1}{(1+x)(1-x)}$

Osnove kombinatorike

Rodovne funkcije

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

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

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

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

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

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

$$B_\lambda(x) = \sum_n \binom{\lambda}{n}x^n = (1+x)^\lambda; \qquad \binom{\lambda}{n} = \frac{\lambda^{\underline{n}}}{n!}$$

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*	$\binom{n+k-1}{k}$	$\binom{n}{k}$

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

Pravila za računanje z dogodki

$$\begin{array}{rl}
\text{idempotentnost} & A\cup A = A = A\cap A \\
\text{komutativnost} & A\cup B = B\cup A \\
 & A\cap B = B\cap A \\
\text{asociativnost} & (A\cup B)\cup C = A\cup(B\cup C) \\
 & (A\cap B)\cap C = A\cap(B\cap C) \\
\text{distibutivnost} & (A\cup B)\cap C = (A\cap C)\cup(A\cap C) \\
 & (A\cap B)\cup C = (A\cup C)\cap(A\cup C) \\
\text{De Morgan} & (\bigcup_{i\in I} A_i)^{\complement} = \bigcap_{i\in I} A_i^{\complement} \\
 & (\bigcap_{i\in I} A_i)^{\complement} = \bigcup_{i\in I} A_i^{\complement} \\
 & A\cup\Omega = \Omega \quad A\cap\Omega = A \\
 & A\cup\emptyset = A \quad A\cap\emptyset = \emptyset \\
 & A\cup A^{\complement} = \Omega \quad A\cap A^{\complement} = \emptyset
\end{array}$$

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

Če 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

$$\bigcup_i A_i = \Omega \qquad A_i\cup A_j = \emptyset, \forall i,j: i\neq j$$

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

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

Lastnosti P:

- $P(\emptyset) = 0$
- P je končno aditivna.
- P je *monotona*: $A\subseteq B \implies P(A)\leq P(B)$
- $P(A^{\complement}) = 1-P(A)$
- 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 $H_1, H_2, H_3, \dots$ 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

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

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{P(H1)P(A|H1)+P(H2)P(A|H2)+\dots}$$

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 $P(A|B) = P(A|B^{\complement})$. Dogodki $A1, A2, A3, \dots$ so neodvisni, če za poljubne različne indekse $i_1, i_2, \dots, i_k$ velja:

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

Slučajne spremenljivke

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

Diskretne porazdelitve

Diskretna enakomerna porazdelitev na n točkah:

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

Binomska porazdelitev

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

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

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

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

Geometrijske porazdelitev

$\mathrm{Geo}(p)$, $p\in(0,1)$

$(X=k)$ 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 ponovitvi.

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

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

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: $X\sim\mathrm{Hip}(s,r,n) = \mathrm{Hip}(r,s,n)$. Velja

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

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 $\mathrm{Bin}(n,p)\approx\mathrm{Poi}(np)$

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

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

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

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

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 poljubna $a\leq b$ velja:

$$P(a\leq X\leq b) = \int_a^b p_X(t)dx$$

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 x velja $p_x(x) = F_X'(x)$.

Slučajni vektorji

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

Porazdelitvena funkcija slučajnega vektorja

$$F_{(X_1,\dots,X_n)}(x_1,\dots,x_n) = P(X_1\leq x_1,\dots,X_n\leq x_n)$$

Neodvisnot slučajnih spremenljivk

Slučajne spremenljivke so neodvisne, če je

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

Torej so dogodki $(X_1\leq x_1),\dots,(X_n\leq x_n)$ neodvisni.

Naj bo (X,Y) diskretno porazdeljen sluč. vektor.

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

potem velja:

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

Naj bo (X,Y) zvezno porazdeljen sluč. vektor z gostoto $p_{(X,Y)}(x,y)$, potem velja:

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

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 bo X zvezno 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 gostoto:

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

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 $Y = h(X)$ porazdeljen:

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

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

Matematično upanje

Diskretno porazdeljena sl. sprem.

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

$$E(X) = \sum_{k=1}^{\infty} x_k p_k \quad \text{če} \quad \sum_{k=1}^{\infty} |x_k| p_k < \infty$$

$$E(h(X)) = \sum_{k=1}^{\infty} h(x_k)p_k$$

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

Zvezno porazdeljena sl. sprem. X z gostoto p_X

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

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

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

Za neko zvezno funkcijo h.

Lastnosti

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

$$|E(X)|\leq E(|X|)$$

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

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

Za poljubne sl. sprem $X_1,\dots,X_n$ velja:

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

Indikator dogodka

Indikator dogodka A je sl. sprem.:

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

Disperzija (varianca)

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

Lastnosti:

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

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:

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

$$X, Y \text{ neodvisni} \implies X, Y \text{ nekorelirani}$$

Če imata X in Y, je nekoreliranost ekvivalentna zvezi:

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

Kovarianca

$$\begin{aligned} K(X,Y) &= E((X - E(X))(Y - E(Y))) \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

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

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

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:

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

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

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

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

$$E(h(X,Y)|Y=y) = E(h(X,y)|Y=y)$$

$$E(h(X,Y)|Y) = \sum_x h(x,Y)P(X = x|Y)$$

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

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 $H_1, H_2, \dots$ 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 velja:

$$\begin{aligned} E(Xg(Y)|Y) &= E(X|Y)g(Y) \\ E(Xg(Y)) &= E(E(X|Y)g(Y)) \\ E(X) &= E(E(X|Y)) \end{aligned}$$

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

$$m_k(a) = E((X - a)^k) \quad \textit{če obstaja}$$

- **Začetni moment** $z_k := m_k(0) = E(X^k)$
- **Centralni moment** $m_k := m_k(E(X)) = E((X - E(x))^k)$
- **Faktorski moment** reda r: $E(X(X-1)\dots(X-r+1)) = G_X^{(r)}(1)$

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

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:

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

Asimetrija

$$A(X) = E(X_S^3) = E\left(\left(\frac{X - E(X)}{\sigma(X)}\right)^3 \right) = \frac{m_3}{m_2^{\frac{3}{2}}}$$

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

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:

$$P(X \leq x_p) \geq p \text{ in } P(X \geq x_p) = 1 - p$$

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

- Mediana: $x_{\frac{1}{2}}$
- 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 $\mathbb{N} \cup \{0\}$:

$$p_k = P(X = k) \qquad k = 0, 1, 2, \dots$$

Rodovna funkcija sl. sprem. X:

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

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

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

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

Izrek o enoličnosti:

$$\begin{aligned} &\forall s \in [-1,1] : G_X(s) = G_Y(s) \\ &\iff P(X = k) = P(Y = k)\ \forall k = 0, 1, 2, \dots \end{aligned}$$

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

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

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

Naj bodo $\forall n \in \mathbb{N}$ sl. sprem $N, X_1, \dots, X_n$ neodvisne. Naj ima N rodovno funkcijo G_N in X_i rodovno funkcijo G_X ($X_1, \dots, X_n$ so enako porazdeljene). Naj bo $S = X_1 + \dots + X_n$. Potem je:

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

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

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

Znane rodovne funkcije

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

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

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

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

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

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

$$B_\lambda(x) = \sum_n \binom{\lambda}{n} x^n = (1+x)^\lambda; \qquad \binom{\lambda}{n} = \frac{\lambda^{\underline{n}}}{n!}$$

Momentno rodovna funkcija

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

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:

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

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

Trditev:

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

Če sta X in Y neodvisni, je:

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

Izreka o velikih številih

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

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

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

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

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

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

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

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 $E(Y_n) = 0$.

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

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

Neenakost Čebiševa

Če ima X disperzijo, je

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

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

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

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.

Izrek Čibiševa: Če so sl. sprem. $X_1, X_2, \dots$ paroma nekorelirane in je $\sup_{n \in \mathbb{N}} D(X_n) < \infty$, potem velja ŠZVŠ.

Izrek Kolmogorova Naj za neodvisne slučajne spremenljivke $X_1, X_2, \dots$ velja pogoj

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

potem za $\{X_n\}_{n \in \mathbb{N}}$ velja KZVŠ.

Zgornji pogoj velja, če je $\sup_n D(X_n) < \infty$

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

Centralni limitni izrek

Naj bo $X_1, X_2, X_n, \dots$ zaporedje sl. sprem. z mat. up. in naj bo $S = X_1 + \dots + X_n$, $Z_n = \frac{S_n - E(S_n)}{n}$. 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 $X_1, X_2, \dots$ neodvisne in enako porazdeljene, velja centralni limitni zakon. *Za velike* n *je* $S_n \sim N(E(S_n), \sigma(S_n))$

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

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

Izrek o zveznosti rodovne funkcije

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

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

potem

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

Funkcija Γ

- $\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} dx, \qquad \forall s > 0$
- $\Gamma(1) = 1$
- $\Gamma(s+1) = s\Gamma(s)$
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
- $\Gamma(n+1) = n! \qquad n \in \mathbb{N}$
- $\Gamma(x)\Gamma(x+1) = \frac{\pi}{\sin(\pi x)}$

Funkcija B

- $B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1} dx, \qquad \forall p, q > 0$
- $B(p,q) = \int_0^\infty \frac{u^{p-1}}{(1+u)^{p+q}} du$
- $B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$
- $\frac{1}{2}B(p,q) = \int_0^{\frac{\pi}{2}} (\sin x)^{2p-1} (\cos x)^{2q-1}$
- simetričnost: $B(p,q) = B(q,p)$

Osnovni pojmi statistike

Na verjetnostnem prostoru $(\Omega, \mathcal{F})$ imamo sl. sprem X.

Vzorec velikosti n je sl. vektor $(X_1, \dots, X_n)$, kjer so X_i paroma neodvisni in porazdeljeni kot X.

Vrednost tega sl. vektorja pri enem naboru meritev je $(x_1, \dots, x_n)$. 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 $(X_1, \dots, X_n)$ 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 **cenilka** parametra.

Y je **nepristranska** cenilka, če $E(Y) = \zeta$, sicer **pristarnskost** 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 $\mathrm{SE}(Y)$ je standardna diviacija $\sigma(Y)$.

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

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 $(X_1, \dots, X_n)$ 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* cenilka za $z_k = E(X^k)$.

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

Metoda maksimalne zaneslijvosti (verjetja)

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

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

Pri danih $x_1, \dots, x_n$ izberemo ξ, da je dosežen maksimum funkcije L.

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

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

Intervalsko ocenjevanje parametrov

Naj bo gostota sl. sprem X odvisna od parametra ξ in naj bo $(x_1, \dots, x_n)$ vzorec.

Interval $[A, B]$ (*ki je odvisen le od vzorca*) je inteval zaupanja za parameter ξ pri **stopnji tveganja** $\alpha \in [0, 1]$, če je verjetnost

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

Številu $1 - \alpha$ rečamo **stopnja zaupanja** A in B pa sta vzorčni statistiki.

Waldov interval zaupanja

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

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

Waldov interval zaupanja za p:

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

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

- Zanima nas μ, σ je znan.

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

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

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

$$\frac{\overline{X} - \mu}{s}\sqrt{n} \sim \mathrm{Student}(n-1) \qquad \Delta = \frac{c\sigma}{\sqrt{n}}$$

$$c = t_{\alpha/2}(n-1) \qquad \overline{X} - \Delta < \mu < \overline{X} + \Delta$$

- Zanima nas σ, μ ni znan.

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

$$c_1 = \chi^2_{1-\alpha/2}(n-1) \qquad c_2 = \chi^2_{\alpha/2}(n-1)$$

$$s\sqrt{\frac{n-1}{c_2}} < \sigma < s\sqrt{\frac{n-1}{c_1}}$$

Preizkušanjue hipotez

Hipoteza je **enostavna**, če natančno določa porazdelitev, 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.

Test Z

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

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

Testna statistika, ki odloča o zavrnitvi hipoteze:

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

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:

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

Test T

$X \sim N(\mu, \sigma)$, 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 $T \sim Student(n-1)$.

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 $(X_1, \dots, X_m)$ in $(Y_1, \dots, Y_n)$.

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 $T \sim \mathrm{Student}(m+n-2)$.

Test χ^2 (Pearson)

Naj ima sl. sprem X porazdelitveno funkcijo 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: $S_1, \dots, S_r$, da je $p_k = P(X \in S_k | H_0) > 0$ za $\forall k = 1, \dots, r$.

Naj bo $(X_1, \dots, X_n)$ slučajni vzorec za sl. sprem X in N_k število vrednosti vzorca, ki padejo v S_k.

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

Pri velikem n ima statistka

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

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 parametra θ, 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 $x_1, \dots, x_n$ dobimo slučajni vektor $(Y_1, \dots, Y_n)$, kjer je $Y_k \sim N(a + bx_k, \sigma)$.

Radi bi ocenili vrednost a in b.

Metoda maksimalne zaneslijvosti

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

Vpeljemo vsote:

$$S_{xx} = \sum_{k=1}^n x_k^2 \quad S_{xY} = \sum_{k=1}^n x_k Y_k \quad S_x = \sum_{k=1}^n x_k \quad S_Y = \sum_{k=1}^n Y_k$$

Potem je

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

Testiranje neodvisnosti

Prilagoditveni test

To je poseben primer Pearsonevega testa χ^2

H_0: dogodka A in B sta neodvisna.

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

kategorije	$A \cap B$	$A \cap B^\complement$	$A^\complement \cap B$	$A^\complement \cap B^\complement$
verjetnost	pq	$p(1-q)$	$(1-p)q$	$(1-p)(1-q)$

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

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

ima porazdelitev $\chi^2(3)$

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

Če pa p in q nista znana, ju ocenimo iz podatkov:

$$\hat{p} = \frac{N_{A\cap B} + N_{A\cap B^\complement}}{n} \qquad \hat{q} = \frac{N_{A\cap B} + N_{A^\complement\cap B}}{n}$$

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 $(A_1, \dots, A_r)$, druga pa s kategorij $(B_1, \dots, B_s)$.

H_0: A_i in B_i sta neodvisna.

Naj bo $p_i = P(A_i)$ in $q_j(B_j)$.

Naj bo X_{ij} opažena frekvenca kategorije A_i in B_j.

Pričakovana frekvenca za X_{ij} je $n\hat{p}_i\hat{q}_j$.

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

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

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

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

Pri dani stopnji zaupanja α iz tabele razberemo c_α, da $P(\chi^2 > c_\alpha) = \alpha$. Kritično območje: $K_\alpha = [c_\alpha, \infty)$

Če je $n\hat{p}\hat{q} < 5$ za kak i, j moramo združiti nekatere razrede.

Neparametrični testi

Test z znaki

analog testa T

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

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

Testiramo hipotezo: $H_0(F_X = F_Y)$

Definirajmo razliko $D_i = X_i - Y_i$ za $i = 1, \dots, n$. Privzamemo, da so vzorčne vrednosti $D_i \neq 0$, sicer jih izpustimo in zmanjšamo n.

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

Naj bo S^+ število pozitivnih D_i, S^- pa število negativnih. Tedaj je $S^+ \sim \mathrm{Bin}(n, \frac{1}{2})$, torej je $p_k = P(S^+ = k) = \binom{n}{k}2^{-k}$.

Pri dani stopnji značilnosti α je kritično območje:

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

kjer je k_α doočena z zahtevama:

$$P(S^+ \leq k_\alpha) = \sum_{k=0}^{k_\alpha} p_k \leq \frac{\alpha}{2} \quad \text{in} \quad P(S^+ \leq k_\alpha + 1) = \sum_{k=0}^{k_\alpha+1} p_k > \frac{\alpha}{2}$$

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

Inverzijski test

Analog primerjalnega T testa

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

$$R_i = \mathrm{rang}(x_i) = k \iff x_i = z_k$$

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 $m+n \geq 20$ in m, n, je V približno normalno porazdeljena:

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

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

Inverzija med X_i in Y_j se pojavi, če ima Y_j manjši rang kot X_i.

Naj bo U št. inverzij v zaporedju $z_1, \dots, z_{n+m}$. Potem je $V = U + \frac{m(m+1)}{2}$, ker vsaka inverzija poveča vsoto rangov za 1.