Verjetnost - uvod

statistična definicija: eksperiment → dogodek A

n-ponovitev

$f_n(A)$ ... frekvenca dogodka A (število ponovitev, v katerih se je A zgodil)

verjetnost dogodka A

$$P(A) = \lim_{n\to\infty} \frac{f_n(A)}{n}$$

zgled: A... pade 6 $\qquad P(A) = ?$

klasična definicija:

dogodek A

izzidi ... elementarni dogodki, ki so enako možni
(privzamemo, da jih je končno mnogo)

zgled: homogena kocka → izzidi: pade 1, 2, 3, 4, 5, 6

A pade 6 $\qquad P(A) = \dfrac{\text{število izzidov, kjer se } A \text{ zgodi}}{\text{št. vseh izzidov}} = \dfrac{1}{6}$

B pade sodo število $\qquad P(B) = \dfrac{3}{6} = \dfrac{1}{2}$

zgled: dve kocki

Kolikšna je verjetnost, da je vsota 7?

2, 3, 4, ..., 7, ..., 11, 12 $\qquad$ niso vse enako možne

A… vsota 7 $\quad \cancel{P(A) = \frac{1}{11}}$

2. kocka \ 1. kocka	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$$P(A) = \frac{6}{36} = \frac{1}{6}$$

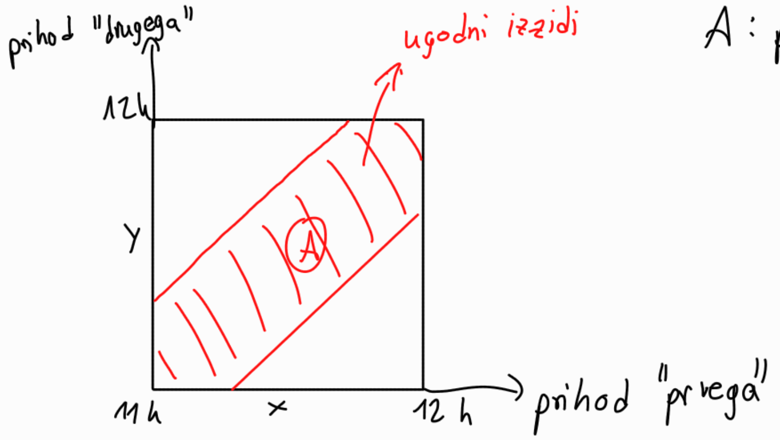
št. vseh izzidov

geometrijska definicija:

Zgled: 2 prijatelja se dogovorita, da se bosta dobila med 11h in 12h.

1. bo čakal 2. 20 min in nato odšel. V vsakem primeru bo odšel ob 12h.

Recimo, da pride vsak slučajno enkrat med 11h in 12h. Kolikšna je verjetnost, da se srečata?



A: prijatelja se srečata

$$|y-x| \leq \frac{1}{3} \quad (20 \text{ min})$$

$$x, y \in [0,1]$$

prihod prvega ob $11+x$

prihod drugega ob $11+y$

$$P(A) = \frac{\text{ploščina ugodnih izzidov}}{\text{ploščina vseh izzidov}} = \frac{1\cdot 1 - \frac{\frac{2}{3}\cdot\frac{2}{3}}{2}\cdot 2}{1\cdot 1} = \frac{1-\frac{4}{9}}{1} = \frac{5}{9}$$

primer: Problem Buffonove igle

Zgled iz fizike: Imamo m posod (stanja) in n kroglic (delci)

$m > n$

Kolikšna je verjetnost, da se pri slučajni porazdelitvi kroglic v posodi zgodi, da je v prvih m posodah v vsaki natanko 1 kroglica?

(a) kroglice so med seboj različne — Maxwell-Boltzmannova statistika za pline

vseh možnosti $= m^n$

ugodnih $= n!$

$$\boxed{P(A) = \frac{n!}{m^n}}$$

(b) kroglice se ne razlikujejo — Boltzmann-Einsteinova statistika za bozone

vseh možnosti $= \binom{n+m-1}{n} = \binom{n+m-1}{m-1}$

ugodnih $= 1$

$$P(A) = \frac{1}{\binom{n+m-1}{n}} = \frac{n! \cdot (m-1)!}{(n+m-1)!}$$

Diracova izključitveno načelo

(c) kroglice enake, v vsaki posodi največ 1 — Fermi-Diracova statistika za fermione

vseh možnosti $= \binom{m}{n}$

ugodnih $= 1$

$$P(A) = \frac{1}{\binom{m}{n}} = \frac{n! \cdot (m-n)!}{m!}$$

Aksiomatska definicija: (Kolmogorov ~1930)

Prostor izzidov Ω (množica)

dogodki so podmnožice v Ω (ne nujno vse)

$\mathcal{F}$ - množica dogodkov ima strukturo σ-algebre :

1) $\mathcal{F} \neq \emptyset$ vsebuje vsaj kakšen dogodek

2) je zaprta za komplemente

$A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

$A^c = \Omega \setminus A$

3) $\mathcal{F}$ je zaprta za štetne (neskončne) unije

$\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

za končne sledi iz lastnosti

ostale lastnosti σ-algebre

posledice 3 lastnosti :

4) $A, B \in \mathcal{F}$
$B \subseteq A$

$A \setminus B$ ... komplement B-ja v A-ju

$$A / B = A \cap B^c \overset{2, 9}{\in} \mathcal{F}$$

5) $\mathcal{F}$ je zaprta za končne unije :

$\{A_i\}_{i=1}^{\infty} \in \mathcal{F}$ če je $\emptyset \in \mathcal{F}$, potem lahko $A_i = \emptyset$ za $i > n$

ali: $A_i = A_1$

$$\Rightarrow \{A_i\}_{i=1}^{n} = \{A_i\}_{i=1}^{\infty} \in \mathcal{F}$$

6) $\emptyset \in \mathcal{F}$
$(\Omega \in \mathcal{F})$

1. $A \in \mathcal{F}$ (ker je $\mathcal{F}$ neprazna)

2. $\Rightarrow A^c \in \mathcal{F}$

$\overset{5}{\Rightarrow} A \cup A^c = \Omega \in \mathcal{F}$

$\Rightarrow \Omega^c = \emptyset \in \mathcal{F}$

7) najmanjša σ-algebra na Ω je $\{\emptyset, \Omega\}$

8) največja σ-algebra na Ω je $\mathcal{P}(\Omega)$

9) $\mathcal{F}$ je zaprta za števne preseke

$$\left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c \overset{2.,3.}{\in} \mathcal{F}$$

de Morganovo pravilo

operacije na dogodkih (to so elementi $\mathcal{F}$)

1) vsota (oziroma unija) $\quad A+B = A \cup B$

2) produkt (oziroma presek) $\quad AB = A \cap B$

lastnosti: - produkt in vsota sta komutativna, asociativna, idempotentna

- veljata distributivnosti

$$A(B+C) = AB + AC$$

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

- de Morganova zakona

$$(A \cap B)^c = A^c \cup B^c$$

$$(A \cup B)^c = A^c \cap B^c$$

tudi za neskončne unije in preseke

3) razlika $\quad A - B = A \setminus B = A \cap B^c$

če je $B \subseteq A$, potem je $A = B \cup (A \setminus B)$

$$B \cap (A \setminus B) = \emptyset$$

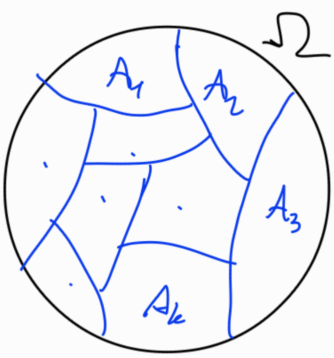
dogodka A in B sta nezdružljiva, če je $A \cap B = \emptyset$

zgled: Najmanjša σ-algebra je $\{\emptyset, \Omega\}$

Če σ-algebra $\mathcal{F}$ vsebuje A, ki ni $\emptyset$ in ni Ω,
potem mora $A^c \in \mathcal{F}$

$\{\emptyset, A, A^c, \Omega\}$ je najmanjša σ-algebra, ki vsebuje A
(σ-algebra generirana z A)

Če je $\mathcal{F}$ zaprta samo za končne unije in komplemente, potem rečemo, da je $\mathcal{F}$ algebra dogodkov (brez σ)



A_i je tako imenovan atom za $\mathcal{F}$, če $A_i \neq \emptyset$ in nobena prava podmnožica A_i ni v $\mathcal{F}$

$\mathcal{F}$ = vse možne unije A_i

def: Če so $\{A_i\}_{i \in I} \subseteq \mathcal{P}(\Omega)$, potem najmanjšo σ-algebro, ki vsebuje $A_i \; \forall i \in I$ imenujemo σ-algebra generirana z $\{A_i\}_{i \in I}$

zgled: $\Omega = \mathbb{N}$

$\{\{n\}, n \in \mathbb{N}\}$ najmanjša σ-algebra, ki vsebuje $\{\{n\}, n \in \mathbb{N}\}$ je $\mathcal{P}(\mathbb{N})$

najmanjša algebra, ki vsebuje $\{\{n\}, n \in \mathbb{N}\}$ pa je množica vseh končnih množic in množic, ki imajo končne komplemente

ta recimo ne vsebuje množice vseh sodih števil, torej $\neq \mathcal{P}(\mathbb{N})$

zgled: $\mathbb{R}$, (a,b) / $[a,b]$ interval (kakršenkoli)

najmanjšo σ-algebro generirano z vsemi odprtimi intervali v $\mathbb{R}$ imenujemo Borelova σ-algebra. Borelova σ-algebra na (a,b) ali na $[a,b]$ je σ-algebra generirana z vsemi odprtimi podintervali

opomba: enako dobimo, če vzamemo vse ~~zaprte~~ intervale

$\mathcal{B}$ oznaka za Borelovo σ-algebro

$\mathcal{B} \subsetneq \mathcal{P}(\mathbb{R})$

def: Če za končno (oz. števno) družino podmnožic A_i velja

kot atomi
$$\begin{cases} A_i \cap A_j = \emptyset \\ A_i \neq \emptyset \\ \bigcup_{i=1}^{n\ (\infty)} = \Omega \end{cases},$$

potem rečemo $\{A_i\}_{i=1}^{n\ (\infty)}$ imenujemo popoln sistem dogodkov

def:

Ω, $\mathcal{F}$ in verjetnost

$P : \mathcal{F} \to [0,1]$ vsakemu dogodku priredi število med 0 in 1

$(\Omega, \mathcal{F}, P)$ verjetnostni prostor je preslikava, za katero velja:

1) $P(A) \geq 0$
2) $P(\Omega) = 1$
3) za poljubno števno družino $\{A_i\}_{i=1}^{\infty}$ paroma nezdružljivih dogodkov velja, da $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

Rečemo, da je P σ-aditivna

Lastnosti verjetnosti:

1. $\boxed{P(\emptyset) = 0}$

dokaz: vzamemo družino $\{\Omega, \emptyset, \emptyset, \dots\}$

$$P(\Omega) = P(\Omega) + \infty \cdot P(\emptyset)$$
$$1 = 1 + \infty \cdot x \quad \Rightarrow x = 0$$

2. $A \in \mathcal{F}$ $\qquad \boxed{P(A^c) = 1 - P(A)}$

dokaz: $A \cup A^c = \Omega$
$A \cap A^c = \emptyset$

$\{A, A^c, \emptyset, \emptyset, \dots\} \to$ unija je Ω

$$P(\Omega) = P(A) + P(A^c) + \infty \cdot P(\emptyset)$$
$$1 = P(A) + P(A^c) + 0$$
$$P(A^c) = 1 - P(A)$$

3. P je aditivna:

$\{A_i\}_{i=1}^n$, $A_i \cap A_j = \emptyset$

$$\underset{P(\emptyset)=0}{\Rightarrow} P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

4. P je monotona

$$A \subseteq B \Rightarrow P(A) \leq P(B)$$

dokaz: $B = A \cup (B \setminus A)$

$$A \cap (B \setminus A) = \emptyset$$

$$P(B) \overset{\text{aditivnost}}{=} P(A) + \underbrace{P(B \setminus A)}_{\geq 0} \Rightarrow P(A) \leq P(B) \quad \square$$

5. P je zvezna: a) naraščajoče

Iz $A_1 \subseteq A_2 \subseteq \dots$ sledi $\boxed{\lim_{n\to\infty} P(A_n) = P\left(\bigcup_{i=1}^{\infty} A_i\right)}$

$\{A_i\}_{i=1}^{\infty}$ $\quad A_i \subseteq A_{i+1} \quad \forall i$

$$A_n = \bigcup_{i=1}^{n} A_n$$

b) padajoče

$A_1 \supseteq A_2 \supseteq \dots$

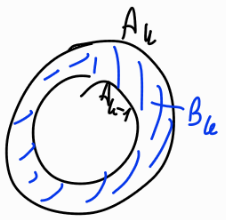
$\{A_i\}_{i=1}^{\infty}$ $\quad A_i \supseteq A_{i+1} \quad \forall i$ sledi $\lim_{n\to\infty} P(A_n) = P\left(\bigcap_{i=1}^{\infty} A_i\right)$

dokaz: Naj bo $B_1 = A_1$,

$B_2 = A_2 \setminus A_1$

$B_3 = A_3 \setminus A_2$

$\vdots$



$B_i \cap B_j = \emptyset$

$$\Rightarrow P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} B_i = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i)$$

$\|$ po konstr.

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

$$= \lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} B_i\right) = \lim_{n\to\infty} P(A_n)$$

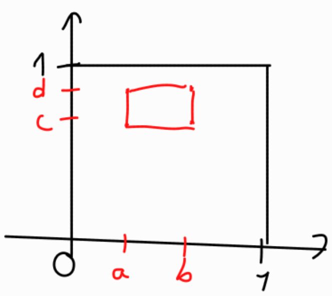
zgled: $\Omega = \{1, 2, \dots, n\}$

$\mathcal{F} = \mathcal{P}(\Omega)$

$P(\{i\}) = \frac{1}{n}$

$\Rightarrow P(A) = \frac{|A|}{n} = \frac{|A|}{|\Omega|}$ nam da klasično definicijo

zgled: $\Omega = [0,1] \times [0,1]$



naj bo $\mathcal{B}$ najmanjša σ-algebra, ki vsebuje vse pravokotnike $[a,b] \times [c,d]$

$a, b, c, d \in [0,1]$

$\mathcal{B}$ imenujemo Borelova σ-algebra

$P : \mathcal{B} \to [0,1)$ je določena z vrednostmi na pravokotnikih $[a,b] \times [c,d]$:

$$P([a,b]\times[c,d]) = (b-a)(d-c)$$ ploščina

to nam da <u>geometrijsko</u> definicijo verjetnosti

pogojna verjetnost

$(\Omega, \mathcal{F}, P)$ (oz. verjetnostni prostor na Ω):

$A \in \mathcal{F}$, $P(A) \neq 0$

$A \neq \Omega$

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

<u>pogojna verjetnost dogodka B pri pogoju A</u>

Vidimo: $P(A/A) = 1$

<u>zgled:</u> V posodi imamo 2 beli in 1 črno kroglico. Iz posode zaporedoma potegnemo po 1 kroglico. Kolikšna je verjetnost, da je 2. potegnjena kroglica bela, če je bila 1. bela?

$B_1 \ldots$ 1. kroglica bela

$B_2 \ldots$ 2. kroglica bela

(a) Vračamo $\quad P(B_2/B_1) = \dfrac{P(B_1\cap B_2)}{P(B_1)} = \dfrac{\frac{4}{9}}{\frac{2}{3}} = \dfrac{2}{3}$

$B_1 \cap B_2$:

1. \ 2.	B'	B''	$\bar{C}$
B'	✓	✓	×
B''	✓	✓	×
$\bar{C}$	×	×	×

$P(B_1\cap B_2) = \frac{4}{9}$

b) ne vračamo $\quad P(B_2/B_1) = \dfrac{P(B_1\cap B_2)}{P(B_1)} = \dfrac{\frac{2}{3}\cdot\frac{1}{2}}{\frac{2}{3}} = \dfrac{1}{2}$

def: Dogodka A in B sta neodvisna, če je $P(A\cap B) = P(A)\cdot P(B)$

zgled: B_1 in B_2 v primeru vračanja kroglic sta neodvisna

---

3. Predavanje

$A, B, C,\quad P(A) > 0,\quad P(A\cap B) > 0$

$$P(B/A) = \frac{P(A\cap B)}{P(A)} \Rightarrow P(A\cap B) = P(B/A)\cdot P(A)$$

$$P(\underbrace{A\cap B}\cap C) = P(C/_{A\cap B})\cdot P(A\cap B)$$

$$\boxed{P(A\cap B\cap C) = P(C/_{A\cap B})\cdot P(B/_A)\cdot P(A)}$$

splošno: $A_1, \dots, A_n,\quad P(A_1\cap\dots\cap A_n) > 0$

$$P(A_1\cap\dots\cap A_n) = P(A_n/_{A_1\cap\dots A_{n-1}})\cdot P(A_{n-1}/_{A_1\cap\dots\cap A_{n-2}})\cdot\ldots\cdot P(A_2/_{A_1})\cdot P(A_1)$$

Denimo, da $H_1, H_2, H_3, \dots, H_k \dots$ (končen ali števno neskončen popoln sistem dogodkov)

$P(H_i) > 0 \qquad H_i\cap H_j = \emptyset$

$\qquad\qquad\quad \bigcup_{i=1}^{\infty} H_i = \Omega$

$$\Rightarrow P(\Omega) = 1 = \sum_{i=1}^{\infty} P(H_i)$$

$A \in \mathcal{F} \quad \Rightarrow A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{\infty} H_i\right) = \bigcup_{i=1}^{\infty} (A \cap H_i)$

$\Rightarrow P(A) = \sum_{i=1}^{\infty} P(A \cap H_i) = \sum_{i=1}^{\infty} P(A/H_i) \cdot P(A)$

To je <u>formula o popolni verjetnosti</u>:

za popoln sistem dogodkov $\{H_i\}_i$ in $A \in \mathcal{F}$ velja

$$\boxed{P(A) = \sum_{i=1}^{\infty} P(A/H_i) \cdot P(H_i)}$$

Bayesova formula

zamenjamo

$$P(H_k/A) = \frac{P(H_k \cap A)}{P(A)} \quad \Rightarrow \quad \boxed{P(H_k/A) = \frac{P(A/H_k) \cdot P(H_k)}{\sum_{i=1}^{\infty} P(A/H_i) \cdot P(H_i)}}$$

<u>Zgled:</u> (a) Imamo 3 stroje, ki serijsko proizvajajo isti izdelek.

1. stroj	proizvede 25%	celotne proizvodnje, verjetnost, da je defekten	1%
2. stroj	35%		2%
3. stroj	40%		3%

Kolikšna je verjetnost, da je slučajno izbran izdelek iz celotne proizvodnje defekten?

D … defekten

A … na 1. stroju $\quad P(A) = \frac{25}{100}$

B … na 2. stroju $\quad P(B) = \frac{35}{100}$

C … na 3. stroju $\quad P(C) = \frac{40}{10}$

A, B, C tvorijo popoln sistem dogodkov

$$P(D) = P(D/A) \cdot P(A) + P(D/B) \cdot P(B) + P(D/C) \cdot P(C) = 0{,}01 \cdot 0{,}25 + 0{,}02 \cdot 0{,}35 + 0{,}03 \cdot 0{,}4$$

$$P(D) = \frac{215}{10000} = 0{,}0215$$

(b) izberemo slučajno defekten izdelek. Kolikšna je verjetnost, da je bil narejen na 3. stroju

$$P(C/D) = \frac{P(D/C)\cdot P(C)}{P(D)} = \frac{0{,}03\cdot 0{,}4}{0{,}0215} = \frac{3\cdot 40\cdot 10000}{100\cdot 100\cdot 215} = \frac{120}{215} = \frac{24}{43}$$

zgled: Test z detektorjem laži, ki je 95% pravilen, če testiranec govori resnico
95% pravilen, če testiranec laže

Recimo 1000 oseb, od teh je 1 lažnivec

Detektor laži naključno izbrano osebo proglasi za lažnivca. Kolikšna je verjetnost, da je res lažnivec?

popoln sistem dogodkov $\begin{cases} L \ldots \text{oseba lažnivec} \\ R \ldots \text{oseba resnicoljub} \end{cases}$

L_d ... proglasi za lažnivca (A)

R_d ... proglasi za resnicoljuba

$$P(L/L_d) = \frac{P(L_d/L)\cdot P(L)}{P(L_d/L)\cdot P(L) + P(L_d/R)\cdot P(R)}$$

$$= \frac{0{,}95\cdot \frac{1}{1000}}{0{,}95\cdot\frac{1}{1000} + 0{,}05\cdot\frac{999}{1000}} = 0{,}02 = 2\%$$

trditev: Če sta A in B neodvisna $P(A|B) = P(A)$ in
$P(B/A) = P(B)$

dokaz: $$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{P(A)\cdot \cancel{P(B)}}{\cancel{P(B)}} = P(A)$$

def: Dogodki $\{A_i\}_{i=1}^{n}$ so popolnoma neodvisni, če je

$$P(A_{i_1}, \ldots, A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j}) \quad \text{za vse } \{i_1,\ldots,i_k\} \subseteq \{1,\ldots,n\}$$

Dogodki so paroma neodvisni, če je poljuben par A_i, A_j neodvisen

Če so paroma neodvisni, potem ni nujno, da so popolnoma neodvisni.

zgled: Met tetraedra (naključno izberemo eno število v $\{1,2,3,4\}$)

$A=\{1,2\}$ $B=\{1,3\}$ $C=\{1,4\}$

želimo pokazati, da so A,B,C paroma neodvisni, niso pa popolnoma neodvisni

$$\left.\begin{array}{l} P(A)=P(B)=P(C)=\frac{2}{4}=\frac{1}{2} \\ P(A\cap B)=P(A\cap C)=P(B\cap C)=\frac{1}{4}=P(A)\cdot P(B)=P(A)\cdot P(C)=P(B)\cdot P(C) \end{array}\right\} \text{so paroma neodvisni}$$

$$\left.\begin{array}{l} P(A\cap B\cap C)=\frac{1}{4} \\ P(A)\cdot P(B)\cdot P(C)=\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{8} \end{array}\right\} \text{niso popolnoma neodvisni}$$

ZAPOREDJE NEODVISNIH POSKUSOV

poskus, opazujemo dogodek A pri ponavljajočem se poskusu. Ponovitve poskusa so med seboj popolnoma neodvisne.

zgled: met kocke

$A\ldots$ $P(A)=p\in(0,1)$ $p+q=1$

$P(A^c)=1-p=:q$

zanima nas $\overbrace{\text{verjetnost, da se pri } n \text{ ponovitvah poskusa } A \text{ zgodi } k\text{-krat}}^{A_n(k)}$ $0\leq k\leq n$

$\{A_n(k)\}_{k=0}^{n}$ je popoln sistem dogodkov

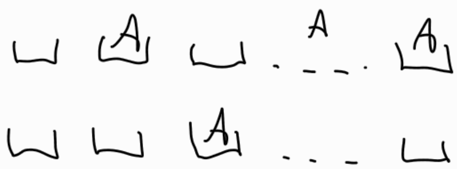
$$P(A_n(k)) \overset{\text{oznaka}}{=:} P_n(k) \overset{\text{formula o pop. ver.}}{=} \binom{n}{k} p^k \cdot q^{n-k} + \underbrace{0}_{\text{kjer se ni zgodil } k\text{-krat}}$$

$\binom{n}{k}$ – kje se je zgodil

p^k – verjetnost, da se je tam zgodil

q^{n-k} – verjetnost, da se drugje ni zgodil

n ponovitev



Bernoullijeva formula

$$\boxed{P_n(k) = \binom{n}{k} \cdot p^k q^{n-k}}$$

zgled: Kaljivost semena je 95%. Kolikšna je verjetnost, da izmed 1000 posejanih semen vzklije natanko 950?

$$P_{1000}(950) = \binom{1000}{950} \cdot 0{,}95^{950} \cdot 0{,}05^{50} = 0{,}05779$$

Za oceno $P_n(k)$ imamo "aproksimacijske" formule, ki so bodisi asimptotične bodisi limitne

1. Poissonova formula

smiselna je za velik n in majhen p $(n \cdot p \approx 0{,}0\ldots)$

$$\boxed{P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad , \text{ kjer je } \lambda = n \cdot p}$$

dokaz: Stirlingova formula: $n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

$$\lambda = n \cdot p \Rightarrow p = \frac{\lambda}{n}$$

potem je $P_n(k) = \dfrac{n(n-1)\cdot\ldots\cdot(n-k+1)}{k!} \cdot p^k \cdot q^{n-k}$

$$= \frac{n(n-1)\cdot\ldots\cdot(n-k+1)}{k!}\cdot\left(\frac{\lambda}{n}\right)^k\cdot\underbrace{\left(1-\frac{\lambda}{n}\right)^{n-k}}$$

$$\lim_{n\to\infty}\Big(= \frac{n(n-1)\cdot\ldots\cdot(n-k+1)}{k!}\cdot\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{e^{-\lambda}}\cdot\frac{\lambda^k}{n^k}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{1}$$

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

zgled: kaljivost semen $\quad P_{1000}(50) \approx \frac{\lambda^{50}}{50!}e^{-\lambda} \approx \frac{\cancel{50^{50}}\cdot \cancel{e^{-50}}}{\sqrt{100\pi}\,\cancel{\frac{50^{50}}{e^{50}}}} = \frac{1}{10\sqrt{\pi}}$

(50: seme ni kalilo)

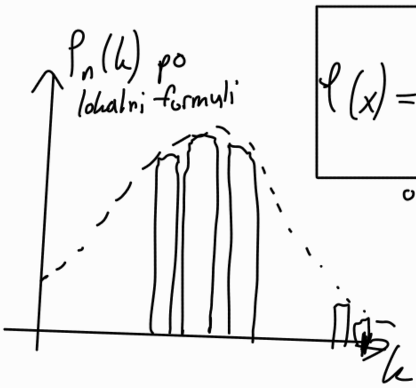
$\lambda = np = 1000\cdot 0{,}05 = 50$

$50! \approx \sqrt{100\pi}\cdot\left(\frac{50}{e}\right)^{50}$

2. Laplaceova lokalna formula

uporabna za velik n, p blizu 0,5

$$P_n(k)\approx \frac{1}{\sqrt{2\pi pqn}}\cdot e^{-\frac{(k-np)^2}{2pqn}}$$



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

osnova za Gaussovo krivuljo

premaknjena za np pomnožena z $\sqrt{npq}$

zgled: kaljivost semen

$$P_{1000}(50)\approx\frac{1}{\sqrt{2\pi\cdot0{,}05\cdot0{,}95\cdot1000}}\cdot e^{-\frac{(50-1000\cdot0{,}05)^2}{2\cdot0{,}05\cdot0{,}95\cdot1000}}$$

$p=0{,}05$

3. Laplaceova integralska formula

Opazujemo dogodek $B_n(k,l)$, da se v n ponovitvah dogodek A zgodi vsaj k-krat in največ l-krat $(k<l)$

$$B_n(k,l) = \bigcup_{j=k}^{l} A_n(j)$$

$$P_n(k,l) := P(B_n(k,l))$$

$$P_n(k,l) = \sum_{h=k}^{l} P_n(h) =$$

Lap. lok. for.

$$\approx \frac{1}{\sqrt{2\pi pqn}} \cdot \sum_{h=k}^{l} e^{-\frac{1}{2}\frac{(h-np)^2}{npq}} =$$

Δx_h $\quad x_h^2$

$$= \frac{1}{\sqrt{2\pi}} \cdot \sum_{h=k}^{l} e^{-\frac{1}{2}x_h^2} \cdot \Delta x_h$$

$$x_h := \frac{(h-np)^2}{\sqrt{npq}}$$

$$\Delta x_h := x_h - x_{h-1} = \frac{1}{\sqrt{npq}}$$

$\downarrow$ če gre Δx_h proti 0

$$\boxed{P_n(k,l) \approx \frac{1}{\sqrt{2\pi}} \cdot \int_a^b e^{-\frac{1}{2}x^2} dx \qquad a = x_k,\ b = x_l}$$

zgled: kaljivost semen $B_{1000}(950, 1000)$ $\downarrow$ vsaj 950 skali

$$x_l = \frac{(50 - 1000 \cdot 0{,}05)^2}{1000 \cdot 0{,}95 \cdot 0{,}05}$$

$$P^{A}_{1000}(950,1000) = P^{\bar{A}}_{1000}(0,50) \approx \frac{1}{\sqrt{2\pi}} \int_{-7{,}36}^{0} e^{-\frac{1}{2}x^2} dx \approx 0{,}5$$

Slučajne spremenljivke

$(\Omega, \mathcal{F}, P)$

$X: \Omega \to \mathbb{R}$ slučajna spremenljivka je realna funkcija na Ω,

ki zadošča:

$$X^{-1}((-\infty, x]) \in \mathcal{F} \text{ za } \forall x \in \mathbb{R}$$

(v teoriji mere je $X: \Omega \to \mathbb{R}$ merljiva) zgled: X predstavlja število pik pri 1 metu kocke

$\Omega = \{\text{pade } i \text{ pik}, \ i=1,\dots,6\}$

$\mathcal{F} = \mathcal{P}(\Omega)$

$X(\text{pade } i \text{ pik}) = i$

$X : \Omega \to \mathbb{R}$ diskretna $|X(\omega_1) - X(\omega_2)| > \delta \quad \forall \omega_1, \omega_2$

potem je X slučajna spremenljivka

def: $X : \Omega \to \mathbb{R}$ slučajna spremenljivka

funkcija $F_X(x) = P[X \leq x] = P(X^{-1}(-\infty, x])$

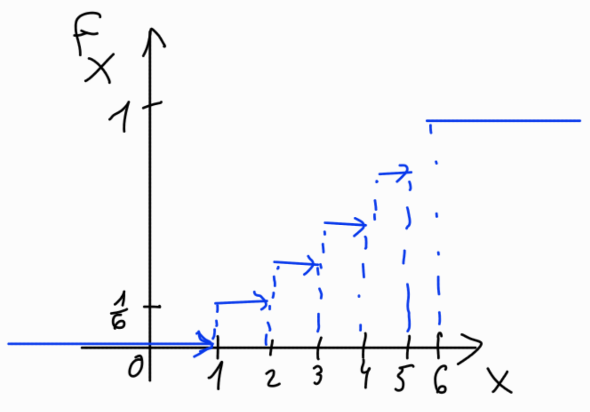
standardna oznaka

porazdelitvena funkcija za X

$F_X : \mathbb{R} \to [0,1]$

zgled: met kocke $\Omega = \{\text{pade } i \text{ pik}, \ i=1,\dots,6\}$

$X(\text{pade } i \text{ pik}) = i \qquad P(\text{pade } i \text{ pik}) = \frac{1}{6} \quad \forall i$



$X: \Omega \to \mathbb{R}$ slučajna spremenljivka

F_X njena porazdelitvena funkcija

lastnosti F_X:

$x_1 \leq x_2$

1) $0 \leq F_X(x) \leq 1$

$F_X(x_1) = P[X \leq x_1]$ $F_X(x_2) = P[X_2 \leq x_1]$

2) F_X naraščajoča $\xrightarrow{\text{dokaz}}$ $[X \leq x_1] \subseteq [X_2 \leq x_1]$

$\overset{\text{monotonost } P}{\Rightarrow} P[X \leq x_1] \leq P[X_2 \leq x_1]$

3) $\lim\limits_{x \to -\infty} F_X(x) = 0$

$\lim\limits_{x \to -\infty} P\left(\bigcap [X \leq x]\right) = P(\emptyset) = 0$

4) $\lim\limits_{x \to \infty} F_X(x) = 1$

$\lim\limits_{x \to \infty} P\left(\bigcup [X \leq x]\right) = P(\Omega) = 1$

5) $F_X(x)$ je z desne zvezna v vsaki točki x

dokaz: $F_X(x+) = \lim\limits_{h \downarrow 0} F_X(x+h) = F_X(x)$

Naj bo $\{x_n\}_{n=1}^{\infty}$ padajoče zaporedje, ki konvergira k x

$$\bigcap_{n=1}^{\infty} [X \leq x_n] = [X \leq x]$$

$$\Rightarrow F_X(x+) = \lim_{n \to \infty} F_X(x_n) = \lim_{n \to \infty} P\left(\bigcap [X \leq x_n]\right) = P[X \leq x] = F_X(x) \quad \square$$

$F_X(x-)$ obstaja za vse x, lahko se zgodi, da je $F_X(x-) < F_X(x)$

$F_X: \mathbb{R} \to [0,1]$ naraščajoča, slika v kompaktno množico

$\Rightarrow$ ima največ števno mnogo točk nezveznosti

6) $F_X(x-) = P[X < x]$

$$[X < x] = \bigcup_{n=1}^{\infty} [X \leq x_n]$$

$\{x_n\}_{n=1}^{\infty}$ naraščajoče zaporedje, ki konvergira k x

$$\lim_{n\to\infty} F_X(x_n) = F_X(x-)$$

$$\lim_{n\to\infty} P[X \leq x_n] = P\left(\bigcup_{n=1}^{\infty} [X \leq x_n]\right) = P([-\infty, x)]$$

najpomembnejši družini slučajnih spremenljivk:

- diskretne slučajne spremenljivke
- zvezne slučajne spremenljivke

$X: \Omega \to \mathbb{R}$ je diskretno slučajno porazdeljena, če je njena zaloga vrednosti končna ali števno neskončna

take slučajne spremenljivke podamo z verjetnostno tabelo ali splošnim pravilom

$$X: \begin{pmatrix} x_1 & \dots & x_n \\ p_1 & \dots & p_n \end{pmatrix} \quad 0 < p_i < 1 \quad \sum_{i=1}^{n\,(\infty)} p_i = 1$$

↓

porazdelitev za X

zgledi: 1) Enakomerna porazdelitev ... $U(n)$

$$X : \begin{pmatrix} 1 & 2 & \dots & n \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}$$

$F_X(x)$ za diskretno porazdelitev je odsekoma konstnantna, v točki x_i ima skok višine p_i $\forall i$

2) Binomska porazdelitev ... $Bin(n,p)$

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

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \left(p + (1-p)\right)^n = 1^n = 1$$

3) Bernoullijeva porazdelitev ... $Ber(p)$ $0<p<1$

$$X : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

4) Poissonova porazdelitev ... $Poi(\lambda)$ $\lambda > 0$ λ imenujemo intenziteta

$$P[X=k] = p_k = \frac{\lambda^k}{k!} e^{-\lambda} \qquad k = 0, 1, \dots$$

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \cdot e^{\lambda} = 1$$

Taylor za e^{λ}

5) Geometrijska porazdelitev ... $Geo(p)$ $0<p<1$, $q = 1-p$

$$p_k = p \cdot q^k \qquad k = 0, 1, \dots$$

(met kocke do prve šestice)

lahko tudi $p_k = p \cdot q^{k-1}$ $k=0,1,\ldots$

$$\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} p \cdot q^k = p \cdot \frac{1}{1-q} = \frac{p}{1-(1-p)} = \frac{p}{p} = 1$$

6) Pascalova (ali negativna binomska) porazdelitev ... $Pas(m, p)$

(Met kocke dokler m-tič ne dobimo šestice) $q = 1-p$

$$p_k = \binom{k-1}{m-1} \cdot p^m \cdot q^{k-m} \quad k = m, m+1, \ldots$$

$\downarrow$

v k-tem metu m-tič pade

za $m=1$ dobimo geometrijsko porazdelitev

7) Hipergeometrijska porazdelitev ... $Hip(n, M, N)$

$$p_k = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \quad k = 0, \ldots, n$$

$n \leq M$
$n \leq N-M$
$N > M$

posoda z N kroglicami, od katerih je N belih, $N-M$ črnih.

Izberemo n kroglic, koliko je belih?

$X = k$ je belih

$$\sum_{k=0}^{n} p_k = \frac{1}{\binom{N}{n}} \sum_{k=0}^{n} \binom{M}{k}\binom{N-M}{n-k} = \frac{\binom{N}{n}}{\binom{N}{n}} = 1$$

Zvezne slučajne spremenljivke

$X : \Omega \to \mathbb{R}$; X je zv. slučajna spr., če $\exists$ taka integrabilna funkcija $p_X(x)$ na $\mathbb{R}$, da je

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

$p_X(x)$ imenujemo gostota slučajne spr. X

komentarji:

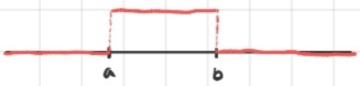
- F_X je zv. funk. $F_X : \mathbb{R} \to [0,1]$
- $p_X(x) \geq 0$ (F_X monotona)
- $\lim\limits_{x\to\infty} F_X(x) = 1 = \int_{-\infty}^{\infty} p_X(t)\, dt$

pomembne zv. porazdelitve slučajnih spr.

1) Enakomerna porazdelitev na intervalu $[a,b]$:
$a<b$

$$p_X(x) = \begin{cases} \frac{1}{b-a} \; ; & a<x<b \\ 0 \; ; & \text{sicer} \end{cases} \qquad F_X(x) = \begin{cases} 0 \; ; & x \leq a \\ 1 \; ; & x \geq b \\ \frac{x-b}{b-a} \; ; & a<x<b \end{cases}$$

$U(0,1)$, $U(a,b)$



2) Normalna oz. Gaussova porazdelitev $\mathcal{N}(\mu, \sigma)$ $\mu \in \mathbb{R}$ $\sigma > 0$

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

$\mathcal{N}(0,1)$: $\Phi(x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt$

standardna normalna porazdelitev

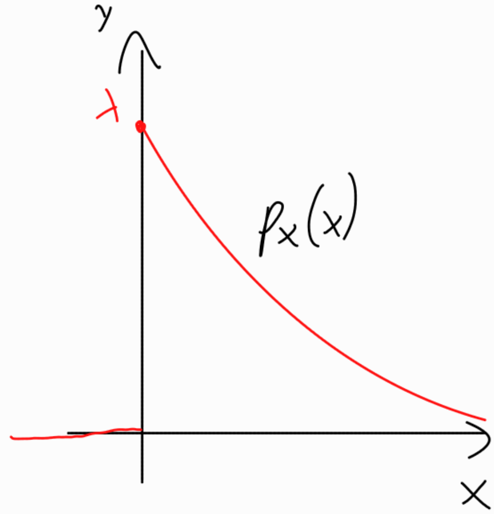
$$F_X(x) = \frac{1}{2} + \Phi\left(\frac{x-\mu}{\sigma}\right)$$

zaradi začetka integrala od 0 naprej

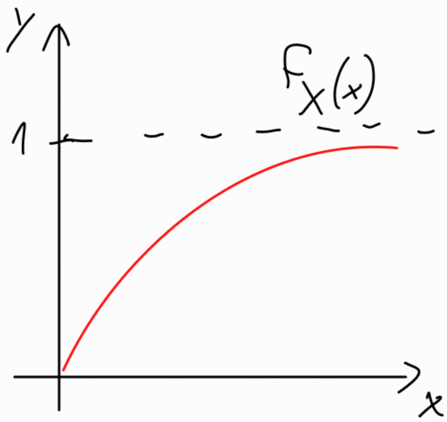
(3) Eksponentna porazdelitev

$Exp(\lambda) \quad \lambda > 0$

$$p_X(x) = \begin{cases} \lambda e^{-\lambda x} & ; x > 0 \\ 0 & ; \text{sicer} \end{cases}$$



$$F_X(x) = \begin{cases} 0 & ; x \leq 0 \\ 1 - e^{-\lambda x} & ; x > 0 \end{cases}$$



(4.) Gama porazdelitev

$\Gamma(b, c) \quad b, c > 0$

$$p_X(x) = \begin{cases} 0 & ; x \leq 0 \\ \frac{c^b}{\Gamma(b)} x^{b-1} e^{-cx} & ; x > 0 \end{cases}$$

$$Exp(\lambda) = \Gamma(1, \lambda)$$

(5.) Hi-kvadrat porazdelitev

$\chi^2(n) \quad n \in \mathbb{N}$

$$p_X(x) = \begin{cases} 0 & ; x \leq 0 \\ \frac{1}{2^{\frac{n}{2}} \cdot \Gamma(\frac{n}{2})} \cdot x^{\frac{n}{2} - 1} e^{-\frac{x}{2}} & ; x > 0 \end{cases}$$

$$\chi^2(n) = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$$

6. Cauchyjeva porazdelitev

$$p_X(x) = \frac{1}{\pi(1+x^2)} \qquad F_X(x) = \frac{1}{\pi} \cdot \arctan(x)$$

7. Paretova porazdelitev

$P(a, \alpha) \quad a>0 \quad \alpha>0$

$$p_X(x) = \begin{cases} 0 & ; x \leq a \\ \alpha \cdot \frac{a^\alpha}{x^{\alpha+1}} & ; x > a \end{cases}$$

8. Studentova porazdelitev $t_\nu \quad \nu > 0 \quad (\nu \in \mathbb{N})$

$$p_X(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu} \cdot \Gamma\left(\frac{\nu}{2}\right)} \cdot \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

t_1 = Cauchyjeva porazdelitev

$t_\infty = N(0,1)$

Slučajni vektorji

$X = (X_1, X_2, \dots, X_n) : \Omega \to \mathbb{R}^n \quad n \geq 2$

$X_i : \Omega \to \mathbb{R}$ slučajna spremenljivka

porazdelitvena funkcija za slučajni vektor X:

$$F_X(x_1, \dots, x_n) = P[X_1 \leq x_1, \dots, X_n \leq x_n]$$

$x = (x_1, \dots, x_n)$

$$0 \le F_X(x) \le 1$$

$$\lim_{\substack{x_i \to \infty \\ \text{za vse } i}} F_X(x) = 1$$

$\tilde{X}$ podvektor v X:

recimo $\tilde{X} = (X_1, X_{10}, X_{11})$

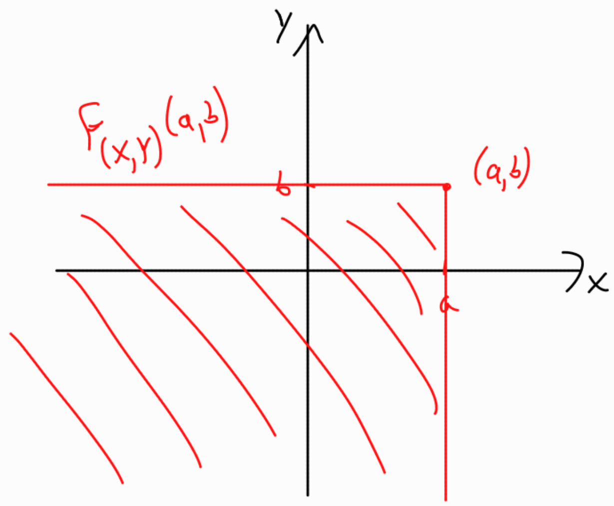
$$F_{\tilde{X}}(\tilde{x}) = \lim_{\substack{x_i \to \infty \\ x_i \notin \tilde{X}}} F_X(x)$$

$$X = (X_1, X_2, X_3)$$

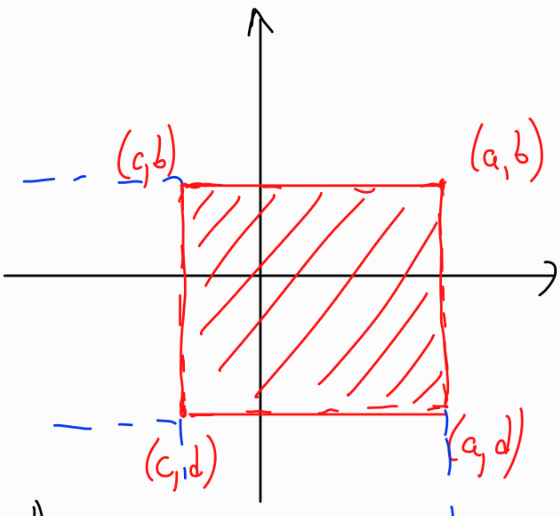
$$F_{(X_1,X_2)}(x_1,x_2) = \lim_{x_3 \to \infty} F_{(X_1,X_2,X_3)}(x_1,x_2,x_3)$$

$F_{X_i}(x_i)$ ... imenujemo robne (marginalne) porazdelitve slučajnega vektorja

$\underset{\sim}{X} = (X, Y)$ pomen $F_{(X,Y)}(a,b)$?



$$F_{(X,Y)}(a,b) = P[X \le a, Y \le b]$$



$$P[c \le X \le a,\ d \le Y \le b]$$

$$= F_{(X,Y)}(a,b) - F_{(X,Y)}(a,d) - F_{(X,Y)}(c,b) + F_{(X,Y)}(c,d) \ge 0$$

porazdelitve slučajnih vektorjev

1.) <u>Diskretne porazdelitve</u>: podamo s tabelo verjetnosti

$$p_{i_1, i_2, \dots, i_n} = P[X_1 = i_1, \dots, X_n = i_n]$$

$i_1 \dots i_n$ po vseh možnih vrednostih

$n = 2$:

$Y \backslash X$	$x_1 \quad \dots \quad x_s$	Y
y_1	$p_{11} \quad \dots \quad p_{1s}$	q_1
$\vdots$	$\vdots \qquad\qquad \vdots$	$\vdots$
y_r	$p_{r1} \quad \dots \quad p_{rs}$	q_r
X		

porazdelitev za Y

porazdelitev za X

$$p_{ij} = P[X = x_j, Y \leq y_i]$$

2.) <u>Zvezne porazdelitve</u>

X je porazdeljen zvezno, če obstaja integrabilna funkcija n spremenljivk

$p_X(x_1, \dots, x_n)$, da je $F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} p_X(t_1, \dots, t_n)\, dt_1 \dots dt_n$

to funkcijo imenujemo <u>gostota slučajnega vektorja X</u>.

a) $p_X(x) \geq 0$

b) $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p_X(t_1, \dots, t_n)\, dt_1 \dots dt_n = 1$

$F_X'(x) = p_X(x)$ za 1 spremenljivko

$$\frac{\partial^n}{\partial x_1 \dots \partial x_n} F_X = p_X$$

$$p_{\hat{X}}(\tilde{x}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p_X(\tilde{x}, \hat{t})\, d\hat{t}_1 \dots d\hat{t}_k \qquad t_1, \dots, t_k \notin \tilde{x}$$

npr. (X, Y)

$p_{X,Y}(x,y)$ gostota

$$p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)\, dy$$

$$p_Y(y) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)\, dx$$

Poseben primer zvezne porazdelitve je večrazsežna normalna porazdelitev

$\mathcal{N}(\mu, \Sigma)$ $\quad \mu \in \mathbb{R}^n$, Σ pozitivno definitna matrika

$$\boxed{p_X(x) = \sqrt{\frac{\det \Sigma}{(2\pi)^n}}\; e^{-\frac{1}{2}(x-\mu)^T \Sigma (x-\mu)}}$$

$\det \Sigma \neq 0$

$\langle \Sigma u, u \rangle > 0$ za $u \neq 0$, $u \in \mathbb{R}^n$

$\| $

$u^T \Sigma u$

za $n=2$:

$$K = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$$

$\sigma_x > 0$

$\sigma_y > 0$

$N(\mu_x, \sigma_x) \sim X$

$N(\mu_y, \sigma_y) \sim Y$

robni porazdelitvi

$\mu = (\mu_x, \mu_y)$

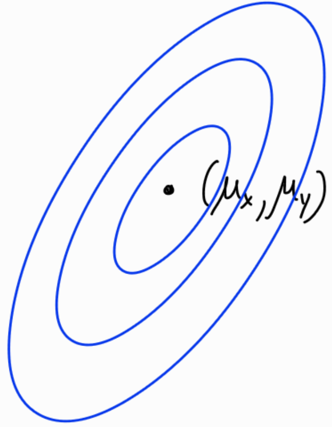
$-1 < \rho < 1$

$$\Sigma = K^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} \frac{1}{\sigma_x^2} & -\frac{\rho}{\sigma_x\sigma_y} \\ -\frac{\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{bmatrix}$$

$$\boxed{p_{X,Y}(x,y) = \frac{1}{2\pi \cdot \sigma_x\sigma_y\sqrt{1-\rho^2}}\; e^{-\frac{1}{2(1-\rho^2)} \cdot \left(\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 \right)}}$$

Nivojnice: $p_{X,Y}(x,y) = C$ so <u>elipse</u>

krožnice $\Leftrightarrow$ $\rho = 0$ in $\sigma_x = \sigma_y$



ρ vrti os

σ_x in σ_y ... polosi

Neodvisnost slučajnih spremenljivk

X ... slučajni vektor

$F_X(x)$ ... njegova porazdelitvena funkcija

potem so slučajne spremenljivke $X_1, \ldots, X_n$ <u>neodvisne</u>, če je

$$F_X(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_2}(x_2)$$

Izrek: Slučajne spremenljivke $X_1, \ldots, X_n$ so neodvisne, natanko tedaj ko so dogodki $[X_1 \in A_1], \ldots, [X_n \in A_n]$ popolnoma neodvisne za poljubne Borelove množice $A_1, \ldots, A_n$

$n=2$:

$X \backslash Y$	y_1	...	y_n	X
x_1	p_{11}	...	p_{1n}	p_1
$\vdots$	$\vdots$		$\vdots$	$\vdots$
x_m	p_{m1}	...	p_{mn}	p_m
Y	q_1	...	q_n	1

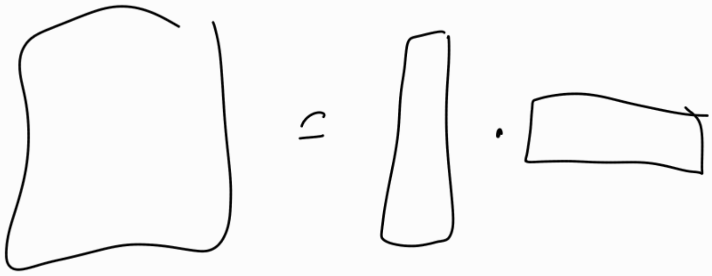
X, Y neodvisna $\Leftrightarrow$ $p_{ij} = p_i \cdot q_j$

uporabimo

$$F_{(X,Y)}(x,y) = F_X(x) \cdot F_Y(y)$$

$$\Updownarrow$$

$$P[X = x_i, Y = y_i] = P[X = x_i] \cdot P[Y = y_j]$$



se da napisati $\iff$ ima rang 1

$$\boxed{X, Y \text{ neodvisni} \iff P = [p_{ij}] \text{ ima rang } 1}$$

(X,Y) zvezen slučajen vektor:

potem sta X in Y neodvisni sluč. spremenljivki

$$\Updownarrow$$

$$p_{(X,Y)}(x,y) = f(x) \cdot g(y)$$

$$F_{(X,Y)}(x,y) = F_X(x) \cdot F_Y(y)$$

Zgled: $n=2$, $N(\mu, \Sigma)$

$$p_{(X,Y)}(x,y) = K \cdot e^{-\frac{1}{2}\left[(x)^2 \ldots \cancel{(x)(y)} \ldots (y)^2\right]}$$

X, Y neodvisni $\iff \rho = 0$

za $n \geq 2$

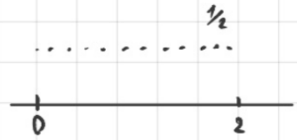
$X_1, \ldots, X_n$ neodvisne $\iff \Sigma$ diagonalna matrika

$$\Sigma = \begin{pmatrix} \sigma_{X_1} & & 0 \\ & \ddots & \\ 0 & & \sigma_{X_n} \end{pmatrix}$$

Zgled: slučajna spr. ki ni niti diskretna niti zvezna:

X dobitek v naslednjem poskusu:

- najprej vržemo kovanec. Če pade grb dobimo 1, sicer pa X izberemo enakomerno na $[0,2]$



$$Z \sim U(0,2) \sim p_Z(z) = \begin{cases} 1/2 & ; 0 < z < 2 \\ 0 & ; \text{sicer} \end{cases}$$

$$P[Z \leq z] = \begin{cases} 0 & z \leq 0 \\ \frac{z}{2} & 0 < z < 2 \\ 1 & z \geq 2 \end{cases}$$

$[X \leq x]$ $0 \leq x < 1$:

$$P[X \leq x] = P[X \leq x \mid grb]\, P[grb] + P[X \leq x \mid cifra]\, P[cifra] = 0 \cdot \frac{1}{2} + \frac{x}{2} \cdot \frac{1}{2} = \frac{x}{4}$$

$1 \leq x \leq 2$:

$$P[X \leq x] = P[X \leq x \mid grb]\, P[grb] + P[X \leq x \mid cifra]\, P[cifra] = 1 \cdot \frac{1}{2} + \frac{x}{2} \cdot \frac{1}{2} = \frac{1}{2} + \frac{x}{4}$$

$$F_X(x) = \begin{cases} 0 & ; \quad x \leq 0 \\ \frac{x}{4} & ; \quad 0 \leq x < 1 \\ \frac{1}{2} + \frac{x}{4} & ; \quad 1 \leq x \leq 2 \\ 1 & ; \quad x \geq 2 \end{cases}$$

Funkcije slučajnih spr.

$X: \Omega \to \mathbb{R}$

$f: \mathbb{R} \to \mathbb{R}$ zvezna

$Y = f(X) = f \circ X$ Ali je Y slučajna spr.?

$$[Y \le y] = [f(X) \le y] = \{\omega \mid f(X)(\omega) \le y\} = \{\omega \mid f(X)(\omega) \in (-\infty, y]\} = \{\omega \mid X(\omega) \in \underbrace{f^{-1}((-\infty, y])}_{\in \mathcal{B} \text{ (borelova)}}\}$$

$\Rightarrow [Y \le y] \in \mathcal{F}$

$\Rightarrow Y$ je slučajna spr.

- Kako poiščemo porazdelitev F_Y, če poznamo F_X in f.

• $F_Y(y) = P[Y \le y] = P[f(X) \le y]$ + f ima zvezen odvod

Če X zv. sl. spr., f strogo naraščajoča in ima zalogo vrednosti na (a,b) za $-\infty \le a < b \le \infty$

V tem primeru obstaja $f^{-1}: (a,b) \to \mathbb{R}$,

Potem velja: $F_Y(y) = P[Y \le y] = P[f(X) \le y] \overset{a<y<b}{=} P[X \le f^{-1}(y)] = F_X(f^{-1}(y))$ $\quad \left(F_Y(y) = 0 \;\; y < a;\; F_Y(y) = 1 \text{ za } y > b \right)$

f naraščajoča in obr.

• Tj. porazdelitvena funkcija F_Y enaka:

$F_Y(y) = F_X(f^{-1}(y))$ in

gostota: $p_Y(y) = p_X(f^{-1}(y)) \cdot (f^{-1}(y))'$

verižno pravilo

• Če f strogo padajoča zv. odv. funk. pa velja: $F_Y(y) = P[Y \le y] = P[f(X) \le y] = P[X \ge f^{-1}(y)] = 1 - P[X < f^{-1}(y)] =$

X zv. sluč. spr.

$= 1 - F_X(f^{-1}(y))$

$\Rightarrow p_Y(y) = -p_X(f^{-1}(y)) \cdot (f^{-1}(y))' \Rightarrow p_Y(y) = p_X(f^{-1}(y)) \, |(f^{-1}(y))'|$

$p_Y(y) = 0$ za $y \le a$ in $y \ge b$

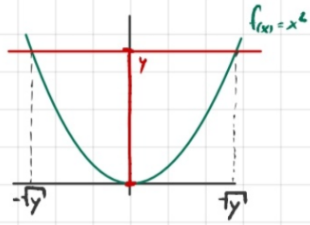
- Zgled: 1) $X \sim N(0,1)$, $Y = aX + b$ $\quad a \ne 0$

$f(x) = ax + b = y \Rightarrow x = \frac{y-b}{a} \Rightarrow f^{-1}(y) = \frac{y-b}{a}$ $\quad f$ bijekcija na $\mathbb{R}$ $\qquad p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$p_Y(y) = p_X\left(\frac{y-b}{a}\right) \cdot \left|\frac{1}{a}\right| = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-b)^2}{a^2}} \cdot \frac{1}{|a|} \Rightarrow Y \sim N(b, |a|)$

2) $Z \sim N(\mu, \sigma)$: $\mu \in \mathbb{R}$, $\sigma > 0$

$Y = \frac{Z-\mu}{\sigma}$ $\rightarrow$ $Y \sim N(0,1)$ standardna normalna porazdelitev



3) $X \sim N(0,1)$, $Y = X^2 \rightarrow f(x) = x^2$ (ni bijekcija), iščemo $p_Y(y) = ?$

- Najprej poiščemo porazdelitveno funkcijo $F_Y(y)$

$$F_Y(y) = P[Y \le y] = P[X^2 \le y] = \begin{cases} 0 & ; y \le 0 \\ \sqrt{\frac{2}{\pi}} \int_0^{\sqrt{y}} e^{-\frac{x^2}{2}} dx & y > 0 \end{cases}$$

Za $y > 0$: $F_Y(y) = P[x^2 \le y] = P[-\sqrt{y} \le X \le \sqrt{y}]$

$= P[X \in [-\sqrt{y}, \sqrt{y}]] =$

$= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{y}}^{\sqrt{y}} e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{y}} e^{-\frac{x^2}{2}} dx =$

- $$p_Y(y) = \begin{cases} \sqrt{\frac{2}{\pi}} \cdot e^{-\frac{y}{2}} \cdot \frac{1}{2} \frac{1}{\sqrt{y}} & ; y > 0 \\ 0 & ; y \le 0 \end{cases}$$

- za $y > 0$: $p_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \cdot y^{-\frac{1}{2}}$ $\Rightarrow$ opazimo $Y \sim \chi^2(1)$

4) $X \sim \begin{pmatrix} -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$

$X^2 \sim \begin{pmatrix} 0 & 1 \\ 1/3 & 2/3 \end{pmatrix}$

$[X^2 = 0] = [X = 0]$

$[X^2 = 1] = [X = 1] \cup [X = -1]$

Funkcije slučajnih vektorjev

- $X = (X_1, X_2, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$

$f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ zvezna (če X zv. naj bo f zv. odv.) $\quad (f = (f_1, \dots, f_m),\ f_i : \mathbb{R}^n \rightarrow \mathbb{R})$

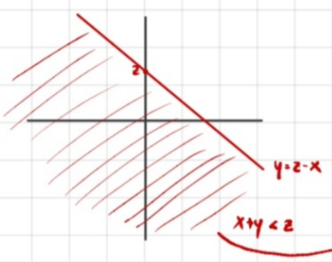
$Y = f(X) : \Omega \rightarrow \mathbb{R}^m$

$F_Y(y) = P[Y_1 \le y_1, Y_2 \le y_2, \dots, Y_m \le y_m] = P[f_1(X) \le y_1, \dots, f_m(X) \le y_m]$

- **m = 1** Dan je zv. slučajni vektor (X, Y) in $Z = X + Y$

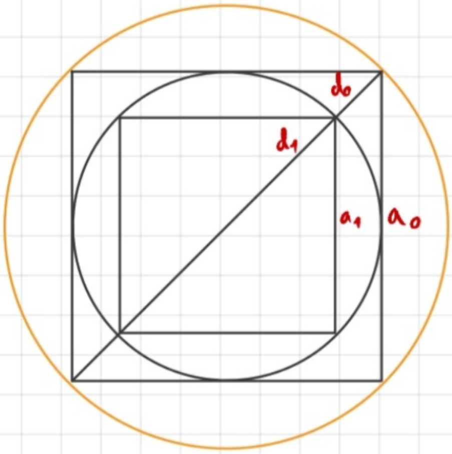
$f(x,y) = x + y$, $f: \mathbb{R}^2 \rightarrow \mathbb{R}$; iščemo gostoto za Z (oz. $F_Z(z)$)

$$F_Z(z) = P[Z \le z] = P[X + Y \le z] = \iint_{x+y \le z} p_{(X,Y)}(x,y)\, dx\, dy = \int_{-\infty}^{\infty} dx \int_{-\infty}^{z-x} p_{(X,Y)}(x,y)\, dy$$



$x + y = z$

$y = z - x$



$d_0 = 10$

$a_0^2 + a_0^2 = d_0^2 \rightsquigarrow a_0 = \frac{\sqrt{100}}{\sqrt{2}} = \frac{10}{\sqrt{2}} = 10\left(\frac{1}{\sqrt{2}}\right)^1$

$d_1 = a_0 = \frac{10}{\sqrt{2}}$

$a_1 = \frac{10}{\sqrt{2}\sqrt{2}} = 10\left(\frac{1}{\sqrt{2}}\right)^2$

$d_n = \frac{10}{2^{\frac{n}{2}}} = \frac{5 \cdot 2}{2^{\frac{n}{2}}} = 5 \cdot 2^{-\frac{n}{2}+1} = 10\left(\frac{1}{\sqrt{2}}\right)^n$

$a_n = d_{n+1} = 5 \cdot 2^{-\frac{n+1}{2}+1} = 10\left(\frac{1}{\sqrt{2}}\right)^{n+1}$

$S = \pi r^2 = \pi \frac{d^2}{4}$ $\qquad S_n = \pi \frac{d_n^2}{4} = \pi \cdot \frac{\left(10\left(\frac{1}{\sqrt{2}}\right)^n\right)^2}{4} =$

$= \pi \frac{100}{4} \left(\frac{1}{\sqrt{2}}\right)^{2n} = 25\pi \, 2^{-\frac{2n}{2}} = 25\pi \, 2^{-n}$

$S = \frac{a_1}{1-k} = \frac{25\pi}{1-\frac{1}{2}} = \frac{25\pi}{\frac{1}{2}} = 50\pi$

zaporedje za polmer: $r_n = \frac{d_n}{2} = 5 \cdot \left(\frac{1}{\sqrt{2}}\right)^n$

stranice kvadrata: $a_n = 10\left(\frac{1}{\sqrt{2}}\right)^{n+1}$

$\sqrt[3]{3\sqrt[3]{3\sqrt[3]{3\sqrt[3]{3}}}}\ldots = \sqrt[3]{3} \cdot \sqrt[3]{3} \cdot \sqrt[27]{3} \cdots = 3^{\frac{1}{3}} \cdot 3^{\frac{1}{3}} \cdot 3^{\frac{1}{27}}$

$a_n = a_{n-1} \cdot 3^{\frac{1}{3^n}}$

$a_0 = 1 \quad a_1 = \sqrt[3]{3} \quad a_2 = \sqrt[3]{3} \cdot \sqrt[3]{3}$

$\lim_{n\to\infty} a_n = \lim_{n\to\infty} a_{n-1} \cdot 3^{\frac{1}{3^n}} = \lim_{n\to\infty} a_{n-1} \cdot \lim_{n\to\infty} 3^{\frac{1}{3^n}} = \lim_{n\to\infty} a_{n-1} \cdot 1 = \lim_{n\to\infty} a_{n-1}$

$\lim_{n\to\infty} a_{n-1} = \cdots = \lim_{n\to\infty} a_{n-2} = \cdots = \lim_{n\to\infty} a_0 = \lim_{n\to\infty} 1 = 1$

$\sqrt[3]{3\sqrt[3]{3\sqrt[3]{3\sqrt[3]{3}}}}\ldots \rightsquigarrow x = \sqrt[3]{3x} \quad /^3$

$x^3 = 3x$

$x^2 = 3$

$x = \sqrt{3}$

$15 \cdot 3 = 45$

$10 \cdot 6 = 60$

funkcije slučajnih vektorjev:

$f: \mathbb{R}^n \to \mathbb{R}^n$

$X = (X_1, \ldots, X_n)$

$f(X) = (f_1(X), \ldots, f_n(X))$

$F_{f(X)}(x) = P[f(X) \leq x] = P[X \leq f^{-1}((-\infty, x_1], \ldots, (-\infty, x_n])]$

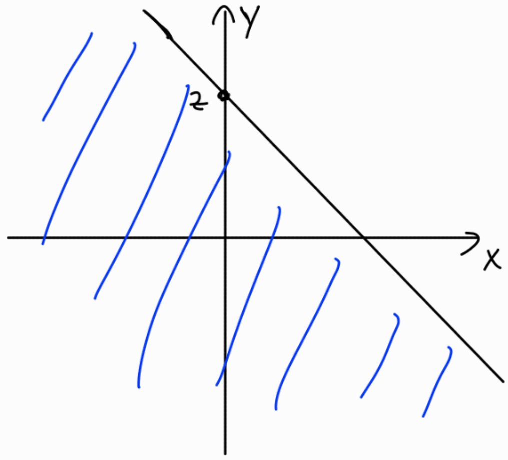
zgled: (X, Y), $f: \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = x + y$; (X,Y) porazdeljen zvezno z gostoto $p_{X,Y}$

$f(X,Y) = Z = X + Y$

$F_Z(z) = ?$ $\quad F_Z(z) = P[Z \leq z] = P[X+Y \leq z]$

Ta verjetnost je integral $p_{X,Y}$ po območju v $\mathbb{R}^2$, kjer je $X+Y \leq z$

$p_{X,Y}(x,y)$



$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} p_{X,Y}(x,y) \, dy \, dx$$

V primeru, da sta X in Y neodvisna, pa dobimo $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$

odvod

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} p_X(x) \cdot p_Y(y) \, dy \, dx$$

$$p_Z(z) = F_Z'(z) = \int_{-\infty}^{\infty} p_X(x) \cdot p_Y(z-x) \cdot 1 = \int_{-\infty}^{\infty} p_X(x) \cdot p_Y(z-x) = p_X * p_Y$$

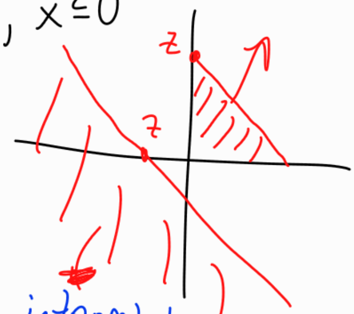
Zgled: X, Y neodvisni slučajni spremenljivki

$X \sim \chi^2(m)$ in
$Y \sim \chi^2(n)$

Kako je porazdeljen $Z = X + Y$

$$p_X(x) = \begin{cases} \dfrac{1}{2^{\frac{m}{2}} \cdot \Gamma(\frac{m}{2})} \cdot x^{\frac{m}{2}-1} e^{-\frac{x}{2}} & , x > 0 \\ 0 & , x \leq 0 \end{cases}$$

v bistvu $\int$ samo po trikotniku



$$p_Z(z) = (p_X * p_Y)(z)$$

integral bo 0 za $z \leq 0$

$$p_Z(z) = \frac{1}{2^{\frac{m+n}{2}} \cdot \Gamma(\frac{m}{2}) \cdot \Gamma(\frac{n}{2})} \int_0^z x^{\frac{m}{2}-1} \cancel{e^{-\frac{x}{2}}} \cdot (z-x)^{\frac{n}{2}-1} \cdot e^{-\frac{(z-\cancel{x})}{2}} \, dx$$

$$= \frac{e^{-\frac{z}{2}}}{2^{\frac{m+n}{2}} \cdot \Gamma(\frac{m}{2}) \cdot \Gamma(\frac{n}{2})} \cdot \int_0^z x^{\frac{m}{2}-1} \cdot (z-x)^{\frac{n}{2}-1} \, dx$$

$x = tz$

$t = \frac{x}{z}$

$dt = \frac{1}{z} dx$

$dx = z \cdot dt$

$$= \frac{e^{-\frac{z}{2}}}{2^{\frac{m+n}{2}} \cdot \Gamma(\frac{m}{2}) \cdot \Gamma(\frac{n}{2})} \cdot \int_0^1 (tz)^{\frac{m}{2}-1} \cdot (z-tz)^{\frac{n}{2}-1} \cdot z \, dt$$

$$= \frac{e^{-\frac{z}{2}}}{2^{\frac{m+n}{2}} \cdot \Gamma(\frac{m}{2}) \cdot \Gamma(\frac{n}{2})} \cdot \int_0^1 z^{\frac{m}{2}-1+\frac{n}{2}-1+1} \cdot t^{\frac{m}{2}-1} \cdot (1-t)^{\frac{n}{2}-1} \, dt$$

$$= \frac{e^{-\frac{z}{2}} z^{\frac{m+n}{2}-1}}{2^{\frac{m+n}{2}} \cdot \Gamma(\frac{m}{2}) \cdot \Gamma(\frac{n}{2})} \cdot \int_0^1 t^{\frac{m}{2}-1} \cdot (1-t)^{\frac{n}{2}-1} \, dt$$

$$B\left(\frac{m}{2}, \frac{n}{2}\right) = \frac{\Gamma(\frac{m}{2}) \cdot \Gamma(\frac{n}{2})}{\Gamma(\frac{m+n}{2})}$$

$$= \frac{e^{-\frac{z}{2}} z^{\frac{m+n}{2}-1}}{2^{\frac{m+n}{2}} \cdot \cancel{\Gamma(\frac{m}{2})} \cdot \cancel{\Gamma(\frac{n}{2})}} \cdot \frac{\cancel{\Gamma(\frac{m}{2})} \cdot \cancel{\Gamma(\frac{n}{2})}}{\Gamma(\frac{m+n}{2})} = \frac{e^{-\frac{z}{2}} \cdot z^{\frac{m+n}{2}-1}}{2^{\frac{m+n}{2}} \cdot \Gamma(\frac{m+n}{2})} \sim \chi^2(m+n)$$

izrek: Če je X porazdeljen $\chi^2(m)$ in sta X in Y neodvisna, potem
Y porazdeljen $\chi^2(n)$

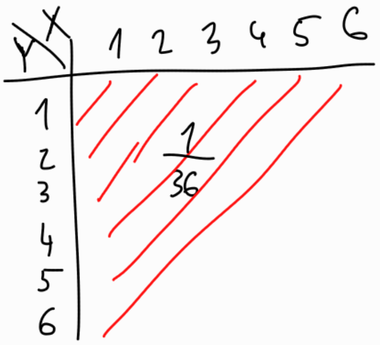
je $X+Y$ porazdeljen $\chi^2(m+n)$

Če je X porazdeljen $N(0,1)$, potem je X^2 porazdeljen $\chi^2(1)$

Če imamo slučajne spremenljivke $X_1, \dots, X_n$ vse porazdeljene $N(0,1)$, ki so med seboj neodvisne, potem je

$$\boxed{X_1^2 + \dots + X_n^2 \sim \chi^2(n)}$$

zgled: X, Y met neodvisnih kock



$$X+Y : \begin{pmatrix} 2 & 3 & 4 & \dots & 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \dots & \frac{1}{36} \end{pmatrix}$$

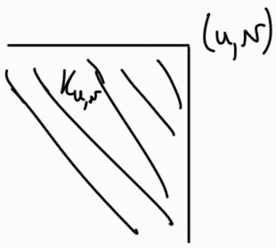
V primeru, da je f zvezno odvedljiva in bijektivna na območju, kjer je $\neq 0$

$f: \mathbb{R}^2 \to \mathbb{R}^2$

$p_{U,V}(u,v) = ?$

$f(X,Y)$

$K_{u,v} =$



(X,Y) zvezni z gostoto $p_{X,Y}(x,y)$

$K_{u,v} \subseteq \mathbb{R}^2$

$U = f_1(X,Y)$

$V = f_2(X,Y)$

$F_{U,V}(u,v) = P[(U,V) \in K_{u,v}] =$

$(U,V) = f(X,Y)$

$= P[f(X,Y) \in K_{u,v}] =$

$$= P\left[(X,Y) \in f^{-1}(K_{u,v})\right] =$$

$$= \iint_{f^{-1}(K_{u,v})} p_{X,Y}(x,y)\,dx\,dy =$$

iščemo pa $p_{U,V}(u,v)$

$$\left.\begin{array}{l} u = f_1(x,y) \\ v = f_2(x,y) \end{array}\right\} \Rightarrow \left.\begin{array}{l} x = x(u,v) \\ y = y(u,v) \end{array}\right\} f^{-1}(u,v) =$$

$$= (x(u,v), y(u,v))$$

Vstavimo novi spremenljivki $(u,v) = f(x,y)$

drug u kot pri $K_{u,v}$

$$= \iint_{K_{u,v}} p_{X,Y}(x(u',v'), y(u',v'))\ |J(u',v')|\,du\,dv \qquad J(u,v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix}_{(u,v)}$$

$$\Rightarrow p_{U,V}(u,v) = p_{X,Y}(x(u,v), y(u,v)) \cdot |J(u,v)|$$

$U = X \qquad X = U$

$V = f(X,Y)$

$y = y(u,v)$

$$J(u,v) = \begin{vmatrix} 1 & 0 \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial y}{\partial v}$$

V tem primeru je $p_V(v) = \int_{-\infty}^{\infty} p_{X,Y}(u, y(u,v)) \cdot \left|\frac{\partial y}{\partial v}\right| du$

V primeru, da je $V = X + Y$

$X = U$

$Y = V - X = V - U$

$y(u,v) = v - u$

$$p_V(v) = \int_{-\infty}^{\infty} p_{X,Y}(u, v-u)\,|1|\,du$$

$$= \int_{-\infty}^{\infty} p_{X,Y}(u, v-u)\,du$$

$X = U$

$V = X \cdot Y$

$Y = \frac{V}{X} = \frac{V}{U}$

$\frac{\partial y}{\partial v} = \frac{1}{u}$

$$p_V(v) = \int_{-\infty}^{\infty} p_{X,Y}\left(u, \frac{v}{u}\right) \cdot \left|\frac{1}{u}\right| du$$

Matematično upanje ("povprečna vrednost")

"pričakovana vrednost"

<u>def:</u> X, diskretno na končni množici

$$X : \begin{pmatrix} x_1, \dots, x_n \\ p_1, \dots, p_n \end{pmatrix}$$

expectation

matematično upanje

$$\boxed{E(X) = \sum_{i=1}^{n} x_i \cdot p_i = \sum_{i=1}^{n} x_i \cdot P[X = x_i]}$$

<u>zgled:</u> a) X enakomerno na $\{1, 2, \dots, n\}$

$$E(X) = \sum_{i=1}^{n} i \cdot P[X=i] = \sum_{i=1}^{n} i \cdot \frac{1}{n} = \frac{1}{\cancel{n}} \cdot \frac{(n+1) \cdot \cancel{n}}{2} = \frac{n+1}{2}$$

b) X enakomerno na $\{x_1, \dots x_n\}$

$$E(X) = \sum_{i=1}^{n} x_i \cdot \frac{1}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

dobimo povprečno vrednost X

c) $Ber(p)$ $\begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$

$$E(X) = 0 \cdot (1-p) + 1 \cdot p = p$$

1) $Bin(n,p) \sim X \qquad P[X = x_i] = \binom{n}{i} p^i \cdot (1-p)^{n-i}$

$$E(X) = \sum_{i=0}^{n} i \cdot \binom{n}{i} p^i \cdot (1-p)^{n-i} =$$

$$= \sum_{i=\cancel{0}\,1}^{n} \frac{\overset{n\cdot(n-1)!}{n!} \cdot \cancel{i}}{\underset{(i-1)!}{\cancel{i!}}\,(n-i)!} \, p^i \cdot (1-p)^i =$$

$$= n \cdot \sum_{i=1}^{n} \binom{n-1}{i-1} p^i \cdot (1-p)^{n-i} = \qquad i' = i-1$$

$$= np \cdot \underbrace{\sum_{i'=0}^{n-1} \binom{n-1}{i'} p^{i'} \, (1-p)^{n-1-i'}}_{1} =$$

$$E(X) = np$$

def: Če je X porazdeljena diskretno na neskončni množici, potem matematično upanje obstaja, če konvergira vrsta

$$\sum_{i=1}^{\infty} |x_i| \cdot p_i < \infty$$

, potem velja

$$E(X) = \sum_{i=1}^{\infty} x_i \cdot p_i$$

zgled: 1) $X \sim Geom(q)$

$$P[X = i] = p \cdot q^i \qquad p = 1-q$$

$i = 0, 1, \ldots$

$$E(X) = \sum_{i=0}^{\infty} i \cdot p \cdot q^i = p \cdot \sum_{i=0}^{\infty} i \cdot q^i = \frac{p \cdot q}{(1-q)^2}$$

$$\sum_{i=0}^{\infty} q^i = \frac{1}{1-q}$$

$$\sum_{i=\cancel{0}\,1}^{\infty} i\, q^i = q \cdot \sum_{i=1} i \cdot q^{i-1} = \frac{q}{(1-q)^2}$$

$$= \frac{p \cdot q}{p^2} = \frac{q}{p}$$

2) $X \sim Poi(n, p)$

$$P[X = n] = \frac{\lambda^n}{n!} e^{-\lambda} \qquad n = 0, 1, \ldots$$

$$E(X) = \sum_{n=\cancel{0}\,1}^{\infty} \cancel{n} \cdot \frac{\lambda^{n-1+1}}{\cancel{n!}\;(n-1)!} e^{-\lambda} = \lambda \underbrace{\sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!}}_{e^{\lambda}} e^{-\lambda} = \lambda$$

zvezne porazdelitve:

X, p_X gostota X

$$\boxed{E(X) = \int_{-\infty}^{\infty} x \cdot p_X(x)\, dx}$$, če obstaja integral $\int_{-\infty}^{\infty} |x| \cdot p_X(x)\, dx$

primeri: a) $X \sim U(a, b)$

$$p_X(x) = \begin{cases} \frac{1}{b-a} & ; a \leq x \leq b \\ 0 & ; \text{sicer} \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} x \cdot p_X(x)\, dx = \int_a^b x \cdot \frac{1}{b-a}\, dx = \frac{1}{b-a} \cdot \frac{x^2}{2} \Big|_a^b$$

$$= \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{\cancel{(b-a)}(b+a)}{2\cancel{(b-a)}} = \frac{a+b}{2}$$

b) $X \sim N(\mu, \sigma)$

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

$$E(X) = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$t = \frac{x-\mu}{\sigma}$ $\quad x = \sigma t + \mu$

$dt = \frac{dx}{\sigma}$

$dx = \sigma \cdot dt$

$$= \frac{1}{\sqrt{2\pi}\,\cancel{\sigma}} \cdot \int_{-\infty}^{\infty} (\sigma t + \mu) \cdot e^{-\frac{t^2}{2}} \cdot \cancel{\sigma}\, dt$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \left(\sigma \int_{-\infty}^{\infty} \underbrace{t e^{-\frac{t^2}{2}}}_{\text{liha}} dt + \mu \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt \right)$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \left(0 + \mu \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt \right)$$

$$= \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt =$$

analiza 3

$$= \frac{\mu}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = \underline{\underline{\mu}}$$

c) $\Gamma(b,c) \sim X \quad b, c > 0$

$$p_X(x) = \frac{c^b}{\Gamma(b)} x^{b-1} e^{-cx}$$

$b = b' - 1$

$b' = b + 1$

$$\boxed{\int_0^{\infty} x^{b'-1} \cdot e^{-cx}} = \frac{\Gamma(b')}{c^{b'}}$$

$$E(X) = \frac{c^b}{\Gamma(b)} \cdot \int_0^{\infty} \underset{x \cdot x^{b-1}}{x^b} \cdot e^{-cx} dx = \frac{c^b}{\Gamma(b)} \int_0^{\infty} x^{b'-1} e^{-cx} dx =$$

$$= \frac{c^b}{\Gamma(b)} \cdot \frac{\Gamma(b')}{c^{b'}} = \frac{\cancel{c^b} \cdot \cancel{\Gamma(b+1)} \cdot b}{\cancel{\Gamma(b)} \cdot c^{\cancel{b+1}}} = \frac{b}{c}$$

d) 1. Zgled slučajnih spremenljivk, ki nimata matematičnega upanja:

$$P\Big[X = \underbrace{\frac{2^n (-1)^n}{n}}_{x_n}\Big] = \frac{1}{2^n} \qquad n = 1, 2, \dots$$

harmonična vrsta

$$\sum_{n=1}^{\infty} |x_n| \, p_n = \sum_{n=1}^{\infty} \frac{\cancel{2^n}}{n} \cdot \frac{1}{\cancel{2^n}} = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

2. $X \sim$ Cauchy

$$p_X(x) = \frac{1}{\pi(1+x^2)}$$

soda

$$E(X) = \int_{-\infty}^{\infty} \frac{|x|}{(1+x^2) \cdot \pi} \, dx = \int_0^{\infty} \frac{2x}{(1+x^2)\pi} \, dx$$

$t = 1 + x^2$

$dt = 2x \, dx$

$$= \frac{1}{\pi} \int_1^{\infty} \frac{1}{t} \, dt = \frac{1}{\pi} \cdot \ln t \Big|_1^{\infty}$$

$$= \frac{1}{\pi} \cdot (\infty - 0) = \infty \quad \text{divergira}$$

X, f zvezna funkcija na zalogi vrednosti X

$f(X)$ ... če je X diskretna slučajna spremenljivka

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

$$E(f(X)) = \sum_{n=1}^{N(\infty)} f(x_n)\, p_n$$

$f(x_n)$

$p_{i_1} + p_{i_2} + \ldots + p_{i_k}$

(če ni injektivna in jih presliku več v isto

$f(X)$ ... če je X zvezna slučajna spremenljivka

p_X, $f: \mathbb{R} \to \mathbb{R}$ zvezno odvedljiva,

potem je $\boxed{E(f(X)) = \int_{-\infty}^{\infty} f(x)\, p_X(x)\, dx}$, če obstaja $\int_{-\infty}^{\infty} |f(x)| \cdot p_X(x)\, dx$

dokaz: (predpostavimo še bijektivnost)

naj bo $f: \mathbb{R} \to \mathbb{R}$ zvezno odvedljiva bijektivna

$$p_{f(X)}(x) = p_X(f^{-1}(x)) \cdot (f^{-1})'(x)$$

$$E(f(X)) = \int_{-\infty}^{\infty} x \cdot p_{f(X)}(x)\, dx = \int_{-\infty}^{\infty} x \cdot p_X(f^{-1}(x)) \cdot ((f^{-1})'(x))\, dx =$$

$t = f^{-1}(x)$

$f(t) \overset{\text{bijekcija}}{=} x$

$$= \int_{-\infty}^{\infty} f(t) \cdot p_X(t) \cdot 1\, dt$$

$t = f^{-1}(x) \quad \Big|'$

$1 = (f^{-1})'(x)$

$$= \int_{-\infty}^{\infty} f(t) \cdot p_X(t)\, dt \qquad \square$$

opomba: Slučajna spremenljivka X ima matematično upanje, natanko tedaj ko ima slučajna spremenljivka $|X|$ matematično upanje

trditev: Če X ima matematično upanje, potem velja

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

dokaz: X diskretna

$$E(|X|) = \sum_n |x_n| p_n = \sum_n |x_n||p_n| = \sum_n |x_n p_n| \overset{\triangle \text{ neenakost}}{\geq} \left|\sum_n x_n p_n\right| = |E(X)|$$

X zvezna:

$$|E(X)| = \left|\int_{-\infty}^{\infty} x \cdot p_X(x)\, dx\right| \leq \int_{-\infty}^{\infty} |x\, p_X(x)|\, dx = \int_{-\infty}^{\infty} |x|\, p_X(x)\, dx \quad \square$$

trditev: Če X ima matematično upanje in je $a \in \mathbb{R}$, potem

$$E(a \cdot X) = a \cdot E(X)$$

trditev: Naj bo $X = (X_1, \ldots, X_n)$ slučajni vektor in $Z = f(X_1, \ldots, X_n)$ za neko funkcijo f (zvezno).

Potem je za diskreten slučajni vektor X

$$E(Z) = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} f(x_1, \ldots, x_n)\, p_{x_1 \ldots x_n}$$

$$p_{x_1 \ldots x_n} = P[X_1 = x_1, \ldots, X_n = x_n]$$

$$E(Z) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\, p_X(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

Če vsote (integrali) absolutno konvergirajo

trditev: Velja $E(X+Y) = E(X) + E(Y)$, če $E(X)$ in $E(Y)$ obstajata

še več, E je linearna preslikava

$$E(a_1 X_1 + \ldots + a_n X_n) = a_1 E(X_1) + \ldots + a_n E(X_n)$$, če vsi $E(X_1), \ldots, E(X_n)$ obstajajo

dokaz: X, Y diskretni sl. spr.

$$E(X+Y) = \sum_i \sum_j (x_i + y_j)\, p_{ij} \qquad p_{ij} = P[X = x_i, Y = y_j]$$

$$= \sum_i x_i \left(\sum_j p_{ij}\right) + \sum_j y_j \left(\sum_i p_{ij}\right)$$

$$= \sum_i x_i\, q_i + \sum_j y_j \cdot r_j$$

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

$$\begin{array}{c|ccc|c}
 & x_1 & \dots & x_n & \\
\hline
y_1 & p_{11} & \dots & p_{1n} & q_1 \\
\vdots & \vdots & & & \vdots \\
y_n & p_{n1} & \dots & p_{nn} & q_n \\
\hline
 & r_1 & \dots & r_n &
\end{array}$$

X, Y zvezni

$$E(X+Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y)\, p_{XY}(x,y)\, dx\, dy =$$

$$= \int_{-\infty}^{\infty} x \underbrace{\left(\int_{-\infty}^{\infty} p_{X,Y}(x,y)\, dy\right)}_{p_X(x)} dx + \int_{-\infty}^{\infty} y \underbrace{\left(\int_{-\infty}^{\infty} p_{X,Y}(x,y)\, dx\right)}_{p_Y(y)} dy$$

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

posledica: X ima matematično upanje.

Potem je $E(X - E(X)) = 0$

$$E(X - E(X)) = E(X) - E(E(X)) =$$

$$= E(X) - E(X) = 0$$

$$Y \sim \begin{pmatrix} E(X) \\ 1 \end{pmatrix}$$

$X_i \sim Ber(p) \quad X_i \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$

$i = 1, \ldots, n$

$Y = X_1 + \ldots + X_n \sim Bin(n, p)$

$E(X_i) = p$

$E(Y) = n \cdot p = np$

trditev: X, Y neodvisni slučajni spremenljivki, ki imata matematično upanje.

Potem je $E(XY) = E(X) \cdot E(Y)$

dokaz: X, Y zvezni

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot y \cdot p_X(x) \cdot p_Y(y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty} x\, p_X(x) \cdot \left(\int_{-\infty}^{\infty} y\, p_Y(y)\, dy\right) dx$$

$$= \int_{-\infty}^{\infty} x\, p_X(x) \cdot E(Y)\, dx =$$

$$= E(Y) \cdot \int_{-\infty}^{\infty} x \cdot p_X(x)\, dx =$$

$$= E(Y) \cdot E(X) = E(X) \cdot E(Y)$$

X, Y diskretni

$p_i = P[X = x_i]$

$q_j = P[Y = y_j]$

$$E(XY) = \sum_i \sum_j x_i y_j \cdot p_i \cdot q_j$$

$$= \sum_i x_i p_i \cdot \sum_j y_j q_j = E(X) \cdot E(Y)$$

izrek: (Cauchy-Schwarzova neenakost) $\langle u, v\rangle \leq \|u\| \|v\|$

X, Y slučajni spremenljivki, ki imata $E(X^2)$ in $E(Y^2)$ $\neq 0$

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

enakost velja, natanko tedaj ko je $|Y| = \sqrt{\frac{E(Y^2)}{E(X^2)}}\, |X|$ z verjetnostjo 1.

dokaz: $a, b > 0$ $\quad ab \leq \frac{1}{2}(a^2+b^2)$

$$\underbrace{\sqrt{ab}}_{\text{geo.}} \leq \underbrace{\frac{1}{2}(a+b)}_{\text{aritm.}}$$

$$(a-b)^2 \geq 0$$
$$a^2 - 2ab + b^2 \geq 0$$
$$a^2 + b^2 \geq 2ab \quad /:2$$
$$\frac{1}{2}(a^2+b^2) \geq ab$$

$=$ natanko tedaj ko $(a-b)^2 = 0$, $a = b$

Če sta U in V nenegativni slučajni spremenljivki, potem velja

$$U \cdot V \leq \frac{1}{2}(U^2 + V^2)$$

če izberemo $U = \alpha \cdot |X|$, $\alpha > 0$, $V = \frac{1}{\alpha}|Y|$, potem imamo $|XY| \leq \frac{1}{2}\left(\alpha^2 |X^2| + \frac{1}{\alpha^2}|Y|^2\right)$

$$UV = |X||Y| = |XY|$$

$$\Rightarrow E(|XY|) = \frac{1}{2} E\left(\alpha^2 |X^2| + \frac{1}{\alpha^2}|Y^2|\right)$$
$$= \frac{1}{2}\left(\alpha^2 E(|X|^2) + \frac{1}{\alpha^2} E(|Y^2|)\right)$$

Če je $X \leq Y$, potem je $E(X) \leq E(Y)$:

$$Y - X \geq 0$$
$$\Rightarrow E(Y-X) \geq 0$$
$$E(Y) - E(X) \geq 0$$
$$E(Y) \geq E(X)$$

$$\alpha^2 = \sqrt{\frac{E(Y^2)}{E(X^2)}}$$

$$E(|XY|) = \frac{1}{2}\left(\sqrt{E(|X|^2) \cdot E(|Y|^2)} + \sqrt{E(|X^2|) \cdot E(|Y^2|)}\right)$$
$$= \sqrt{E(|X|^2) \cdot E(|Y|^2)}$$

enačaj velja, kadar $U=V$ z verjetnostjo 1:

$$\alpha \cdot |X| = \frac{1}{\alpha}|Y|$$

$$\alpha^2 |X| = |Y|$$

$$\sqrt{\frac{E(Y^2)}{E(X^2)}} = |Y| \qquad \square$$

<u>def:</u> Če sta X in Y taki, da imata matematično upanje, potem rečemo, da sta <u>nekorelirani</u>, če velja $E(XY) = E(X) \cdot E(Y)$

<u>posledica:</u> X, Y neodvisni $\Rightarrow$ X, Y nekorelirani

obratno NE VELJA!

<u>zgled:</u> $U : \begin{pmatrix} 0 & \frac{1}{2} & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$

Naj bo $X = \sin(\pi \cdot U)$ $\quad X : \begin{pmatrix} 0 & 1 \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}$ $\quad E(X) = \frac{1}{3}$

$Y = \cos(\pi \cdot U)$ $\quad Y : \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$ $\quad E(Y) = 0$

$E(X) \cdot E(Y) = 0$

$Y \backslash X$	0	1	
-1	$\frac{1}{3}$	0	$\frac{1}{3}$
0	0	$\frac{1}{3}$	$\frac{1}{3}$
1	$\frac{1}{3}$	0	$\frac{1}{3}$
	$\frac{1}{3}$	$\frac{1}{3}$	

$U=1$

$U=\frac{1}{2}$

$U=0$

$XY : \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ $\quad E(XY) = 0$

$E(XY) = E(X) \cdot E(Y) \Rightarrow$ sta nekorelirani

nista pa neodvisni

Momenti slučajne spremenljivke X so matematična upanja

$E(X), E(X^2), E(X^3), \ldots$ (če obstajajo)

disperzija (varianca) in kovarianca

def: X slučajna spremenljivka, za katero obstaja $E(X^2)$. Potem gledamo

$D(X) = \mathrm{Var}(X) = E\left(\left(X - E(X)\right)^2\right)$ imenujemo disperzija ali varianca slučajne spremenljivke X

$E(E(X)) = E(X)$

velja: $D(X) = E\left(\left(X - E(X)\right)^2\right) = E\left(X^2 - 2X\,E(X) + E(X)^2\right) =$

$= E(X^2) - 2\,E(X)^2 + E(X)^2 =$

$= \boxed{E(X^2) - E(X)^2}$

lastnosti disperzije: ("meri" razpršenost)

1) $D(X) \geq 0$

2) $D(\alpha X) = \alpha^2 \cdot D(X)$

3) $D(X) \leq E\left((X-a)^2\right)$ za poljuben $a \in \mathbb{R}$

$=$, ko $a = E(X)$

dokaz: $E\left((X-a)^2\right) = E\left(X^2 - 2aX + a^2\right) = \underbrace{E(X^2)}_{D(X)} - 2a\,E(X) + a^2 + E(X)^2 - \underline{E(X)^2}$

$= D(X) + \underbrace{\left(E(X) - a\right)^2}_{\geq 0} \geq D(X)$ $\square$

pregled disperzij nekaterih porazdelitev:

1) $Bin(n,p)$: $E(X) = n \cdot p$, $D(X) = n \cdot p \cdot q$
$q = 1-p$

2) $Poi(\lambda)$: $E(X) = \lambda = D(X)$

3) $Geo(p)$: $E(X) = \frac{1}{p}$ $D(X) = \frac{q}{p^2}$

4) $Pas(m,p)$: $E(X) = \frac{m}{p}$ $D(X) = \frac{mq}{p^2}$

5) Enakomerna $U(a,b)$: $E(X) = \frac{a+b}{2}$, $D(X) = \frac{(b-a)^2}{12}$

$U(0,1)$ $E(X) = \frac{1}{2}$

$$D(X) = \int_0^1 \left(x - \frac{1}{2}\right)^2 dx =$$
$$= \int_0^1 \left(x^2 - x + \frac{1}{4}\right) dx =$$
$$= \left(\frac{1}{3}x^3 - \frac{1}{2}x^2 + \frac{1}{4}x\right)\Big|_0^1 = \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{1}{12}$$

6) $\Gamma(b,c)$: $E(X) = \frac{b}{c}$ $D(X) = \frac{b}{c^2}$

7) $\chi^2(n)$: $E(X) = n$ $D(X) = 2n$

8) $Exp(\lambda)$: $E(X) = \frac{1}{\lambda}$ $D(X) = \frac{1}{\lambda^2}$

9) $N(\mu,\sigma)$ $E(X) = \mu$ $D(X) = \sigma^2$

$$D(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx =$$

$y = \frac{x-\mu}{\sigma}$

$dy = \frac{1}{\sigma} dx$

$(x-\mu) = \sigma y$

$$= \frac{1}{\sqrt{2\pi}\cancel{\sigma}} \int_{-\infty}^{\infty} (\sigma y)^2 e^{-\frac{1}{2}y} \cancel{\sigma} dy$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-\frac{1}{2}y} dy$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = \sigma^2$$

<u>def:</u> <u>standardni odklon</u> slučajne spremenljivke X je $\sigma(X) = \sqrt{D(X)}$

$N(\mu, \sigma)$: $\sigma(X) = \sigma$

<u>def:</u> (X,Y) slučajni vektor, obstajata $E(X^2)$, $E(Y^2)$. <u>Kovarianca</u> slučajnih spremenljivk X in Y je $\boxed{K(X,Y) = E\big((X-E(X))\cdot(Y-E(Y))\big)}$

opazimo: $K(X,Y) = E\big(XY - XE(Y) - E(X)Y + E(X)\cdot E(Y)\big)$

$$= E(XY) - E(Y)E(X) - E(X)E(Y) + E(X)\cdot E(Y)$$

$$\boxed{= E(XY) - E(X)E(Y)}$$

<u>lastnosti:</u>

1) $K(X,X) = D(X)$
2) če sta X, Y nekorelirani $\Rightarrow K(X,Y) = 0$
3) $K(X,Y) = K(Y,X)$
4) $K(aX+bY, Z) = a\cdot K(X,Z) + b\cdot K(Y,Z)$ linearnost
5) $|K(X,Y)| \leq \sqrt{D(X)\cdot D(Y)} = \sigma(X)\cdot\sigma(Y)$

To dobimo iz Cauchy-Schwarz neenakosti za $X-E(X)$ in $Y-E(Y)$

<u>dokaz:</u> $|K(X,Y)| = \big|E\big((X-E(X))\cdot(Y-E(Y))\big)\big| \leq E\big(|X-E(X)|\cdot|Y-E(Y)|\big)$

$$\overset{C.S.}{\leq} \sqrt{E\big((X-E(X))^2\big)\cdot E\big((Y-E(Y))^2\big)}$$

$$= \sqrt{D(X)\cdot D(Y)}$$

$= $ velja, natanko tedaj ko je $\left|Y-E(Y)\right| = \frac{\sigma(Y)}{\sigma(X)}\left|X-E(X)\right|$

6) $$D(X+Y) = D(X)+D(Y)+2K(X,Y)$$

Če sta X in Y nekorelirani, velja $D(X+Y) = D(X)+D(Y)$

<u>dokaz:</u> $D(X+Y) = E((X+Y)^2) - E(X+Y)^2$

$$= E(X^2) + 2\cdot E(XY) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2$$

$$= D(X) + D(Y) + 2\cdot K(X,Y)$$

7) Z indukcijo dobimo

$$D\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} D(X_i) + 2\cdot \sum_{1\leq i<j\leq n} K(X_i, Y_j)$$

Nova predavanja

→ def: X ima $E(X), D(X)$, potem slučajno spr. $X_S = \dfrac{X - E(X)}{\sigma(X)}$ imenujemo standardizacija slučajne spr. X.

Velja: $E(X_S) = \frac{1}{\sigma(X)} E(X - E(X)) = \frac{1}{\sigma(X)} (E(X) - E(X)) = \underline{0}$

$D(X_S) = D\left(\frac{1}{\sigma(X)} (X - E(X))\right) = \frac{1}{\sigma^2(X)} D(X - E(X)) = \dfrac{D(X)}{D(X)} = \underline{1}$

zgled: $X \sim N(\mu, \sigma) \Rightarrow X_S = \frac{X - \mu}{\sigma} \sim N(0,1)$

→ def: X, Y slučajni spr., ki imata disperzijo, potem velja $|K(X_S, Y_S)| \leq 1$

C.S. $|K(X,Y)| \leq \sigma(X) \cdot \sigma(Y)$

$$K(X,Y) = E((X - E(X))(Y - E(Y))) = \sigma(X)\, \sigma(Y)\, \underset{\underset{E((X_S - 0)(Y_S - 0))}{\parallel}}{E(X_S, Y_S)} = \sigma(X)\, \sigma(Y)\, K(X_S, Y_S)$$

če je $E(X) = E(Y) = 0$, je $K(X,Y) = E(XY)$

$|K(X,Y)| = |\sigma(X) \cdot \sigma(Y)| \cdot |K(X_S, Y_S)| \overset{C.S.}{\leq} \sigma(X)\, \sigma(Y) \cdot \overset{=1}{\sigma(X_S)} \cdot \overset{=1}{\sigma(Y_S)}$ pokrajšamo $\sigma(X) \cdot \sigma(Y)$

$\Rightarrow |K(X_S, Y_S)| \leq 1$

→ def: X, Y sluč. spr. ki imata disperzijo

Potem: $r(X,Y) = E(X_S Y_S) = K(X_S, Y_S) = \dfrac{K(X,Y)}{\sigma(X)\sigma(Y)}$ KORELACIJSKI KOEFICIENT SLUČAJNIH SPR. X in Y

Lastnosti:
1) $-1 \leq r(X,Y) \leq 1$
2) če X, Y nekorelirani (posebej če sta neodvisni) potem $r(X,Y) = 0$
3) $r(X,Y) = 1 \Leftrightarrow Y = \dfrac{\sigma(X)}{\sigma(Y)} (X - E(X)) + E(Y)$ z verjetnostjo 1.

$|r(X,Y)| = 1 \Leftrightarrow$ imamo $=$ v C.S. neenakosti

4.) $r(X,Y) = -1 \Leftrightarrow Y = -\dfrac{\sigma(X)}{\sigma(Y)} (X - E(X)) + E(Y)$ z verjetnostjo 1

5.) $|r(X,Y)| = 1 \Leftrightarrow Y = aX + b$ za neka skalarja a, b z verjetnostjo 1.

zgled: $(X,Y) \sim N(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$

$$-1 < \rho < 1, \quad \sigma_x, \sigma_y > 0, \quad \mu_x, \mu_y \in \mathbb{R}$$

$$p_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}\,\sigma_x \sigma_y \sqrt{1-\rho^2}}\, e^{-\frac{1}{2}\left(\frac{(x-\mu_x)^2 - 2\rho(x-\mu_x)(y-\mu_y) + (y-\mu_y)^2}{(1-\rho^2)\sigma_x\sigma_y}\right)}$$

izračunajmo korelacijski koeficient $r(X,Y)$:

$$\overbrace{x^2 - 2x(\rho y) + (\rho y)^2}^{(x-\rho y)^2} \underbrace{-(\rho y)^2 + y^2}_{y^2(1-\rho^2)}$$

$$(X_s, Y_s) = N(0,0,1,1,\rho)$$

$$r(X,Y) = E(X_s, Y_s) = \frac{1}{\sqrt{2\pi}^2 \cdot \sqrt{1-\rho^2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x y\, e^{-\frac{1}{2}\left(\frac{x^2 - 2xy\rho + y^2}{(1-\rho^2)}\right)} dx\, dy$$

dimenzija slučajnega vektorja

$$= \frac{1}{\sqrt{2\pi}^2 \cdot \sqrt{1-\rho^2}} \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} y \cdot e^{-\frac{1}{2}y^2} \cdot x \cdot e^{-\frac{1}{2}\frac{(x-\rho y)^2}{1-\rho^2}} dx\right) dy$$

$$= \frac{1}{\sqrt{2\pi}^2 \sqrt{1-\rho^2}} \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} x\, e^{-\frac{1}{2}\left(\frac{x-\rho y}{\sqrt{1-\rho^2}}\right)^2} dx\right) \cdot y\, e^{-\frac{1}{2}y^2} dy$$

brez ²   upanje od $N(\rho y,\ 1-\rho^2) = \rho y$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \rho y \cdot y \cdot e^{-\frac{1}{2}y^2} dy = \underline{\underline{\rho}}$$

disperzija $N(0,1) = 1$

posledica:

$r(X,Y) = 0$ za normalni vektor (X,Y) $\iff$ X,Y sta neodvisna.

X, Y

	x_1	x_2
y_1	p_{11}	p_{12}
y_2	p_{21}	p_{22}

$$\begin{array}{cc} q_1r_1 & q_1r_2 \\ \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \\ q_2r_1 & q_2r_2 \end{array} = \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} \begin{bmatrix} r_1 & r_2 \end{bmatrix} \iff \text{rang } 1$$

X, Y sta neodvisna $\Leftrightarrow$ rang matrike $\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$ ima rang 1

$\Updownarrow$ $r(X,Y)=0$

POGOJNE PORAZDELITVE
in
POGOJNO MATEMATIČNO UPANJE

(X,Y) slučajni vektor

B nek dogodek, $P(B)>0$

X slučajna spremenljivka

pogojna porazdelitvena funkcija

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

1.) Diskretni primer (X,Y):

I, J števni

$\{x_i\}_{i\in I}, \{y_j\}_{j\in J}$

$p_{ij} = P[X=x_i, Y=y_j]$

$p_i = P[X=x_i]$

$q_j = P[Y=y_j]$

Za B vzamemo dogodek $[Y=y_j]$

$$F_X(x|y_j) = \sum_{x_i\le x} p_{ij} \cdot \frac{1}{q_j}$$

pogojna porazdelitvena funkcija za X glede na $[Y=y_j]$

$\frac{1}{P[Y=y_j]}$

pogojni porazdelitveni funkciji $F_X(x|y_j)$ pripada diskretna porazdelitev:

$P[X=x_i | Y=y_j] > 0$ za nekatere x_i, sicer $F_X(x|y_j)$ nima skokov vpeljemo oznako

$$p_{i|j} := P[X = x_i \mid Y = y_j] = \frac{\mathbb{P}[X = x_i, Y = y_j]}{P[Y = y_j]} = \frac{p_{ij}}{q_j}$$

sin/cos primer:

	$Y \backslash X$	-1	0	1	
$y_1 = 0$	0	0	$\frac{1}{3}$	0	$\frac{1}{3}$
$y_2 = 1$	1	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{2}{3}$

$$p_{i|0} \sim \begin{pmatrix} -1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$p_{i|1} \sim \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

zgled: X, Y neodvisni

$\forall j: \quad p_{i|j} = \frac{p_{ij}}{q_j} = \frac{p_i \cdot \cancel{q_j}}{\cancel{q_j}} = p_i$ pogojne porazdelitve so enake robni porazdelitvi

za pogojno porazdelitev $(p_{i|j})_{i \in I}$ lahko izračunamo njeno matematično upanje:

$$E(X \mid Y = y_j) = \sum_{i \in I} x_i \, p_{i|j}$$

vsa matematična upanja $E(X \mid Y = y_j)$ nam dajo novo slučajno spremenljivko

$$E(X|Y) : \begin{pmatrix} E(X \mid Y = y_j) \\ q_j \end{pmatrix}_{j \in J}$$

imenujemo jo pogojno matematično upanje X glede na Y

zgled:

$Y \backslash X$	0	1	
0	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{3}$
1	$\frac{3}{4}$	$\frac{5}{12}$	$\frac{2}{3}$
	$\frac{1}{2}$	$\frac{1}{2}$	

$E(X|Y) \sim ?$

$$E(X \mid Y = 0) = \frac{3}{4} \cdot 0 + \frac{1}{4} \cdot 1 = \frac{1}{4}$$

$$E(X \mid Y = 1) = \frac{\frac{3}{4} \cdot 0 + \frac{5}{12} \cdot 1}{\frac{2}{3}} = \frac{5 \cdot \cancel{3}}{2 \cdot \cancel{12} \cdot 4} = \frac{5}{8}$$

$$E(X|Y) \sim \begin{pmatrix} \frac{1}{4} & \frac{5}{8} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

$$E(E(X|Y)) = \frac{1}{4}\cdot\frac{1}{3} + \frac{5}{8}\cdot\frac{2}{3} = \frac{1}{12} + \frac{5}{12} = \frac{1}{2} = E(X)$$

Nova predavanja

1.)

<u>trditev:</u>

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

<u>dokaz:</u> $$E(X|Y=y_j) = \sum_i x_i \cdot p_{i|j}$$

$$P(E(X|Y) = E(\underbrace{X|Y=y_j}_{\in \mathbb{R}})) = q_j$$

$$E(E(X|Y)) = \sum_j \Big(\sum_i x_i \cdot p_{i|j}\Big) \cdot q_j =$$

$$= \sum_j \sum_i x_i \cdot \frac{p_{ij}}{\cancel{q_j}} \cdot \cancel{q_j} =$$

$$= \sum_i x_i \sum_j p_{ij} = \sum_i x_i \cdot p_i = E(X)$$

2) Če sta X in Y neodvisni slučajni spremenljivki, potem je

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

<u>dokaz:</u> za $\forall j$ $$E(X|Y=y_j) = \sum_i x_i \cdot p_{i|j} = \sum_i x_i \cdot \frac{p_{ij}}{q_j} = \sum_i x_i \cdot \frac{p_i \cdot q_j}{q_j} =$$

$$= \sum_i x_i \cdot p_i = E(X)$$

zgled: Mečemo kocko, dokler ne dobimo 6. Naj bo Y število metov do prve 6 in X naj bo število enic, ki je padlo.

Poišči $E(X \mid Y)$.

$y = 1, 2, \dots$

$$P[Y = y] = \left(\frac{5}{6}\right)^{y-1} \cdot \frac{1}{6}$$

$$E(X \mid Y = y) \qquad (X \mid Y=y) \sim \begin{pmatrix} 0 & 1 & \cdots & (y-1) \\ ? & ? & & ? \end{pmatrix}$$

$$\text{Bin}\left(y-1, \frac{1}{5}\right)$$

$A \dots$ pade 1

$\overline{A} \dots$ ne pade 1

$$\underbrace{\sqcup \quad \sqcup \quad \dots \quad \sqcup}_{y-1}$$

$p = \frac{1}{5}$ binomska porazdelitev

$$P\left[E(X \mid Y=y) = k\right] = \binom{y-1}{k} \cdot \left(\frac{4}{5}\right)^{y-1-k} \cdot \left(\frac{1}{5}\right)^{k}$$

$$E(X \mid Y=y) = (y-1) \cdot \frac{1}{5}$$

$$E(X \mid Y) = \begin{pmatrix} E(X \mid Y = y_j) \\ q_j \end{pmatrix}_{\forall j}$$

$$P\left(E(X \mid Y) = (y-1) \cdot \frac{1}{5}\right) = \left(\frac{5}{6}\right)^{y-1} \cdot \frac{1}{6}$$

$$\underline{E(X \mid Y=y)} = \boxed{\varphi(Y)} = \frac{1}{5}(y-1)$$

regresijska funkcija

Velja, da je $E(X \mid Y) = \varphi(Y)$

Slučajna spremenljivka $E(X \mid Y)$ je vedno oblike $\varphi(Y)$ za neko funkcijo φ

V zgledu je $E(X|Y) = \frac{1}{5}(Y-1)$

3) $E(f(Y)|Y) = f(Y)$

Če je $X = f(Y)$, je $E(X|Y) = X$

$$p_{i|j} = \begin{cases} 1; & i=j \\ 0; & i \neq j \end{cases}$$

p_{ij}	$f(y_1)$	....	$f(y_m)$
y_1	p_1		0
$\vdots$		$\ddots$	
y_m	0		p_1

če je bijektivna

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

5) $E(aX + bZ | Y) = a E(X|Y) + b E(Z|Y)$

6) $X \geq 0$, potem je tudi $E(X|Y) \geq 0$

7) $E(\underbrace{E(X|Y)}_{\varphi(Y)} | Y) = \underbrace{E(X|Y)}_{\varphi(Y)}$ idempotentnost

<u>zgled:</u> Y ... število metov do prve šestice

X ... število enic pri tem

$E(X|Y) = \frac{1}{5}(Y-1)$

poišči $E(X^2|Y)$

$E(X|Y=y) = \frac{1}{5}(y-1)$ $\quad\quad$ $Bin(y-1, \frac{1}{5})$

$$\mathrm{Var}(X \mid Y=y) = \frac{4}{25}(y-1)$$

$$\mathrm{Var}(X \mid Y=y) = E(X^2 \mid Y=y) - E(X \mid Y=y)^2$$

$$E(X^2 \mid Y=y) = \mathrm{Var}(X \mid Y=y) + E(X \mid Y=y)^2$$

$$= \frac{4}{25}(y-1) + \frac{1}{25}(y-1)^2 =$$

$$= \frac{4}{25}y - \frac{4}{25} + \frac{1}{25}y^2 - \frac{2}{25}y + \frac{1}{25} =$$

$$= \frac{1}{25}y^2 + \frac{2}{25}y - \frac{3}{25} =$$

$$E(X^2 \mid Y=y) = \frac{1}{25}(y^2 + 2y - 3)$$

$$E(X^2 \mid Y) = \frac{1}{25}(Y^2 + 2Y - 3)$$

Zgled: $X_1, X_2, \ldots$ neodvisne enako porazdeljene slučajne spremenljivke
škode
imajo matematično upanje μ in varianco σ^2

N diskretna slučajna spremenljivka, ki zavzame vrednosti v $\mathbb{N}$ in je neodvisna od $X_1, X_2, \ldots$

Naj bo $S = \sum_{i=1}^{N} X_i$ (slučajna vsota)

S ↓ skupna škoda

$E(N)$ in $E(N^2)$ obstajata. Kolikšna sta $E(S)$ in $\mathrm{Var}(S)$?

$$S_n = \sum_{i=1}^{n} X_i \qquad n \in \mathbb{N}$$

poiščimo $E(S \mid N=n)$:

$$E(S \mid N=n) = E\left(\sum_{i=1}^{n} X_i\right) = n \cdot \mu$$

$$E(S \mid N) = \mu \cdot N$$

$$E(E(S \mid N)) = E(S)$$

$$\| $$

$$E(\mu \cdot N) = \mu \cdot E(N)$$

$$\boxed{E(S) = \mu \cdot E(N)}$$

$$E(S^2 \mid N=n) = Var(S^2 \mid N=n) + E(S \mid N=n)^2 =$$

$$= Var(S_n) + (n \cdot \mu)^2 = n \cdot \sigma^2 + n^2 \mu^2$$

$$E(S^2 \mid N) = \sigma^2 \cdot N + \mu^2 N^2 =$$

$$Var(S) = E(S^2) - E(S)^2 =$$

$$= E(E(S^2 \mid N)) - (\mu \cdot E(N))^2 =$$

$$= E(\sigma^2 N + \mu^2 N^2) - \mu^2 E(N)^2 =$$

$$= \sigma^2 E(N) + \mu^2 E(N^2) - \mu^2 E(N)^2 =$$

$$= \sigma^2 E(N) + \mu^2 \cdot Var(N)$$

<u>zgled:</u> Kokoš znese N jajc, kjer je $N \sim Pois(\lambda)$. Iz vsakega jajca se izvali piščanec z verjetnostjo $p \in (0,1)$ neodvisno drug od drugega.

Naj bo K število piščancev. Poišči naslednja matematična upanja.

$E(K \mid N)$, $E(K)$, $E(N \mid K)$

$$E(K \mid N=n) \sim Bin(n,p)$$

$$E(K \mid N=n) = n \cdot p$$

$$E(K \mid N) = p \cdot N$$

$$E(K) = E(E(K \mid N)) =$$

$$= E(p \cdot N) = p \cdot E(N) =$$

$$= p \cdot \lambda$$

$$E(N \mid K) = ?$$

$n \geq k$

$$P(N=n \mid K=k) = \frac{P(K=k \mid N=n) \cdot P(N=n)}{P(K=k)} = \frac{\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \cdot \frac{\lambda^n \cdot e^{-\lambda}}{n!}}{P(K=k)}$$

$$P(K=k) = \sum_{n=k}^{\infty} P(K=k \mid N=n) \cdot P(N=n)$$

$$= \sum_{n=k}^{\infty} \binom{n}{k} p^k \cdot (1-p)^{n-k} \cdot \frac{\lambda^n \cdot e^{-\lambda}}{n!}$$

$$= e^{-\lambda} \cdot \sum_{n=k}^{\infty} \frac{\cancel{n!} \cdot p^k (1-p)^{n-k} \cdot \lambda^{n-k} \cdot \lambda^k}{k!(n-k)! \cdot \cancel{n!}}$$

$$= \frac{e^{-\lambda}}{k!}(p\lambda)^k \underbrace{\sum_{n=k}^{\infty} \frac{(1-p)^{n-k} \cdot \lambda^{n-k}}{(n-k)!}}_{e^{\lambda(1-p)}} =$$

$$= \frac{(p\lambda)^k}{k!} \cdot e^{-\lambda+\lambda-p\lambda} =$$

$$\boxed{P(K=k) = \frac{(p\lambda)^k}{k!} e^{-p\lambda}}$$

$$K \sim Pois(p\lambda)$$

Novo predavanje

$$P(N=n \mid K=k) = \frac{\binom{n}{k} \cancel{p^k} (1-p)^{n-k} \cdot \frac{\lambda^n \cdot e^{-\lambda}}{n!}}{\frac{\cancel{p^k} \cdot \lambda^k}{k!} e^{-p\lambda}} =$$

$$= \frac{\cancel{n!}\, (1-p)^{n-k} \lambda^{n-k} \cdot \cancel{k!}}{\cancel{k!}\,(n-k)! \cdot e^{-p} \cdot \cancel{n!}} = \frac{(1-p)^{n-k} \cdot \lambda^{n-k} \cdot e^{-\lambda+\lambda p}}{(n-k)!}$$

$$= \frac{\lambda^{n-k} \cdot (1-p)^{n-k}}{(n-k)!} \cdot e^{-\lambda(1-p)} = \frac{(\lambda \cdot (1-p))^{n-k}}{(n-k)!} \cdot e^{-\lambda(1-p)}$$

$$N \mid K=k \sim k + \mathrm{Poi}(\lambda \cdot (1-p))$$

$$E(N \mid K=k) = k + E(\mathrm{Poi}(\lambda \cdot (1-p)))$$
$$= k + \lambda(1-p)$$

$$E(N \mid K) = K + \lambda(1-p)$$

pogojno matematično upanje za zvezne porazdelitve

(X,Y) $\quad p_{X,Y}(x,y)$ gostota

$p_X(x), p_Y(y)$ robni gostoti

$$E(X \mid Y) = ?$$

Vzamemo $(y, y+h]$ polodprt interval, $h > 0$

y, h tako, da je

$$P(y \leq Y \leq y+h) > 0$$

$$\parallel$$

$$\int_{y}^{y+h} p_Y(t)\, dt$$

dogodek

$$B_y = [y < Y \leq y+h] \qquad \mathbb{P}(B_y) = F_Y(y+h) - F_Y(y)$$

iščemo pogojno porazdelitveno funkcijo

$$F_X(x \mid B_Y) = \mathbb{P}(X \leq x \mid B_Y) \overset{\text{Bayes}}{=} \frac{P(X \leq x,\, y < Y \leq y+h)}{\mathbb{P}(y < Y \leq y+h)}$$

$$= \frac{F_{X,Y}(x, y+h) - F_{X,Y}(x,y)}{F_Y(y+h) - F_Y(y)}$$

privzamemo, da obstaja limita, ko gre $h \downarrow 0$

$$F(X \mid Y=y) = \lim_{h \downarrow 0} \frac{\left(F_{X,Y}(x, y+h) - F_{X,Y}(x,y)\right) \cdot \frac{1}{h}}{\left(F_Y(y+h) - F_Y(y)\right) \cdot \frac{1}{h}}$$

je zvezna porazdelitev, njeno gostoto imenujemo pogojna gostota X pri pogoju $Y=y$

$$= \frac{\frac{\partial}{\partial y} F_{X,Y}(x,y)}{p_Y(y)} = \frac{\int_{-\infty}^{x} p_{X,Y}(s,y)\, ds}{p_Y(y)}$$

odvod porazdelitvene funkcije je gostota

$$\boxed{F(X \mid Y) = \frac{1}{p_Y(y)} \cdot \int_{-\infty}^{x} p_{X,Y}(s,y)\, ds}$$

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

Poiščemo $E(X|Y=y) = \dfrac{1}{p_Y(y)} \displaystyle\int_{-\infty}^{\infty} x \cdot p_{X,Y}(x,y)\, dx = f(y)$

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

lastnosti: enake kot pri diskretnem primeru

1) $E(X|Y)$ je linearna v X

2) X, Y neodvisni $\Rightarrow E(X|Y) = E(X)$

3) če je $X = f(Y) \Rightarrow E(X|Y) = X = f(Y)$

izrek: Naj bo $f: \mathbb{R} \to \mathbb{R}$ funkcija.

Potem za slučajni spremenljivki X, Y velja

$$\boxed{E\left((X - f(Y))^2\right) \geq E\left((X - E(X|Y))^2\right)}$$

enačaj $\Leftrightarrow f(Y) = E(X|Y)$

zgled: $(X,Y) \sim N(\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$

Kaj je $E(X|Y)$?

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

$$= \frac{\sigma_y\sqrt{2\pi}}{\sigma_x\sigma_y\, 2\pi\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \cancel{\left(\frac{y-\mu_y}{\sigma_y}\right)^2}\right) + \underbrace{\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2 \frac{1-\rho^2}{1-\rho^2}}}$$

$$\frac{1}{2(1-\rho^2)}\cancel{\left(\frac{y-\mu_y}{\sigma_y}\right)^2} - \frac{\rho^2}{2(1-\rho^2)}\left(\frac{y-\mu_y}{\sigma_y}\right)^2$$

$$= \frac{1}{\sigma_y\cdot\sqrt{2\pi}\sqrt{1-\rho^2}}\cdot e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_x}{\sigma_x} - \rho\frac{y-\mu_y}{\sigma_y}\right)^2}$$

$$\underbrace{\left(\frac{x-\mu_x}{\sigma_x} - \rho\frac{y-\mu_y}{\sigma_y}\right)}_{\frac{x-\left(\mu_x + \frac{\rho\sigma_x}{\sigma_y}(y-\mu_y)\right)}{\sigma_x}}$$

$$\sim N\left(\mu_x + \frac{\rho\sigma_x}{\sigma_y}(y-\mu_y),\ \sigma_x\sqrt{1-\rho^2}\right)$$

$$E(X \mid Y=y) = \mu_x + \frac{\rho\cdot\sigma_x}{\sigma_y}(y-\mu_y)$$

$$E(X \mid Y) = \mu_x + \frac{\rho\cdot\sigma_x}{\sigma_y}(Y-\mu_y)$$

$$= \underbrace{\frac{\rho\sigma_x}{\sigma_y}}_{a} Y + \underbrace{\left(\mu_x - \frac{\rho\sigma_x}{\sigma_y}\mu_y\right)}_{b} = aY + b$$

višji momenti

X slučajna spremenljivka $a \in \mathbb{R}$, $k \in \mathbb{N}$

$m_k(a) = E((X-a)^k)$

$\downarrow$

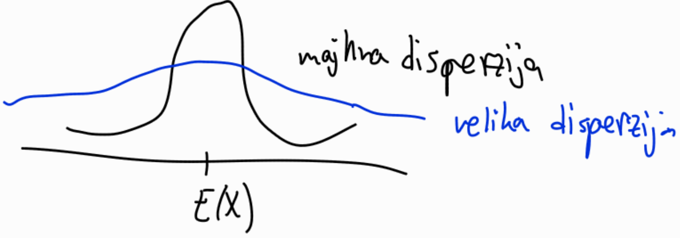
k-ti moment glede na a

Če je $a=0$, potem $m_k(0)$ imenujemo k-ti <u>začetni moment</u>.

Če je $a=E(X)$, potem $m_k(E(X))$ imenujemo k-ti <u>centralni moment</u>.

$m_1(0) = E(X)$

$m_2(E(X)) = Var(X)$



<u>trditev:</u> Če obstaja $m_n(a)$, potem obstaja $m_k(b)$ $\forall b$ in $\forall\ 1 \leq k \leq n$

<u>oznake:</u> $z_k = m_k(0)$

$m_k = m_k(E(X))$

<u>asimetrija</u> slučajne spremenljivke X je $A(X) = \dfrac{m_3}{m_2^{\frac{3}{2}}}$

$A(\lambda X) = A(X)$ za vsak $\lambda \in \mathbb{R} \setminus \{0\}$

<u>sploščenost</u> (kurtozis) slučajne spremenljivke X je $K(X) = \dfrac{m_4}{m_2^2}$

<u>zgled:</u> $X \sim N(\mu, \sigma)$, $A(X) = 0$, $K(X) = 3$

vrstilne karakteristike

<u>Mediana</u> slučajne spremenljivke X je vsaka vrednost $x \in \mathbb{R}$, za katero velja

$$\mathbb{P}(X \geq x) \geq \tfrac{1}{2} \quad \text{in} \quad \mathbb{P}(X \leq x) \geq \tfrac{1}{2}$$

$$\mathbb{P}(X \geq x) = 1 - \mathbb{P}(X < x) = 1 - \underbrace{F_X(x-)}_{\text{leva limita } F \text{ v točki } x}$$

torej je mediana vsako število x, za katero je

$$F_X(x-) \leq \tfrac{1}{2} \leq F_X(x)$$

Zgled: splošno za $0 < p < 1$ definiramo kvantil reda p:

x_p je vsako število, za katero velja

$$F_X(x_p-) \leq p \leq F_X(x_p)$$, oziroma

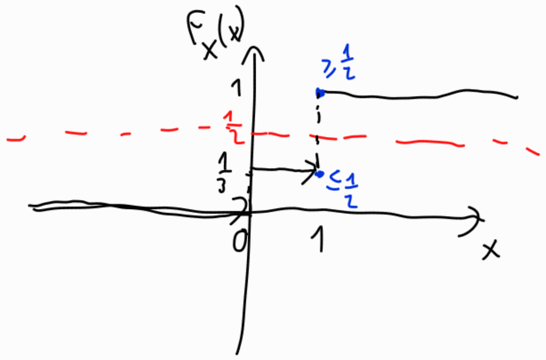
$$P(X \leq x_p) \geq p \quad \text{in} \quad P(X \geq x_p) \geq 1-p$$

$x_{\frac{1}{2}}$ je mediana

Nova predavanja – 2. semester

Zgled:

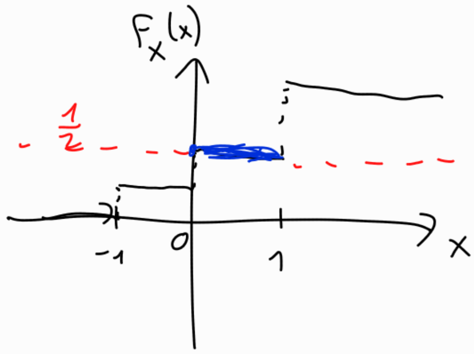
1. $X : \begin{pmatrix} 0 & 1 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$



$x_{\frac{1}{2}} = 1$

$E(X) = \frac{2}{3}$

2. $X : \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$

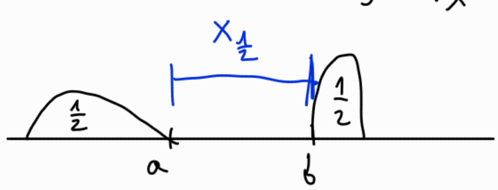


$\forall x_m \in [0,1)$ je mediana

3. $N(\mu, \sigma)$

$x_{\frac{1}{2}} = \mu$

4. graf p_X



Kvartili $x_{\frac{1}{4}}, x_{\frac{3}{4}}$

kvintili $x_{\frac{1}{5}}, x_{\frac{2}{5}}, \ldots$

percentili: $x_{1\%}, x_{20\%}, \ldots$

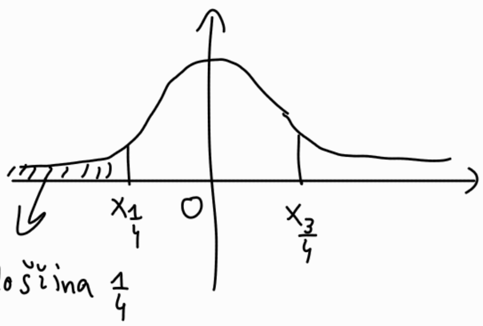
(semiinter) kvartilni razmik

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

npr. če X nima $E(|X|^2)$

zgled: 1. $X \sim N(0,1)$, $s = ?$

$\sigma = 1$



$$x_{\frac{1}{4}} = -x_{\frac{3}{4}}$$

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

$$s \doteq 0{,}67\ldots$$

2. Cauchyjeva porazdelitev

$$p_X(x) = \frac{1}{\pi(1+x^2)}$$

$$x_{\frac{1}{4}} = -x_{\frac{3}{4}}$$

$$s = x_{\frac{3}{4}} = -x_{\frac{1}{4}}$$

$$\frac{1}{\pi}\int_{-\infty}^{x_{\frac{1}{4}}} \frac{1}{1+x^2}\,dx = \frac{1}{4}$$

$$\frac{1}{\pi}\cdot \arctan(x)\Big|_{-\infty}^{x_{\frac{1}{4}}} = \frac{1}{4}$$

$$\frac{1}{\pi}\left(\arctan\left(x_{\frac{1}{4}}\right) + \frac{\pi}{2}\right) = \frac{1}{4}$$

$$\arctan\left(x_{\frac{1}{4}}\right) = \left(\frac{1}{4} - \frac{1}{2}\right)\pi$$

$$\arctan\left(x_{\frac{1}{4}}\right) = -\frac{\pi}{4}$$

$$x_{\frac{1}{4}} = \tan\left(-\frac{\pi}{4}\right)$$

$$x_{\frac{1}{4}} = -1$$

$$\boxed{s = 1}$$

Rodovne funkcije

X diskretna in zavzame vrednosti v $\mathbb{N}_0$

$p_k = \mathbb{P}(X=k)$

$$G_X(s) = \sum_{k=0}^{\infty} p_k \cdot s^k$$

rodovna funkcija slučajne spremenljivke X

$G_X(0) = p_0$

$G_X(1) = \sum_{k=0}^{\infty} p_k = 1$

G_X konvergira v točki $s=1$

$\Rightarrow$ konvergenčni radij $R \geq 1$

$\Rightarrow G_X$ je definirana vsaj na $[-1,1]$

Velja $G_X(s) = \underbrace{\sum_{k=0}^{\infty} p_k \quad s^k}_{E(s^X)} = E(s^X)$

zgledi: 1) $X \sim Geo(p)$

$p_k = \mathbb{P}(X=k) = p \cdot q^{k-1}$; $k \geq 1$

$$G_X(s) = \sum_{k=1}^{\infty} p \cdot q^{k-1} \cdot s^k = p \cdot s \sum_{k=1}^{\infty} q^{k-1} s^{k-1} \underset{k'=k-1}{=}$$

$$= ps \cdot \sum_{k'=0}^{\infty} q^{k'} s^{k'} = p \cdot s \cdot \frac{1}{1-qs} = \frac{ps}{1-qs} = \frac{p \cdot s}{1-(1-p)s}$$

2) $X \sim Poi(\lambda)$

$p_k = \frac{\lambda^k}{k!} e^{-\lambda}$ $k = 0,1,2,\ldots$

$$G_X(s) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \cdot s^k = e^{-\lambda} \cdot \overbrace{\sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!}}^{\text{Taylor za eksponentno}} = e^{-\lambda} \cdot e^{\lambda s} = e^{\lambda(s-1)}$$

izrek: X in Y porazdeljeni po $\mathbb{N}_0$

Potem velja $P(X=k) = \mathbb{P}(Y=k) \ \forall k \quad \Leftrightarrow \quad G_X(s) = G_Y(s)$ za $\forall s \in [-1,1]$

oziroma katerikoli interval

$$G_X(s) = \sum_{k=0}^{\infty} \underbrace{\frac{G_X^{(k)}(0)}{k!}}_{p_k} s^k$$

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

izrek: Naj bo $G_X(s)$ rodovna funkcija slučajne spremenljivke X.

Potem je

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

$$\| $$

$$\lim_{s \uparrow 1} G_X^{(n)}(1)$$

dokaz: $G_X(s) = \sum_{k=0}^{\infty} p_k s^k$

$$G_X^{(n)}(s) = \sum_{k=0}^{\infty} p_k \cdot k \cdot (k-1) \cdot \ldots \cdot (k-n+1) \cdot s^{k-n}$$

$$\lim_{s \uparrow 1} G_X^{(n)}(s) = \sum_{k=0}^{\infty} p_k \cdot k \cdot (k-1) \cdot \ldots \cdot (k-n+1)$$

$$= E(X \cdot (X-1) \cdot \ldots \cdot (X-n+1))$$

posledica: $E(X) = G_X'(1-)$

$$D(X) = \underbrace{G_X''(1-) + G_X'(1-)}_{E(X^2)} - \underbrace{(G_X'(1-))^2}_{E(X)^2}$$

dokaz: $E(X) = G_X'(1-)$ $n=1$

$$E(X\cdot(x-1)) = G_X''(1-)$$
$$\overset{\shortparallel}{E(X^2)-E(X)}$$

$$D(X) = E(X^2) - E(X)^2$$
$$= G_X''(1-) + E(X) - E(X)^2$$
$$= G_X''(1-) + G_X'(1-) - (G_X'(1-))^2$$

zgled: $X \sim Geo(p)$

$$G_X(s) = \frac{ps}{1-qs}$$

$$G_X'(s) = \frac{p\cdot(1-qs) - ps\cdot(-q)}{(1-qs)^2} = \frac{p - pqs + pqs}{(1-qs)^2} = \frac{p}{(1-qs)^2}$$

$$G_X'(1-) = \frac{p}{(1-q)^2} = \frac{p}{(1-1+p)^2} = \boxed{\frac{1}{p} = E(X)}$$

izrek: G_X rodovna funkcija za X in G_Y rodovna funkcija za Y

če sta X in Y neodvisni, potem je

$$G_{X+Y}(s) = G_X(s)\cdot G_Y(s)$$

dokaz:

$$G_{X+Y}(s) = E(s^{X+Y}) = E(s^X\cdot s^Y) = E(s^X)\cdot E(s^Y) = G_X(s)\cdot G_Y(s)$$

posledica: $X_1, \dots, X_n$ neodvisne

Potem je $G_{X_1+\dots+X_n}(s) = G_{X_1}(s)\cdot\ldots\cdot G_{X_n}(s) = \prod_{i=1}^{n} G_{X_i}(s)$

Če so $X_1, \dots, X_n$ enako porazdeljene

$$G_{X_1+\dots+X_n} = (G_{X_1}(s))^n$$

izrek: Naj bodo slučajne spremenljivke $X_1, \dots, X_n, \dots$ $n \in \mathbb{N}$ neodvisne in enako porazdeljene. N neka druga slučajna spremenljivka neodvisna od vseh ostalih. Potem je rodovna funkcija $S = \sum_{i=1}^{N} X_i$

$$G_S(s) = G_N(G_X(s)), \quad G_X(s) = G_{X_n}(s) \quad \forall n \in \mathbb{N}$$

dokaz: $$\mathbb{P}(S=k) = \sum_{n=1}^{\infty} \mathbb{P}(S=k, N=n) = \sum_{n=1}^{\infty} \mathbb{P}(N=n, X_1+\dots+X_n=k) =$$

$$= \sum_{n=1}^{\infty} P(N=n) \cdot \mathbb{P}(X_1+\dots+X_n=k)$$

$$G_S(s) = \sum_{k=0}^{\infty} \mathbb{P}(S=k) \cdot s^k = \sum_{k=0}^{\infty} \left(\sum_{n=1}^{\infty} \mathbb{P}(N=n) \cdot \mathbb{P}(X_1+\dots+X_n)=k \right) \cdot s^k$$

zamenjamo vrstni red vsot $$= \sum_{n=1}^{\infty} \left(\sum_{k=0}^{\infty} \mathbb{P}(N=n) \cdot P(X_1+\dots+X_n)=k \right) \cdot s^k$$

$$= \sum_{n=1}^{\infty} \mathbb{P}(N=n) \cdot \sum_{k=0}^{\infty} \mathbb{P}(X_1+\dots+X_n=k) \cdot s^k$$

$$= \sum_{n=1}^{\infty} \mathbb{P}(N=n) \cdot G_{X_1+\dots+X_n}(s) =$$

$$= \sum_{n=1}^{\infty} \mathbb{P}(N=n) \cdot \left(G_X(s)\right)^n$$

enako porazdeljene in neodvisne

$$= G_N(G_X(s))$$

zgled: Kokoš znese N jajc, $N \sim Poi(\lambda)$

$X_n : \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$ $\qquad K \dots$ število piščancev

$$K = X_1 + X_2 + \dots + X_N$$

$G_K(s) = ?$ $\qquad G_N(s) = e^{\lambda(s-1)}$

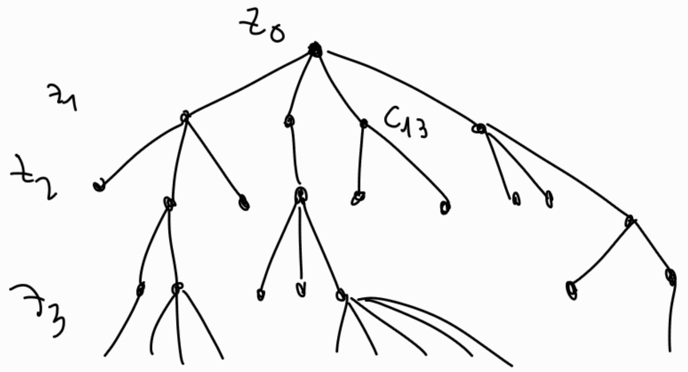
$G_X(s) = q + ps$

$$G_K(s) = G_N(G_X(s))$$
$$= G_N(q+ps)$$
$$= e^{\lambda(q+ps-1)}$$
$$= e^{\lambda(1-p+ps-1)} = \underbrace{e^{p\lambda(s-1)}}$$

to je rodovna funkcija Poissonove porazdelitve

$K \sim Poi(p\lambda)$

procesi razvejanja (branching processes)



Recimo imamo osebek, ki ima potomce. Zanima nas število potomcev v posamezni generaciji, ali bo izumrla, itd...

imamo z_i $i=1,\dots,n$

$Z\dots$ število potomcev, ki jih ima osebek

$z_1\dots$ število potomcev, ki jih ima 1. osebek

$z_2\dots$ število potomcev v 2. generaciji

$c_{ij}\dots$ število potomcev i-te generacije j-tega osebka

velja $c_{ij} \sim Z$

če gledamo Z_n

z_1
$z_2 = c_{11} + \dots + c_{1 z_1}$
$\vdots$

$$Z_n = C_{n-1,1} + \dots + C_{n-1,Z_{n-1}}$$

$$= \sum_{i=1}^{Z_{n-1}} C_{n-1,i}$$

označimo $G(s) = G_Z(s)$ po prejšnjem velja:

$$G_{Z_2}(s) = G_{Z_1}(G_Z(s)) = G(G(s))$$

trditev: $G_{Z_n}(s) = G_{Z_{n-1}}(G(s)) = \underbrace{G(G(\dots G}_{n\text{-krat}}(s)\dots)$

trditev: $E(Z_n) = E(Z)^n$

dokaz: $S = \sum_{i=1}^{N} X_i \rightsquigarrow E(S) = E(X)\,E(N)$

$$E(Z_2) = E(Z)\,E(Z)$$

$$E(Z_3) = E(Z)\cdot E(Z_2) = E(Z)^3$$

$$\vdots$$

$$E(Z_n) = E(Z)^n$$

dogodek $[Z_{n-1} = 0] \subseteq [Z_n = 0]$ ... dogodek izumrtja

dogodek izumrtja: $\overset{\text{extinction}}{E} = \bigcup_{n \geq 1} [Z_n = 0]$

$$\mathbb{P}(E) = \lim_{n\to\infty} \mathbb{P}(Z_n = 0) =: \varepsilon$$

izrek: ε verjetnost izumrtja je najmanjša pozitivna negibna točka rodovne funkcije G_Z.

dokaz:

↳ 1 je vedno negibna točka:

$$G(s) = \sum_{i=0}^{\infty} p_i s^i \quad (\text{vsota } p_i \text{ je } 1)$$

↳ označimo $\varepsilon_n = P(Z_n = 0) = G_{Z_n}(0)$ $\quad (\varepsilon = \lim_{n\to\infty} \varepsilon_n)$

vemo: $G_{Z_n}(s) = G_{Z_{n-1}}(G(s)) \underset{\downarrow}{=} G(G_{Z_{n-1}}(s))$

samo ponovimo n-krat G

vemo, da je G zvezna, odvedljiva, naraščajoča na $[0,1]$

↳ za naraščajoča pogledamo odvod in vidimo, da je pozitiven

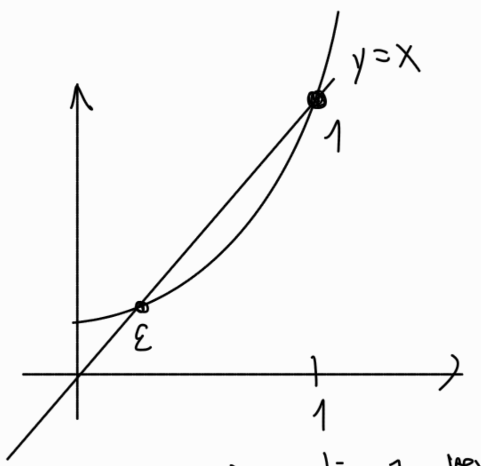
$$\begin{array}{ccc} \lim G(G_{Z_{n-1}}(0)) & = & \lim_{n\to\infty} G_{Z_n}(0) \\ \| & & \| \\ G(\varepsilon) & = & \varepsilon \end{array}$$

$\Rightarrow$ ε negibna točka $\in [0,1]$

↳ očitno je ε_n naraščajoče zaporedje, ker so dogodki vsebovani drug v drugemu

vidimo, da $\varepsilon_n \leq 0$ in $\varepsilon_0 = 0$

$\Rightarrow$ ker G zvezna, naraščajoča, je ε pozitivna negibna točka in je tudi najmanjša, ker začnemo od 0 in imamo naraščajočo funkcijo.



edini 2 možni negibni točki

izumrtje z verjetnostjo ε

izrek: $\varepsilon = 1 \iff E(Z) \leq 1$

Momentno-rodovna funkcija $M_X(t) = E(e^{tX})$

X slučajna spremenljivka

def: pri navadni: $G_X(s) = E(s^X)$

↳ za nenegativne celoštevilsko porazdeljene slučajne spremenljivke velja:

$$M_X(t) = E((e^t)^X) = G_X(e^t)$$

↳ če pa X zvezno porazdeljena:

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} p_X(x)\, dx \quad \ldots \text{ Laplaceova transformacija } p_X$$

zgled: $X \sim N(0,1)$ $\quad M_X(t) = ?$

$$M_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{x^2}{2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2 - 2tx + t^2) + \frac{t^2}{2}} dx =$$

$$= \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{t^2}{2}} \cdot \underbrace{\int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}} dx}_{N(t,1) \to 1} = e^{\frac{t^2}{2}}$$

izrek: če $M_X(t)$ obstaja za $t \in (-\delta, \delta)$ $\delta > 0$, potem vsi momenti

$z_k := E(X^k)$ obstajajo za $\forall k \in \mathbb{N}_0$ in velja

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

torej $M_X^{(k)}(0) = z_k \quad \forall k$

k-ti odvod

dokaz: $M_X(t)$, $t \neq 0 \in (-\delta, \delta)$

$$M_X(t) = E(e^{tX})$$
$$= E\left(\sum_{k=0}^{\infty} \frac{t^k X^k}{k!}\right)$$
$$\overset{\text{Fubinijev izrek}}{=} \sum_{k=0}^{\infty} \frac{t^k}{k!} \cdot E(X^k)$$

trditev: $M_{aX+b}(t) = e^{tb} \cdot M_X(at)$

dokaz:
$$M_{aX+b}(t) = E\left(e^{t(aX+b)}\right)$$
$$= E\left(e^{atX} \cdot e^{tb}\right)$$
$$= e^{tb} \cdot E\left(e^{atX}\right) = e^{tb} \cdot M_X(at)$$

izrek: Če sta X in Y neodvisni slučajni spremenljivki, potem je

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

dokaz:
$$M_{X+Y}(t) = E(e^{tX} \cdot e^{tY}) = M_X(t) \cdot M_Y(t)$$

zgled: $X \sim N(a, \sigma_1) \Rightarrow X = \sigma_1 Z + a$ če $Z \sim N(0,1)$

$Y = N(b, \sigma_2) \Rightarrow Y = \sigma_2 Z + b$

$$M_X(t) = e^{ta} \cdot M_Z(\sigma_1 t) = e^{ta} \cdot e^{\frac{(\sigma_1 t)^2}{2}} = e^{\frac{1}{2}(\sigma_1^2 t^2 + 2at)}$$

$$M_Y(t) = e^{\frac{1}{2}(\sigma_2 t^2 + 2bt)}$$

$$M_{X+Y} = e^{\frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2 + 2(a+b)t)}$$

↳ ker $M_X(t)$ enolično določa porazdelitev X, kadar konvergira na $(-\delta, \delta)$ (ker je Taylorjeva vrsta enolična) je

$$X+Y \sim N(a+b, \sqrt{\sigma_1^2+\sigma_2^2}^{\,2})$$

zgled: Poišči vse začetne momente za X, ki je porazdeljena N(0,1)

$$M_X(t) = e^{\frac{t^2}{2}} = \sum_{k=0}^{\infty} \frac{t^{2k}}{k!\,2^k} = \sum_{k=0}^{\infty} \frac{z_{2k}}{2k!} t^{2k}, \text{ kjer } z_{2k} = \frac{(2k)!}{k!\,2^k} \Rightarrow z_{2k+1} = 0$$

$$z_{2k} = \frac{(2k)!}{2\cdot 4 \cdot \ldots \cdot 2k} = \frac{(2k)!}{(2k)!!} = 1\cdot 3 \cdot \ldots \cdot (2k-1) = (2k-1)!!$$

Šibki in krepki zakon velikih števil

ŠZVŠ KZVŠ

def: imamo neko zaporedje $\{X_n\}_{n=1}^{\infty}$ slučajnih spremenljivk konvergira verjetnostno proti X, če obstaja neka slučajna spr. X, da je

$$\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1 \quad \text{za } \forall \varepsilon > 0$$

oznaka: $X_n \xrightarrow{P} X$

oziroma $\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0$

def: Zaporedje slučajnih spremenljivk $\{X_n\}_{n=1}^{\infty}$ konvergira skoraj gotovo, če

$$\mathbb{P}\left[\lim_{n\to\infty} X_n = X\right] = 1$$

oznaka: $X_n \xrightarrow{s.g.} X$

↳ dogodek $[\lim_{n\to\infty} X_n = X]$

$X: \Omega \to \mathbb{R}$ ⇝ $= [\omega \in \Omega, \lim_{n\to\infty} X_n(\omega) = X(\omega)]$

↳ torej za vsak izzid gledamo, da je limita vrednosti X_n enaka tisti pri X.

Nova predavanja

↳ dogodek $[\omega \in \Omega, \lim_{n\to\infty} X_n(\omega) = X(\omega)$ lahko drugače zapišemo:

$[\omega \in \Omega, \forall k \in \mathbb{N} \ \exists m \in \mathbb{N} \ \forall n \geq m : |X_n(\omega) - X(\omega)| \leq \frac{1}{k}|$

$= \bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \{\omega \in \Omega ; |X_n(\omega) - X(\omega)| < \frac{1}{k}\}$

trditev: Če $X_n \xrightarrow{s.g.} X$ potem $\forall \varepsilon > 0$ velja:

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

dokaz: $C_m := \{\omega \in \Omega ; |X_n(\omega) - X(\omega)| < \varepsilon \quad \forall n \geq m\}$

$= \bigcap_{n=m}^{\infty} \{\omega \in \Omega ; |X_n(\omega) - X(\omega)| < \varepsilon\}$

vidimo: $C_1 \subseteq C_2 \subseteq C_3 \subseteq \ldots$

$\Rightarrow [\omega, \lim_{n\to\infty} X_n(\omega) = X(\omega)] \subseteq \bigcup_{m=1}^{\infty} C_m$

↓

Ker $X_n \xrightarrow{s.g.} X$ ima to verjetnost 1 $\Rightarrow 1 \leq P\left(\bigcup_{m=1}^{\infty} C_m\right)$

$$\Rightarrow P\left(\bigcup_{m=1}^{\infty} C_m\right) = 1 \quad \Rightarrow \lim_{n\to\infty} P(C_n) = 1$$

izrek: Če $X_n \xrightarrow{s.g.} X$, potem $X_n \xrightarrow{P} X$

dokaz: $P(|X_n - X| < \varepsilon \;\; \forall n \geq m) \leq P(|X_m - X| < \varepsilon) \leq 1$

po prejšnji trditvi $= 1$

$$\Rightarrow \quad 1 \leq P(|X_m - X| < \varepsilon) \leq 1$$

$$\Rightarrow \quad P(|X_m - X| < \varepsilon) \xrightarrow{m\to\infty} 1$$

opomba: Obratno ne velja

def: X_n je zaporedje slučajnih spremenljivk, za katere obstaja upanje.

definiramo: $S_n := \sum_{k=1}^{n} X_k$, $\quad Y_n := \frac{1}{n}(S_n - E(S_n)) = \frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i))$

za $\{X_n\}_{n=1}^{\infty}$ velja šibki zakon velikih števil, če $Y_n \xrightarrow{P} 0$,

oziroma $\lim_{n\to\infty} P(|Y_n| < \varepsilon) = 1 \quad \forall \varepsilon > 0$

reformulacija:

$$E(Y_n) = \frac{1}{n}\left(E(S_n) - E(E(S_n))\right) = 0$$

$$\lim_{n\to\infty} P\left(\frac{|S_n - E(S_n)|}{n} < \varepsilon\right) = 1 \text{ ali } \lim_{n\to\infty} P\left(\frac{1}{n}\left|\sum_{i=1}^{n} X_i - E(X_i)\right| < \varepsilon\right) = 1$$

def: $\{X_n\}_{n=1}^{\infty}$ slučajne spremenljivke, vse imajo upanje. Potem za $\{X_n\}_{n=1}^{\infty}$ velja **krepki zakon velikih števil**, če $Y_n \xrightarrow{s.g.} 0$

tj: $P(\lim_{n\to\infty} Y_n = 0) = 1$ oz.

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

posledica: Če velja KZVŠ $\Rightarrow$ ŠZVŠ

zgled: X_n met kocke: $X_n \sim U_d(6)$

$$Y_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - E(X_i))$$

$$= \frac{1}{n} \sum_{i=1}^{n} X_i - n \cdot E(X_i)$$

$$= \frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{n} \cdot n \cdot E(X_1)$$

$$= \frac{1}{n} \sum_{i=1}^{n} X_i - \frac{7}{2}$$

$$E(X_1) = \sum_{i=1}^{6} i \cdot p_i = \sum_{i=1}^{6} i \cdot \frac{1}{6} = \frac{1}{6} \sum_{i=1}^{6} i = \frac{7 \cdot \cancel{6}}{2 \cdot \cancel{6}} = \frac{7}{2}$$

$$\rightsquigarrow Y_n + \frac{7}{2} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

KZVŠ potem pomeni:

$$P\left(\frac{1}{n} \sum_{i=1}^{n} X_i \to \frac{7}{2}\right) = 1$$

ŠZVŠ:

$$\lim_{n\to\infty} P(|Y_n| < \varepsilon) = 1$$

$$\lim_{n\to\infty} P\left(\left|\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) - \frac{7}{2}\right| < \varepsilon\right) = 1$$

izrek (Neenakost Markova):

X slučajna spremenljivka, ki ima upanje. Potem za $\forall a > 0$ velja:

$$P(|X| \geq a) \leq \frac{E(|X|)}{a}$$

dokaz: (za zvezne slučajne spremenljivke)

$$E(|X|) = \int_{-\infty}^{\infty} |x| \cdot p_X(x)\, dx \geq \int_{|x| \geq a} |x|\, p_X(x)\, dx \geq \int_{|x| \geq a} a\, p_X(x)\, dx = a \cdot P(|X| \geq a)$$

zmanjšamo območje integracije

izrek (Neenakost Čebiševa):

X sluč. spr. za katero obstaja varianca. Potem velja:

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

dokaz: Uporabimo neenakost Markova za $(X - E(X))^2$

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

izrek (izrek Markova):

$\{X_n\}_{n=1}^{\infty}$ zaporedje sluč. spr., ki imajo varianco.

če velja $\lim_{n\to\infty} \frac{D(S_n)}{n^2} = 0$, potem velja ŠZVŠ za X_n

dokaz: Želimo $\lim_{n\to\infty} P(|Y_n| > \varepsilon) = 0$

$$Y_n = \frac{S_n - E(S_n)}{n}, \quad E(Y) = 0 \qquad D\left(\frac{S_n}{n}\right)$$

prištevanje konstante ne spremeni disperzije

$$P(|Y_n - \underset{0}{E(Y_n)}| > \varepsilon) \overset{\text{Čebišev}}{\leq} \frac{D(Y_n)}{\varepsilon^2} = \frac{D(S_n)}{n^2 \varepsilon^2} \xrightarrow{n\to\infty} 0$$

predpostavka

posledica (izrek Čebiševa):

$\{X_n\}_{n=1}^{\infty}$ zap. sluč. spr., ki imajo varianco. Za njih velja $\sup_{n\in\mathbb{N}} D(X_n) < \infty$

disperzije navzgor omejene

in so nekorelirane. Potem velja ŠZVŠ.

dokaz: $D(S_n) = \sum_{i=1}^{n} D(X_i)$

nekorelirane

označimo $M := \sup_{n\in\mathbb{N}} D(X_n)$

potem je $\frac{D(S_n)}{n^2} = \frac{\sum_{i=1}^{n} D(X_i)}{n^2} \leq \frac{nM}{n^2} = \frac{M}{n} \xrightarrow{n\to\infty} 0$

izrek Markova

$\Rightarrow$ velja ŠZVŠ

zgled: A dogodek v Bernoullijevem zaporedju

$X_n = \begin{pmatrix} 0 & 1 \\ p & q \end{pmatrix}$ $\quad p+q=1 \quad 0<p,q<1$

ponovitev poskusa

X_n so neodvisne, imajo upanje in disperzijo $D(X_n) = pq$

$\Rightarrow$ velja ŠZVŠ

$P\left(\left|\frac{S_n - nq}{n}\right| < \varepsilon\right) \xrightarrow{n\to\infty} 1$

$P\left(\left|\frac{1}{n}\left(\sum_{i=1}^{n} X_i\right) - q\right| < \varepsilon\right) \xrightarrow{n\to\infty} 1$

torej povprečna vrednost konvergira proti q

izrek (izrek Kolmogorova):

$\{X_n\}_{n=1}^{\infty}$ zaporedje neodvisnih slučajnih spremenljivk, za katere je

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

Brez dokaza

zgled: V Bernoullijevem zaporedju velja KZVŠ.

opomba: Če so vsi X_n enako porazdeljeni in imajo upanje, potem velja KZVŠ, če $D(X_n) < \infty$.

Centralni limitni izrek

normaliziramo

def: $\{X_n\}_{n=1}^{\infty}$ zap. sluč. spr., ki imajo varianco. Označimo $Z_n := \frac{S_n - E(S_n)}{\sigma^2(S_n)}$

$\sigma^2(S_n) = D(S_n)$

$E(Z_n) = 0 \quad , D(Z_n) = 1$

izrek (centralni limitni izrek):

X_n so neodvisne in enako porazdeljene sluč. spr., ki imajo varianco.
Potem porazdelitvene funkcije $F_{Z_n}(x) = P(Z_n \leq x)$ konvergirajo k

$$\lim_{n\to\infty} F_{Z_n}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

predpostavka $E(X_i^2) < \infty$

oziroma $Z_n \to Z \sim N(0,1)$

$F_{Z_n} \to F_Z(x) \quad \forall x \in \mathbb{R}$

dokaz: Če privzamemo, da imajo X_n momentno rodovno funkcijo na $(-\delta, \delta)$, $\delta > 0$

$$M_{Z_n}(t) = E(e^{tZ_n})$$

želimo pokazati, da $\lim_{n\to\infty} M_{Z_n}(t) = e^{\frac{t^2}{2}} = M_{N(0,1)}(t)$

$M_X(t)$ obstaja na $(-\delta, \delta)$ za $\delta > 0$.

X enako porazdelitev kot vsi X_i

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

$E(X) = \mu$

$\sigma(X) = \sigma$

$U_i := X_i - \mu$

$E(U_i) = 0$

$\sigma(U_i) = \sigma$

$M_{U_i}(t) = e^{-\mu t} M_{X_i}(t)$

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

$D(S_n) = n \cdot \sigma^2$

$\sigma(S_n) = \sqrt{n} \cdot \sigma$

$$Z_n = \frac{S_n - E(S_n)}{\sigma(S_n)} = \frac{\sum_{i=1}^{n}(X_i - E(X_i))}{\sigma(S_n)} = \frac{\sum_{i=1}^{n} U_i}{\sigma(S_n)} = \frac{\sum_{i=1}^{n} U_i}{\sigma\sqrt{n}}$$

$$\Rightarrow M_{Z_n}(t) = E(e^{Z_n t}) = E\left(e^{\left(\sum_{i=1}^{n} U_i\right)\cdot \frac{t}{\sigma\sqrt{n}}}\right) = M_U\left(\frac{t}{\sigma\sqrt{n}}\right)^n$$

$$= M_U(t) = 1 + 0\cdot t + \frac{\sigma^2}{2!}t^2 + O(t^3)$$

$$= \left(1 + \frac{\sigma^2}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + O(t^3)\right)^n =$$

$$= \left(1 + \frac{1}{2n}t^2 + O(t^3)\right)^n$$

$$= e^{\frac{t^2}{2}}$$

$$\left(1+\frac{1}{n}\right)^n \to e$$

$$\left(1+\frac{a}{n}\right)^n = \left(\left(1+\frac{a}{n}\right)^{\frac{n}{a}}\right)^a \to e^a$$

Fourierjeva transformacija

p_X gostota

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx}\, p_X(x)\, dx$$

karakteristična funkcija slučajne spremenljivke

za splošno (ni nujno zvezna):

$$\varphi_X(t) = E(e^{itX})$$

V splošnem CLI dokažemo s pomočjo konvergence karakterističnih funkcij Z_n k karakteristični funkciji $N(0,1)$.

opomba: $\{X_i\}_{i=1}^{\infty}$ neodvisne enako porazdeljene $E(X_i^2) < \infty$

$$Z_n \xrightarrow{Z} Z \sim N(0,1)$$

$$Z_n = \frac{S_n - E(S_n)}{\sigma(S_n)} = \frac{S_n - n\cdot\mu \;/:n}{\sigma\cdot\sqrt{n} \;\;/:n} = \frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} \Rightarrow \frac{S_n}{n} = Z_n \cdot \frac{\sigma}{\sqrt{n}} + \mu$$

$\frac{S_n}{n} \to ?$

povprečje X_i

$$P(Z_n \leq x) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \sim N(0,1)$$

$$P\left(\frac{S_n}{n} \leq \frac{\sigma}{\sqrt{n}} x + \mu\right) \underset{\text{za velike } n}{\approx} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

primer:

Laplaceova formula je posledica CLI:

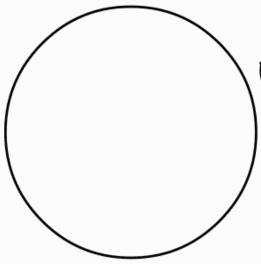
$$P(\alpha \leq S_n \leq \beta) = \phi\left(\frac{\alpha - \overset{n\cdot\mu}{np}}{\underset{\sqrt{n}\cdot\sigma}{\sqrt{npq}}} \leq Z_n \leq \frac{\beta - np}{\sqrt{npq}}\right) \underset{\text{za velike } n}{\approx} \phi\left(\frac{\beta - np}{\sqrt{npq}}\right) - \phi\left(\frac{\alpha - np}{\sqrt{npq}}\right)$$

$$S_n := X_1 + \ldots + X_n$$

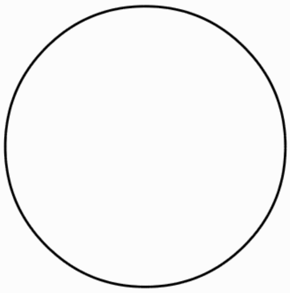
$$X_i \sim \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$$

STATISTIKA

Uvod v matematično statistiko:



populacija
- izobrazba
- pojav bolezni



tehnika
- pogostost napak

Ne analiziramo celotne populacije, pač pa samo nek vzorec, ki pa mora biti reprezentativen (izbran nepristransko)

$X_1, X_2, \dots, X_n$ izbira enega predstavnika in meritev opazovane količine predstavlja 1 slučajno spremenljivko

X_i so enako porazdeljene in neodvisne. $n \ll$ velikost populacije

$E(X_i^2) < \infty$

Vrednost x_i, ki nam jo da i-ta izbira predstavnika, imenujemo realizacija slučajne spremenljivke. Nabor vseh realizacij imenujemo vzorec.

$(x_1, x_2, \dots, x_n)$

Želimo pridobiti informacijo o populaciji iz vzorca.

Želimo vedeti povprečno vrednost $E(X)$

standardni odklon $\sigma(X)$

Povprečna vrednost / srednja vrednost $\overline{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ vzorčno povprečje vzorčni modus : najbolj pogosta vrednost

vzorčna mediana : vzorec uredimo in vzamemo srednjega $X_{\frac{n}{2}}$

To so ocene za pričakovano vrednost $E(X)$

ocene za standardni odklon :

s_0, kjer je $\cancel{s_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2}$ vzorčni odklon

s popravljeni vzorčni odklon

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

kao 1 podatek izgubimo

vzorčne statistike in cenilke

$(X_1, X_2, \ldots, X_n)$ slučajni vektor X_i neodvisne, enako porazdeljene

vzorčna statistika je simetrična funkcija slučajnega vektorja

$$Y_n = g(X_1, \ldots, X_n) \qquad \text{npr: } \overline{X} = g(X_1, \ldots, X_n) = \frac{1}{n}(X_1 + \ldots + X_n)$$

Velikokrat z Y ocenjujemo vrednost nekega parametra porazdelitve.

Takrat Y imenujemo cenilka parametra ξ. $\rightarrow$ mogoče zeta namesto ksi

cenilka Y je nepristranska cenilka za parameter ξ, če je $E(Y) = \xi$

cenilka Y je dosledna cenilka za parameter ξ, če $Y_n \xrightarrow{n\to\infty} \xi$ verjetnostno :

$$P(|Y-\xi| < \varepsilon) \xrightarrow{n\to\infty} 1$$

trditev: $\overline{X} = \frac{1}{n}(X_1 + \dots + X_n)$ je nepristranska cenilka za $\mu := E(X)$

dokaz: $E(\overline{X}) = \frac{1}{n} E(X_1 + \dots + X_n) = \frac{1}{n}\left(E(X_1) + \dots + E(X_n)\right) = \frac{1}{n} \cdot n \cdot E(X) = \mu$

trditev: $\overline{X}$ je dosledna cenilka za μ

dokaz:
$$D(\overline{X}) = D\left(\frac{1}{n} \cdot \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \cdot \sum_{i=1}^{n} D(X_i) = \frac{n}{n^2} \cdot D(X) = \frac{1}{n} D(X) = \frac{\sigma^2}{n} \to 0$$

$$\overset{\text{Čebišev}}{\Rightarrow} \quad X_n \xrightarrow{P} \mu \quad \square$$

def: $X_1, \dots, X_n \qquad \mu, \sigma^2$

$x_1, \dots, x_n$ $\quad$ standardna napaka vzorca

$$SE(x) = \sigma(X) = \sigma$$

$\downarrow$ standardni odklon

standardna napaka cenilke (statistike) Y je

$$SE(Y) = \sigma(Y)$$

zgled: $SE(\overline{X}) = \frac{\sigma}{\sqrt{n}}$

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

$\overline{X} \sim ?$ $\qquad \overline{X} \sim (\mu, \frac{\sigma}{\sqrt{n}})$

$Z \sim N(0,1)$ $\qquad$ neodvisne $Z_1, Z_2, \dots, Z_n \sim N(0,1)$

$Z^2 \sim \chi^2(1)$ $\qquad Z_1^2 + \dots + Z_n^2 \sim \chi^2(n)$

$T = \frac{X}{Z} \sim$ Student $T(n, \beta)$ $\qquad \frac{\text{normalna}}{\chi^2}$

trditev: Če je Y_n cenilka za ξ in velja $\underset{\uparrow}{E(Y_n)} \xrightarrow{n\to\infty} \xi$ in $D(Y_n) \xrightarrow{n\to\infty} 0$,

asimptotično nepristranska

potem je Y dosledna cenilka za ξ

dokaz: $Y_n \xrightarrow{P} \xi$

$$\mathbb{P}(|Y_n - \xi| > 0) \xrightarrow{n\to\infty} 0$$

fiksirajmo $\varepsilon > 0$:

Potem $\exists n_0 : \forall n > n_0$ velja $|E(Y_n) - \xi| < \frac{\varepsilon}{2}$

ker je $[\omega : |Y_n(\omega) - \xi| \geq \varepsilon] \subseteq [|Y_n - \xi| \geq \varepsilon] \subseteq$

$\pm E(Y_n)$

* stranski račun

$$\varepsilon \leq |Y_n - \xi| = |Y_n - E(Y_n) + E(Y_n) - \xi| \overset{\Delta}{\leq} |Y_n - E(Y_n)| + |E(Y_n) - \xi|$$

$$\subseteq [|Y_n - E(Y_n)| + \underbrace{|E(Y_n) - \xi|}_{\leq \frac{\varepsilon}{2}} \geq \varepsilon] \subseteq [|Y_n - E(Y_n)| \geq \tfrac{\varepsilon}{2}]$$

$$\Rightarrow |Y_n - E(Y_n)| \geq \frac{\varepsilon}{2}$$

po 2. predpostavki

$$\Rightarrow P(|Y_n - \xi| \geq \varepsilon) \leq \mathbb{P}(|Y_n - E(Y_n)| \geq \tfrac{\varepsilon}{2}) \overset{\text{Čebišev}}{\leq} \frac{D(Y_n)}{\frac{\varepsilon^2}{4}} = \frac{4D(Y_n)}{\varepsilon^2} \xrightarrow{n\to\infty} 0$$

$\Rightarrow Y_n$ dosledna

vzorčna disperzija: $X_1, \dots, X_n$ iščemo cenilko za σ^2

$x_1, \dots, x_n$

$$S_0^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})^2$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{X})^2$$ popravljena vzorčna disperzija

primeri: Predpostavimo $X \sim N(\mu, \sigma)$

$$\left(S^2 = \frac{n}{n-1} S_0^2 \right)$$

$$(\bar{X} - X_n) = \frac{1}{n} \sum_{i=1}^{n-1} X_i + \left(\frac{1}{n} - 1\right) X_n$$

$$D(\bar{X} - x) = \left((n-1) \cdot \frac{1}{n^2} + \left(\frac{1-n}{n}\right)^2 \right) \sigma^2$$

$$= \left(1 - \frac{1}{n^2}\right) \sigma^2$$

2?

$$\Rightarrow (X_i - \bar{X}) \sim N\left(0, \left(1 - \frac{1}{n^2}\right)^{\frac{1}{2}} \sigma\right)$$

Vpeljemo novo pomožno statistiko

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{(X_i - \bar{X})^2}{\sigma} \right) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \bar{X})^2$$

$z \sim N(0,1)$

izkaže se $\chi^2 \sim \chi^2(n-1)$

$$\chi^2 = z_1^2 + \dots + z_n^2 \qquad z_i \sim N(0,1)$$

$$\Rightarrow E(\chi^2) = n-1$$

trditev: S_0^2 je asimptotično nepristranska cenilka za σ^2

S^2 je nepristranska cenilka za σ^2

dokaz: $S_0^2 = \frac{n}{n-1} S^2$

$$\sigma^2 \chi^2 = (n-1) \cdot S^2$$

$$E(\sigma^2 \chi^2) = E((n-1) S^2)$$

$$\sigma^2 \cdot (n-1) = (n-1) \cdot E(S^2)$$

$$\Rightarrow E(S^2) = \sigma^2 \quad S^2 \text{ je nepristranska}$$

$$E(S_0^2) = \frac{n}{n-1} S^2 = \frac{n}{n-1} \sigma^2 \xrightarrow{n\to\infty} 1 \quad S_0^2 \text{ asimptotično nepristranska}$$

posledica:

S^2 in S_0^2 sta dosledni cenilki za σ^2

dokaz: $E(S^2) = \sigma^2 \to \sigma^2$

$$E(S_0^2) = \frac{n}{n-1} \sigma^2 \to \sigma^2$$

$$D(S^2) = D\left(\frac{\sigma^2}{n-1} \chi^2\right) = \frac{\sigma^4}{(n-1)^2} D(\chi^2) = \frac{\sigma^4}{(n-1)^2} 2\cancel{(n-1)} = \frac{2\sigma^4}{n-1} \xrightarrow{n\to\infty} 0$$

$$D(S_0^2) = \frac{\sigma^4}{n^2} D(\chi^2) = \frac{2(n-1)\sigma^4}{n^2} \xrightarrow{n\to\infty} 0 \quad \Rightarrow S^2 \text{ in } S_0^2 \text{ sta dosledni} \quad \square$$

Studentova t-porazdelitev z n prostostnimi stopnjami

$$T = \frac{Z}{S} \qquad Z \sim N(0,1), \quad S^2 \sim \chi^2(n)$$

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

$$\lim_{n\to\infty} p_T(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

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

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

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

$$\sigma^2 \sim S^2$$ T-statistika

$$T = \frac{\overline{X} - \mu}{S}\sqrt{n} \sim t(n-1)$$ Statistika T je porazdeljena Studentovo $t(n-1)$

Nova predavanja - metode za pridobivanje cenilk

1. metoda momentov

imamo vzorec $X_1, \dots, X_n$

def: k-ti vzorčni moment (začetni) je

$$Z_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

torej $Z_1 = \overline{X}$

ker je $E(Z_k) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^k\right) = \frac{1}{n} \cdot \sum_{i=1}^{n} E(X_i^k) = E(X^k)$

sledi, da Z_k nepristranska cenilka za $E(X^k)$

po izreku Čebiševa: z_k je dosledna cenilka za $E(X^k)$

Recimo, da imamo X zvezno poraz. z gostoto $p(x, \xi_1, \dots, \xi_e)$

vprašanje: ali se da izraziti te parametre?

$$z_k = E(X^k) \quad k\text{-ti začetni moment}$$

$$z_k = \int_{-\infty}^{\infty} x^k p(x, \xi_1, \dots, \xi_e)\, dx$$

recimo, da lahko parametre ξ izrazimo z začetnimi momenti

$$\xi_1 = \mathcal{C}_1(z_1, \dots, z_m)$$
$$\vdots$$
$$\xi_e = \mathcal{C}_e(z_1, \dots, z_m)$$

Cenilka C_e za ξ_e je potem po metodi momentov

$$C_e = \mathcal{C}_e(z_1, \dots, z_m)$$

vse te cenilke so dosledne in asimp. nepristranske

zgled: $X \sim N(\mu, \sigma)$

$$z_1 = E(X) = \mu$$

$$z_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \Rightarrow \bar{X} \text{ je cenilka po metodi momentov za } \mu$$

$$z_2 = E(X^2) = D(X) + E(X)^2$$
$$= \sigma^2 + \mu^2 = \sigma^2 + z_1^2$$
$$\Rightarrow \sigma^2 = z_2 - z_1^2 = \mathcal{C}(z_1, z_2)$$

$\Rightarrow$ cenilka po metodi momentov za σ^2 je

$$z_2 - z_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{2}{n}\left(\sum_{i=1}^{n} X_i^2\right)\bar{X} + \bar{X}^2$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 - 2X_i\bar{X} + \bar{X}^2 \right) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 = S_0^2$$

Zgled: $X \sim U_c(a,b)$

Kaj sta cenilki za a in b po metodi momentov?

$$z_1 = \frac{1}{b-a} \int_a^b x \, dx = \frac{1}{2(b-a)} (b^2 - a^2) = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}$$

$$z_2 = \frac{1}{b-a} \int_a^b x^2 \, dx = \frac{1}{3} \cdot \frac{b^3 - a^3}{b-a} = \frac{a^2 + ab + b^2}{3}$$

$$b = 2z_1 - a$$

$$3z_2 = a^2 + 2az_1 - a^2 + 4z_1^2 - 4az_1 + a^2 = a^2 + (-2z_1)a + 4z_1^2 - 3z_2 = 0$$

$$\Rightarrow a_{1,2} = \frac{2z_1 \pm \sqrt{4z_1^2 - 16z_1^2 + 12z_2}}{2} = z_1 \pm \sqrt{3z_2 - 3z_1^2} = z_1 \pm \sqrt{3}\sqrt{z_2 - z_1^2}$$

$$a < b \rightsquigarrow a = z_1 - \sqrt{3}\sqrt{z_2 - z_1^2} = \bar{X} - \sqrt{3}\sqrt{S_0^2} = \bar{X} - \sqrt{3} S_0$$

$$b = z_1 + \sqrt{3}\sqrt{z_2 - z_1^2} = \bar{X} + \sqrt{3}\sqrt{S_0^2} = \bar{X} + \sqrt{3} S_0$$

Zgled:

$-2, 0, 1, 2, 4$ vzorec

predpostavimo enakomerno porazdelitev

$$a = \bar{X} - \sqrt{3}\sqrt{S_0} \qquad b = \bar{X} - \sqrt{3}\, S_0$$

$$\bar{X} = \frac{1}{5} \cdot 5 = 1$$

$$S_0^2 = \frac{1}{5}(9 + 1 + 0 + 1 + 9) = 4 \qquad S_0 = 2$$

$$a = 1 - \sqrt{3}\,2 \doteq -2{,}46 \qquad b = 1 + \sqrt{3} \cdot 2 \doteq 4{,}46$$

2. metoda največjega verjetja (maximal likelihood method)

privzamemo, da ima X gostoto $p(x, \xi_1, \dots, \xi_\ell)$

gledamo vzorec $X_1, \dots, X_n$

def: Funkcija verjetja $L(x_1, \dots, x_n, \xi_1, \dots, \xi_\ell) = \prod_{i=1}^{n} p(x_i, \xi_1, \dots, \xi_\ell)$

in njen logaritem $\ell(x_1, \dots, x_n, \xi_1, \dots, \xi_\ell) = \ln L$

pri danih $x_1, \dots, x_n$ iščemo maksimum te funkcije:

$$\xi_{i\,max} = \varphi_i(x_1, \dots, x_n)$$

cenilka za ξ_i je $C = \varphi_i(X_1, \dots, X_n)$ in je dosledna

zgled: $X \sim Exp(\lambda)$ $\quad p(x, \lambda) = \lambda \cdot e^{-\lambda x}, \quad x > 0$

$$L(x_1, \dots, x_n, \lambda) = \lambda^n e^{-\lambda(x_1 + \dots + x_n)}$$

$$\ell(x_1, \dots, x_n, \lambda) = n \cdot \ln \lambda - \lambda \cdot \sum_{i=1}^{n} x_i$$

$$\frac{\partial \ell}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \qquad \frac{\partial^2 \ell}{\partial \lambda^2} = -\frac{n}{\lambda^2} < 0 \Rightarrow \max$$

$$0 = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \Rightarrow \lambda = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x_i} = \frac{1}{\overline{x}}$$

cenilka za λ: $\Lambda = \frac{1}{\overline{X}}$

DN: cenilka za λ po metodi momentov

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

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

$$L(x_1,\dots,x_n,\mu,\sigma) = \frac{1}{\sigma^n\sqrt{(2\pi)^n}} e^{-\frac{1}{2\sigma^2}\left((x_1-\mu)^2+\dots(x_n-\mu)^2\right)}$$

$$\ell(x_1,\dots,x_n,\mu,\sigma) = -n\ln\sigma - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\left(\sum_{i=1}^n (x_i-\mu)^2\right)$$

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{2\sigma^2}\sum_{i=1}^n 2(x_i-\mu) = 0 \quad \Rightarrow \mu = \bar{X}$$

$$\frac{\partial \ell}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^n (x_i-\mu)^2 = 0$$

$$\Rightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^n (x_i-\mu)^2 = \frac{1}{n}\sum_{i=1}^n (x_i-\bar{x})^2 = S_0^2$$

Zgled:

$X \sim Ber(p)$

cenilka za p po metodi največjega verjetja?

$$P(X=x) = p^x q^{1-x}, \quad x\in\{0,1\}$$

$$L(x_1,\dots,x_n) = p^{x_1+\dots+x_n}\cdot q^{n-(x_1+\dots+x_n)}$$

$$\ell(x_1,\dots,x_n) = \left(\sum_{i=1}^n x_i\right)\cdot\ln(p) + \left(n-\sum_{i=1}^n x_i\right)\cdot\ln(1-p)$$

$\ln p \overset{?}{=} \dfrac{\log_p p}{\log_p e}$

$$\frac{\partial \ell}{\partial p} = \frac{1}{p}\cdot\sum_{i=1}^n x_i - \frac{n-\sum_{i=1}^n x_i}{1-p} = 0$$

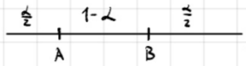
$\log_p e = \dfrac{\ln e}{\ln p}$

$$\Rightarrow p = \frac{\sum_{i=1}^n x_i}{n} = \bar{X}$$

$$x_i\in\{0,1\} \Rightarrow x = \sum_{i=1}^n x_i \in\{0,1,\dots,n\} \Rightarrow p = \frac{x}{n}$$

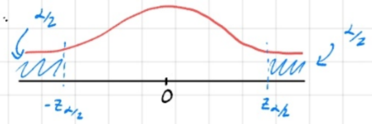
Intervalske ocene parametrov $X_1, X_2, \ldots, X_n$, $X \sim$ znana porazdelitev ξ

- interval $[A, B]$ je interval zaupanja za ξ s stopnjo tveganja α (tipično $0.05, 0.01, \ldots$) če je verjetnost
  vzorčni statistiki
  (zaupanja $1-\alpha$)

$P(\xi \in [A,B]) = 1-\alpha$ oz. $P(\xi \notin [A,B]) = \alpha$ (vzamemo $P[\xi \in (-\infty, A)] = P[\xi \in (B, \infty)] = \frac{\alpha}{2}$ )

- zgledi: ① $X \sim N(\mu, \sigma)$: predpostavimo, da je σ znan, iščemo interval zaupanja za μ

$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$

določamo α: 

$Z = \frac{\bar{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1)$

$P[-z_{\alpha/2} < Z < z_{\alpha/2}] = 1-\alpha$

$-z_{\alpha/2} < Z < z_{\alpha/2}$

$-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma}\sqrt{n} < z_{\alpha/2}$

$|\bar{X}-\mu| < \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}$

$\shortparallel$

$|\mu - \bar{X}|$

$\underbrace{\bar{X} - \frac{z_{\alpha/2}\,\sigma}{\sqrt{n}}}_{A} < \mu < \underbrace{\bar{X} + \frac{z_{\alpha/2}\,\sigma}{\sqrt{n}}}_{B} \Rightarrow \left[\bar{X} - \frac{z_{\alpha/2}\,\sigma}{\sqrt{n}}, \bar{X} + \frac{z_{\alpha/2}\,\sigma}{\sqrt{n}}\right]$ je interval zaupanja pri stopnji tveganja α

npr. $n = 36$ $\bar{X} = 2{,}6$ $\sigma = 0{,}3$ $\alpha = 0{,}05$ $z_{0{,}025} = 1{,}56$

$\Rightarrow A = 2{,}6 - \frac{1{,}56 \cdot 0{,}3}{6} = 2{,}5$ $B = 2{,}7$

② $N(\mu, \sigma)$ iščemo interval zaup. za μ pri neznanem σ

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ ← dobra cenilka za σ

$T = \frac{\bar{X}-\mu}{S}\sqrt{n} \sim t(n-1)$ $\rightsquigarrow$ vzamem $t_{\alpha/2}$ tako da: $P[|T| < t_{\alpha/2}] = 1-\alpha$

interval zaupanja: $\left[\bar{X} - \frac{t_{\alpha/2} \cdot S}{\sqrt{n}}, \bar{X} + \frac{t_{\alpha/2} \cdot S}{\sqrt{n}}\right]$

konkreten zgled: življenjska doba žarnic (dnevi): 9,8 10,2 10,4 10,0 10,2 9,6 9,8

iščemo interval zaupanja za μ, pri $\alpha = 0{,}05$:

$\bar{x} = 10{,}0$

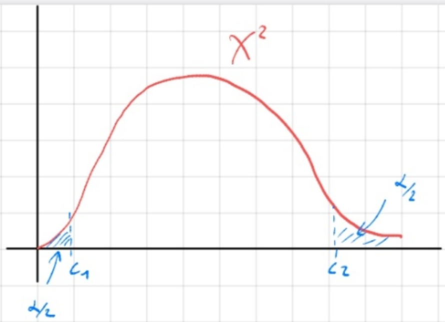
$s = \frac{1}{6}\sum(x_i - \bar{x})^2 = 0{,}283$

$a = \bar{x} - t_{0{,}025}\frac{s}{\sqrt{n}} = 9{,}74$ $\Rightarrow$ interval zaupanja $[9.74, 10.26]$

$b =$ $= 10{,}26$

$t_{0{,}025} = 2{,}45$

$(t(6))$



$P[c_1 < X^2 < c_2] = 1-\alpha$

3. $N(\mu, \sigma)$ Iščemo interval zaupanja pri stopnji tveganja α za σ^2

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

$$\chi^2 = \frac{n-1}{\sigma^2} S^2$$

$$c_1 < \chi^2 < c_2$$

$$c_1 < \frac{n-1}{\sigma^2} S^2 < c_2$$

$$\frac{c_1}{(n-1)S^2} < \frac{1}{\sigma^2} < \frac{c_2}{(n-1)S^2}$$

$$\frac{(n-1)S^2}{c_2} < \sigma^2 < \frac{(n-1)S^2}{c_1}$$ ⇝ interval zaupanja je $\left[\frac{(n-1)S^2}{c_2}, \frac{(n-1)S^2}{c_1}\right]$

za žarnice: $(n-1)s^2 = 6 \cdot 0{,}283^2 = 0.481$

za $X^2(6)$ je $c_1 = 1{,}24$, $c_2 = 14{,}45$ ⇝ $\left[\frac{0{,}481}{14{,}45}, \frac{0{,}481}{1{,}24}\right]$ ⇝ za σ^2 je int. zaupanja $[0.003, 0.388]$

oz. za σ je $[0.182, 0.623]$

4. $X \sim Ber(p)$ iščemo interval zaupanja za p pri α:

$X_1, \dots, X_n$

$$S_n = \sum_{i=1}^{n} X_i = n\bar{X}$$

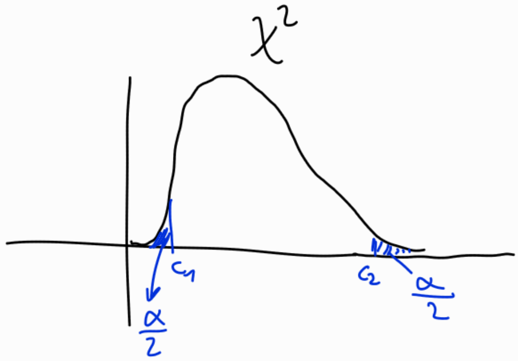
$\bar{X}$ je nepristranska in dopustna cenilka za p

Po CLI je $Z = \frac{S_n - np}{\sqrt{npq}} \approx N(0,1)$

$\bar{X} = \frac{1}{n} S_n$ tj. $S_n \sim N(np, \sqrt{npq})$

$\bar{X} = \frac{S_n}{n} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$

③ $N(\mu, \sigma)$ Iščemo interval zaupanja pri stopnji tveganja α za σ^2



$$\chi^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2 \sim \chi^2(n-1)$$

$$\chi^2 = \frac{n-1}{\sigma^2} s^2$$

$$\mathbb{P}[c_1 < \chi^2 < c_2] = 1-\alpha$$

$$c_1 < \chi^2 < c_2$$

$$c_1 < \frac{n-1}{\sigma^2} s^2 < c_2 \quad / : (n-1)s^2$$

$$\frac{c_1}{(n-1)s^2} < \frac{1}{\sigma^2} < \frac{c_2}{(n-1)s^2}$$

$$\frac{(n-1)s^2}{c_2} < \sigma^2 < \frac{(n-1)s^2}{c_1}$$

interval zaupanja je

$$\left[\frac{(n-1)s^2}{c_2}, \frac{(n-1)s^2}{c_1}\right]$$

---

primer žarnice:

$$(n-1)s^2 = 6 \cdot 0{,}283^2 = 0{,}481$$

za $\chi^2(6)$ je $c_1 = 1{,}24$

$c_2 = 14{,}45$

Potem je $a = \frac{0{,}481}{14{,}45} \doteq 0{,}033$

$b = \frac{0{,}481}{1{,}24} = 0{,}388$

za σ^2 je interval zaupanja $[0{\cdot}033, 0{\cdot}388]$

↓ korenimo

oz. za σ je

$[0{\cdot}182, 0{\cdot}623]$

(4) $X \sim Ber(p)$

iščemo interval zaupanja za p pri stopnji tveganja α

$X_1, X_2, \ldots, X_n$ $\qquad S_n = \sum_{i=1}^{n} X_i = n\bar{X}$

$\bar{X}$ je nepristranska dosledna cenilka za p.

Po CLI je $\quad Z = \dfrac{S_n - np}{\sqrt{npq}} \approx N(0,1)$

$$\bar{X} = \frac{1}{n} S_n$$

$$S_n \approx N(np, \sqrt{npq})$$

$$\bar{X} = \frac{S_n}{n} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$$

$$|\bar{X} - p| < Z_{\frac{\alpha}{2}} \cancel{\sqrt{\frac{pq}{n}}} \quad \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$$

$$\left[\bar{X} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}},\ \bar{X} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right]$$

interval zaupanja za p

konkreten zgled: volitve 2000

Na vzorcu 2207, Bush 47%

$\alpha = 0{,}05 \qquad p_{Bush}$

$Z_{0,05} = 1{,}36$

interval zaupanja za p_{Bush} je $\quad a = 0{,}47 - 1{,}96\sqrt{\dfrac{0{,}47 \cdot 0{,}53}{2207}} = 0{,}45$

$$b = 0{,}47 + 1{,}96\sqrt{\frac{0{,}47 \cdot 0{,}53}{2207}} = 0{,}49$$

$$[0{\cdot}45,\ 0{\cdot}49]$$

preizkušanje statističnih hipotez

Statistična hipoteza je kakršnakoli izjava o porazdelitvi.

$X_1, X_2, \dots, X_n$ $\qquad X$

hipoteza je enostavna, če natanko določa porazdelitev,
sicer rečemo, da je sestavljena

$N(\mu, \sigma)$ $\quad \sigma$ poznamo, hipoteza o μ je enostavna
$\quad \sigma$ ne poznamo, hipoteza o μ je sestavljena

preizkušamo ničelno hipotezo H_0 proti nasprotni alternativi H_1

npr. $H_0\ (\mu = 0)$ $\qquad H_1\ (\mu \neq 0)$ dvostranski test
$H_0\ (\mu = 0)$ $\qquad H_1\ (\mu > 0)$ enostranski test

H_0 je lahko pravilna ali nepravilna.

Na voljo imamo samo vzorec $x_1, \dots, x_n$

Iz vzorca lahko sklepamo samo o statistični značilnosti hipoteze.

Vzorec je konsistenten s hipoteza ali pa ne.

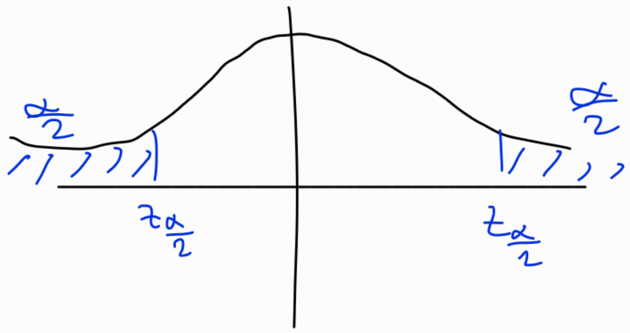
α ... stopnja značilnosti

$\alpha = 0{,}05\ ,\ 0{,}01\ ,\ 0{,}005$

Statističnim testom rečemo tudi testi značilnosti:

primeri testov: $N(\mu, \sigma)$

1) **z-test**: test za μ pri znanem σ



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

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

ne zavrnemo

$$z = \frac{\bar{x}-\mu_0}{\sigma}\sqrt{n} \in \left[-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right]$$

$$\notin \left[-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right]$$

zavrnemo

primer: Konkreten zgled. Izdelovalec vrvic trdi, da je povprečna sila, pri kateri se vrvica strga 150N s standardnim odklonom 5N.

$\sigma = 5$ $\quad H_0(\mu_0 = 150)$

imamo vzorec 50 vrvic: vzorčno povprečje 148 N

vzamemo stopnjo zaupanja $\alpha = 0{,}01$

test značilnosti: $H_0(\mu_0 = 150) : H_1(\mu \neq 150)$

$$z_{0{,}005} = 2{,}58$$

$$Z = \frac{\bar{X}-150}{5}\sqrt{50}$$

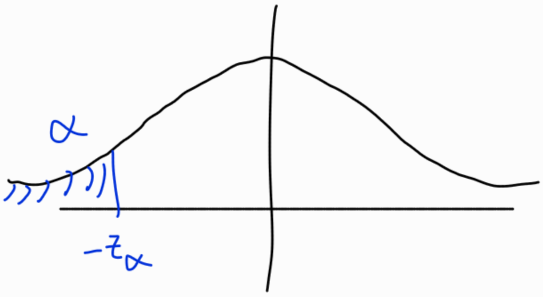
$$z = \frac{148-150}{\cancel{5}}\cancel{5}\sqrt{2} = -2\sqrt{2} \doteq -2{,}82$$

$$-2{,}82 \notin [-2{,}58, 2{,}58]$$

H_0 zavrnemo

V konkretnem primeru je smiseln tudi enostranski test

$H_0(\mu_0 = 150) : H_1(\mu < 150)$



$-z_\alpha > -z_{\frac{\alpha}{2}}$

$-z_\alpha = -2{,}33$

$z = -2{,}82 < -2{,}33$

H_0 zavrnemo

2) T-test : $N(\mu, \sigma)$, σ neznan

Za odločanje o značilnosti hipotez uporabimo T-statistiko

$$T = \frac{\bar{X} - \mu}{S} \sqrt{n}$$

izračunamo T na vzorcu in preverimo ali velja $t \notin [-t_{\frac{\alpha}{2}}, t_{\frac{\alpha}{2}}]$

v primeru dvostranskega testa $H_0 (\mu = \mu_0)$

primer: vrvice, le da ne zaupamo podatku $\sigma = 5$ (σ ni znan)

Na vzorcu dobimo $s = 5{,}01$

$H_0(\mu_0 = 150) : H_1(\mu \neq 150)$

$t_{\frac{\alpha}{2}} = 2{,}68$

$t = \frac{148 - 150}{5{,}01} \sqrt{50} \doteq -2{,}82$ H_0 zavrnemo, ker $-2{,}82 < -2{,}68 = -t_{\frac{\alpha}{2}}$

Nova predavanja

zgled: Pri testiranju zdravila za znižanje krvnega tlaka 10 bolnikov so izmerili sistolični tlak pred in po zdravljenju. Razlike "pred-po" so:

pred 140 po 120 pred-po = 20

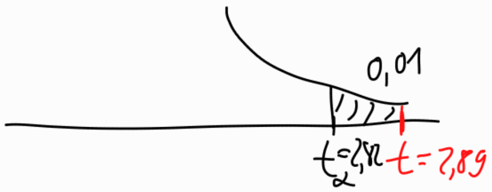
$-8, 2, 4, 3, 14, 19, 22, 32, 35$ mmHg

Test $H_0(\mu=0)$: $H_1(\mu>0)$ (testiramo pozitiven vpliv zdravila)

μ matematično upanje razlik

razlike porazdeljene $N(\mu, \sigma)$, $\alpha = 0{,}01$ $t_\alpha = 2{,}82$

uporabimo T-test : $T = \dfrac{\overline{X}-\mu}{S}\sqrt{n} \sim t(n-1)$



t na vzorcu

$\overline{x} = 12{,}9$

$s = 18{,}1$

$$t = \frac{12{,}9}{81{,}1}\sqrt{10} = 2{,}89$$

ne ovržemo

p-vrednost vzorca je najmanjša stopnja značilnosti, pri kateri še zavrnemo H_0

$p = 0{,}0089 = 0{,}89\%$

pri $\alpha = 0{,}01$ je $t_\alpha = 2{,}82$

3) Studentov primerjalni test

Imamo 2 vzorca $X_1, \dots, X_n$ $X \sim N(\mu_X, \sigma_X)$

$Y_1, \dots, Y_m$ $Y \sim N(\mu_Y, \sigma_Y)$

privzamemo σ_X, σ_Y

$H_0 (\mu_X = \mu_Y) : H_1 (\mu_X \neq \mu_Y)$

vzamemo $T = \frac{\bar{X} - \bar{Y}}{\underbrace{S}_{\text{kombiniran}}}$

definiramo skupno vzorčno disperzijo, ker $\sigma_X = \sigma_Y$

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\bar{X} - X_i)^2 \qquad S^2 = \frac{(n-1) S_X^2 + (m-1) S_Y^2}{\underbrace{n+m-2}_{\text{izgubimo 1 prostostno stopnjo pri } x \text{ in } y}}$$

$$S_Y^2 = \frac{1}{m-1} \sum_{j=1}^{m} (\bar{Y} - Y_i)^2$$

Za test uporabimo t-statistiko

$X \sim N(\mu_X, \sigma)$ $\bar{X} \sim N(\mu_X, \frac{\sigma}{\sqrt{n}})$

$Y \sim N(\mu_Y, \sigma)$ $\bar{Y} \sim N(\mu_Y, \frac{\sigma}{\sqrt{m}})$ $D(X+Y) = D(X) + D(Y)$

$$D(\bar{X} - \bar{Y}) = \frac{\sigma^2}{n} + \frac{\sigma^2}{m}$$

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \sigma \sqrt{\frac{1}{m} + \frac{1}{n}}\right) \quad \text{korenimo} \quad = \sigma^2 \cdot \left(\frac{1}{n} + \frac{1}{m}\right)$$

$$z = \frac{\bar{X} - \bar{Y}}{\sigma \sqrt{\frac{m+n}{m \cdot n}}} = \frac{\bar{X} - \bar{Y}}{\sigma} \sqrt{\frac{mn}{m+n}}$$

$$T = \frac{\bar{X} - \bar{Y}}{S} \sqrt{\frac{m \cdot n}{m+n}} \sim t(m+n-2)$$

dokazi:

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

$$\frac{S_Y^2 (m-1)}{\sigma^2} \sim \chi^2(m-1)$$

$$\frac{S^2 (n+m-2)}{\sigma^2} \sim \chi^2(n+m-2)$$

Zgled: Dve zdravili proti nespečnosti preizkušajo na dveh vzorcih velikosti 10.

Dodatno število ur spanja pri prvem je

pri prvem: 1·9, 0·8, 1·1, 0·1 −0·1 4·4 5·5 1·6 4·6 3·4 $n=10$

pri drugem: 0·7 −1·6 −0·2 −1·2 −0·1 3·4 3·7 0·8 0·0 2·0 $m=10$

dvostranski test $H_0(\mu_X=\mu_Y) : H_1(\mu_X \neq \mu_Y)$

$\alpha=0{,}05$ σ enaki

t-test

$$T = \frac{\bar{X}-\bar{Y}}{S}\sqrt{\frac{n \cdot m}{n+m}}$$

$$T = \frac{\bar{X}-\bar{Y}}{S}\sqrt{5} \sim t(18)$$

$$t_{\frac{\alpha}{2}} = 2{,}10$$

vzorčne vrednosti:

$\bar{X} = 2{,}33$

$\bar{Y} = 0{,}75$

$S_X^2 = 4{,}4$

$S_Y^2 = 3{,}0$

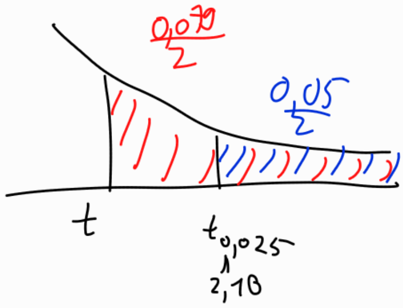
$S^2 = 3{,}70$

$$t = \frac{2{,}33 - 0{,}75}{\sqrt{3{,}7}}\sqrt{5} = 1{,}84$$

$t < t_{\frac{\alpha}{2}}$

hipoteze ne moremo zavrniti

p-vrednost: 7,9%



Pri prejšnjem testu smo privzeli $\sigma_X = \sigma_Y$.

kako preveriti smiselnost te predpostavke?

$$H_0(\sigma_X = \sigma_Y) : H_1(\sigma_X \neq \sigma_Y)$$

f-statistika:

$\sigma_X = \sigma_Y$

$$F = \frac{\frac{S_X^2}{\sigma_X^2}}{\frac{S_Y^2}{\sigma_Y^2}} = \frac{S_X^2}{S_Y^2}$$

$$F = \frac{\frac{(n-1) S_X^2}{\sigma_X^2 (n-1)}}{\frac{(m-1) S_Y^2}{\sigma_Y^2 (m-1)}} = \frac{\chi^2_{n-1} \cdot (m-1)}{\chi^2_{m-1} \cdot (n-1)} \sim F(n-1, m-1)$$

χ^2_{n-1}

χ^2_{m-1}

Fischer- Snedecorjeva porazdelitev z n-1 in m-1 prostostnimi stopnjami

$$f_{d,e}(x) = \frac{1}{B\left(\frac{d}{2}, \frac{e}{2}\right)} \cdot \left(\frac{d}{e}\right)^{\frac{d}{2}} x^{\frac{d}{2}-1} \left(1 + \frac{d}{e} x\right)^{-\frac{d+e}{2}}$$

F-porazdelitev z d in e prostostnimi stopnjami

beta funkcija

Zgled: Ali je predpostavka $\sigma_X = \sigma_Y$ smiselna pri zdravilih proti nespečnosti?

$\alpha = 0{,}05$

$$F = \frac{S_X^2}{S_Y^2} = \frac{4{,}4}{3{,}0} = 1{,}46$$

$m-1 = n-1 = 9$

$f_{9,9}$ porazdelitev

$f_{0,05} = 3{,}18$ $\quad f_{0,025} > f_{0,05} > 1{,}46 \Rightarrow H_0$ ne moremo zavrniti

Neparametrični testi

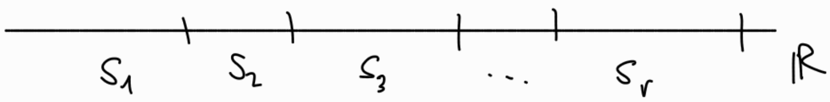
$X_1, X_2, \ldots, X_n \qquad X \sim F$

porazdelitvena funkcija za X

F ni znana. Testiramo $H_0\,(F = F_0) : H_1\,(F \neq F_0)$

za neko dano porazdelitveno funkcijo F

1) χ^2 - test



$\mathbb{R}$ razdelimo na r disjunktnih podmnožic, tako da $p_j = P[X \in S_j] > 0$ in $\sum_{j=1}^{r} p_j = 1$

privzamemo, da ima X porazd. funk. F_0

N_j ... število vrednosti vzorca, ki padejo v S_j

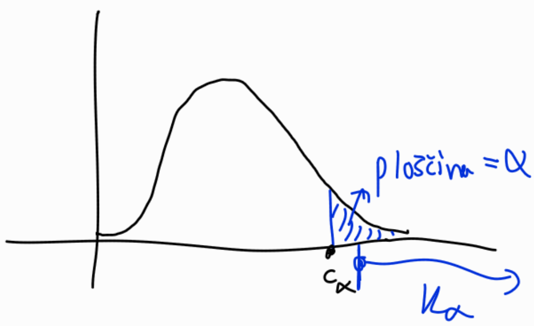
$$\sum_{j=1}^{r} N_j = n$$

$$N_j \sim Bin(n, p_j)$$

$$E(N_j) = n \cdot p_j$$

pri velikih n ima vzorčna statistika

$$\chi^2 = \sum_{j=1}^{r} \frac{(N_j - np_j)^2}{n \cdot p_j} \sim \chi^2(r-1)$$



Zgled: preizkušamo poštenost igralne kocke.

Pri 120 metih smo dobili naslednje rezultate

$p_j = \frac{1}{6}$ za $j = 1,2,3,4,5,6$

$S_j = \{j\}$

vrednost	1	2	3	4	5	6
frekvenca	20	22	17	18	19	24

$n \cdot p_j = 20 \quad \forall j$

$$\chi^2 = \frac{(20-20)^2}{20} + \frac{(22-20)^2}{20} + \ldots + \frac{(20-24)^2}{20} = \frac{34}{20} = 1{,}7$$

$\chi^2 \sim \chi^2(r-1) = \chi^2(5)$

$c_{0,05} = 11{,}1 \qquad K_\alpha = [11{,}1, \infty)$

H_0 ne moremo zavrniti

Zgled: podatki o potresih vsaj 8 po Richterju v letih 1969 – 2001: $n = 33$ (let)

število potresov v 1 letu	0	1	2	3	4 ali več
št let	15	13	4	1	0

preizkušamo domnevo $X \sim \text{Pois}(1{,}5)$ $\qquad X_1, \ldots, X_{33}$

$$p_i = \frac{(1{,}5)^i}{i!} e^{-1{,}5}$$

$n \cdot p_0 = 33 \cdot e^{-1{,}5} \doteq 7{,}4$

$n \cdot p_1 = 33 \cdot 1{,}5 \cdot e^{-1{,}5} \doteq 11{,}0$

$n \cdot p_2 = 33 \cdot \frac{1{,}5}{2} e^{-1{,}5} \doteq 8{,}3$

$\left.\begin{array}{l} n \cdot p_3 = 4{,}1 \\ n \cdot p_4 = 1{,}6 \end{array}\right\}$ pričakovane frekvence naj bodo vsaj 5, drugače združimo frekvence razredi : 0,1,2,3 ali več

S_1, S_2, S_3, S_4

št. potresov	0	1	2	3 ali več	
št. let	15	13	4	1	
pričakovana frekvenca	7,4	11,0	8,3	6,3	vsota frekvenc mora biti 33

$6{,}3 \downarrow n \sum_{j=3}^{\infty} p_j$

4 razredi

$$\chi^2 = \frac{(15-7{,}4)^2}{7{,}4} + \frac{(13-11)^2}{11} + \frac{(4-8{,}3)^2}{8{,}3} + \frac{(1-6{,}3)^2}{6{,}3} = 14{,}9 \qquad \chi^2(3)$$

$\alpha = 0{,}05 \quad c_{0{,}05} = 7{,}82$

$14{,}9 > 7{,}82$ smo v kritičnem območju $\Rightarrow$ zavrnemo H_0

$\overline{X}$ je nepristranska cenilka za λ

$$\overline{x} = \frac{15\cdot 0 + 13\cdot 1 + 4\cdot 2 + 3}{33} = \frac{24}{33} = 0{,}73 \neq 1{,}5$$

bolj smiselno bi bilo testirati $X \sim Pois(0{,}73)$

Pri tem λ_0 hipoteze ne zavrnemo

če so $p_i(\lambda)$ odvedljive funkcije, potem ima statistika

$$\chi^2 = \sum_{i=1}^{r} \frac{(N_i - np_i(\hat{\lambda}))}{np_i(\hat{\lambda})}$$

, kjer je $\hat{\lambda}$ cenilka za λ po metodi največjega verjetja

$\downarrow$

asimptotično porazdelitev $\chi^2(r-2)$

zgled s potresi: $\hat{\lambda} = \bar{x} = 0{,}73$

$np_2 < 5$

št. potresov	0	1	2 ali več
št. let	15	13	5
pričakovana frekvenca	15,9	11,6	5,5

$$\chi^2 = \frac{(15{,}9-15)^2}{15{,}9} + \frac{(11{,}6-13)^2}{11{,}6} + \frac{(5{,}5-5)^2}{5{,}5} = 0{,}27 \qquad \chi^2 \sim \chi^2(1)$$

$\alpha = 0{,}05$ $\quad C_\alpha = 3{,}84$ $\quad 0{,}28 < 3{,}84 \Rightarrow$ Hi~~potez~~ze ne zavrnemo

2) test z znaki (predznaki)

na isti populaciji merimo vrednosti slučajne spremenljivke X in Y

$(X_1, \dots, X_n)$, $(Y_1, \dots, Y_n)$

preizkušamo hipotezo $H_0(F_X = F_Y)$ (porazdelitvi enaki)

poiščemo razlike $D_i = X_i - Y_i$

izpustimo tiste D_i, ki so enaki 0 (ustrezno zmanjšamo n)

naj bo S^+ število pozitivnih razlik D_i

če je $F_X = F_Y$, potem je $S^+ \sim Bin(n, \frac{1}{2})$

$$p_i = P[S^+ = i] = \binom{n}{i} \left(\tfrac{1}{2}\right)^n$$

pri dani stopnji zaupanja α je kritično območje $\{i \mid i \leq k_\alpha \text{ ali } i \geq n-k_\alpha\}$,

kjer je k_α določen tako, da je $\sum_{i=0}^{k_\alpha} p_i \leq \frac{\alpha}{2}$ in $\sum_{i=0}^{k_\alpha+1} p_i > \frac{\alpha}{2}$

Pri velikih n je $S^+ \sim N\left(\frac{n}{2}, \frac{\sqrt{n}}{2}\right)$

$$Z = \frac{S^+ - \frac{n}{2}}{\frac{\sqrt{n}}{2}} = \frac{2S^+ - n}{\sqrt{n}} \sim N(0,1)$$

zgled: na podatkih o učinkovitosti zdravila za zmanjšanje krvnega tlaka

testiramo $H_0\,(F_X = F_Y)$ X ... pred jemanjem

Y ... po jemanju

razlike so $\underset{[\text{pred}-\text{po}]}{D_i}$: $-8 \quad 0 \quad 2 \quad 4 \quad 9 \quad 14 \quad 19 \quad 22 \quad 32 \quad 35$

zmanjšani $n = 9$

$S^+ = 8$

$$P(S^+ \le 1) = p_0 + p_1 = \binom{9}{0}\left(\frac{1}{2}\right)^9 + \binom{9}{1}\cdot\left(\frac{1}{2}\right)^9 = 0{,}02 \le 0{,}025$$

$$P(S^+ \le 2) = \binom{9}{0}\left(\frac{1}{2}\right)^9 + \binom{9}{1}\left(\frac{1}{2}\right)^9 + \binom{9}{2}\left(\frac{1}{2}\right)^9 = 0{,}09 > 0{,}025$$

$\Rightarrow k_\alpha = 1$ kritična množica je $\{0, 1, 9, 8\} = K_{0,05}$

$8 \in K_{0,05}$

H_0 zavrnemo (zdravilo je učinkovito)

3.) test z rangi (inverzijski test) oz. test Wilcoxon-Mann-Whitney

$X_1, X_2, \dots, X_m \quad , \quad Y_1, \dots, Y_n \qquad \underset{\text{BŠS}}{(m \le n)}$

$H_0\,(F_X = F_Y)$

imamo $m+n$ podatkov in jih razvrstimo po velikosti:

$$z_1 \le z_2 \le \dots \le z_{m+n}$$

rang: $1 \quad 2 \quad \dots \quad m+n$

$R_i = \text{rang } X_i = j$, kjer je $x_i = z_j$

vsota rangov $V = \sum_{i=1}^{m} R_i$

V zavzame vrednosti med $1+2+\dots m = \frac{m\cdot(m+1)}{2}$

$$(n+1)+\dots+(n+m) = mn + \frac{m\cdot(m+1)}{2}$$

Ob predpostavki, da je $m+n \geq 20$ in $m \geq 4$ $(n \geq 4)$

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

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

zgled: Zdravili proti nespečnosti

$m = n = 10$ (podatki o podaljšanju spanja pri posameznem zdravilu na 2 populacijah)

vzorčne vrednosti razvrstimo po velikosti:

z_i: −1,6 −1,2 −0,2 −0,1 (× 4,5) −0,1 0,0 0,1 (7 ×) 0,7 0,8 (9,5 ×) 0,8 1,1 (11 ×) 1,6 (12 ×) 1,9 (13 ×) 2,0
3,4 (× 15,5) 3,4 3,7 4,4 (18 ×) 4,6 (19 ×) 5,5 (20 ×)

$$V = 4{,}5 + 7 + 9{,}5 + 11 + 12 + 13 + 15{,}5 + 18 + 19 + 20 = 129{,}5$$

$$Z = \frac{\sqrt{3}}{10\sqrt{21}}(2\cdot 129{,}5 - 210) = 1{,}85$$

pri $\alpha = 0{,}05$ je $z_{0{,}025} = 1{,}96$ $Z < z_{0{,}025} \Rightarrow$ ne zavrnemo hipoteze

---

Zakaj je to tudi inverzijski test?

Inverzija med X_i in Y_j se pojavi, če je Y_j manjši od X_i (ima manjši rang)

Naj bo U število inverzij v podatkih

velja: $V = U + \frac{m(m+1)}{2}$ $(m \leq n)$

$V \sim N(\quad , \quad)$ z-test je isti

MANJKA OD NEČUCA

LINEARNA REGRESIJA

X neodvisna spremenljivka

Y, ki je odvisna od X

Y slučajna spremenljivka

$\mu(x)$ $(= E(Y))$ je funkcija X

Dan imamo vzorec parov $(x_1, y_1), \ldots, (x_n, y_n)$

$Y \sim N(\mu(x), \sigma)$

$Y = \underbrace{ax + b}_{\mu(x)} + \varepsilon$ $\quad (\varepsilon \sim N(0, \sigma))$

σ je neodvisen od x

a, b sta iskana parametra

Iščemo statistiki za a in b

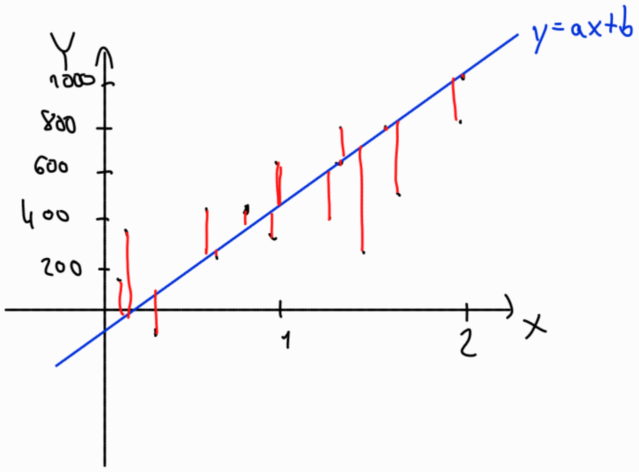
<u>Zgled:</u> Hubble (1929) je ugotavljal hitrost oddaljevanja galaksij od Zemlje kot funkcijo oddaljenosti galaksije od Zemlje.

x ... oddaljenost

Y ... hitrost oddaljevanja

$n = 24$

oddaljenost (parsek)	0,032	0,034	0,214	...	2,0	2,0
hitrost (km/s)	170	290	-130		800	1090



Iščemo a, b, ki minimizirata vsoto kvadratov razdalj | (LSE)

cenilki za a in b po MNK:

$(x_1,y_1), \ldots, (x_n,y_n)$ dan vzorec

iščemo min funkcije $F(a,b) = \sum_{i=1}^{n} (y_i - (ax_i+b))^2$

$$\frac{\partial F}{\partial a} = 0, \quad \frac{\partial F}{\partial b} = 0$$

$$2\sum_{i=1}^{n} (y_i - ax_i - b)(-x_i) = 0 \quad /: -2$$

$$2\sum_{i=1}^{n} (y_i - ax_i - b)(-1) = 0 \quad /: -2$$

$$\sum_{i=1}^{n} (y_i - ax_i - b)\cdot x_i = 0$$

$$\sum_{i=1}^{n} (y_i - ax_i - b) = 0$$

$$\sum_{i=1}^{n} y_i x_i - a\sum_{i=1}^{n} x_i^2 - b\sum_{i=1}^{n} x_i = 0$$

$$\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i - n\cdot b = 0 \quad /: n$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\sum_{i=1}^{n} y_i x_i - a\sum_{i=1}^{n} x_i^2 - b\cdot n\cdot\bar{x} = 0$$

$$\bar{y} - a\bar{x} - b = 0 \quad /\cdot -n\bar{x}$$

$$\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} - a\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) = 0$$

$\Rightarrow$ dobimo a

$$a = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

$$b = \bar{y} - a\bar{x}$$

Zgled: Hubble je dobil $y = 465x - 43$

MLE

Kaj pa statistično? Po metodi največjega verjetja (MNV) (maximal likelihood)

$$p_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y-ax-b}{\sigma}\right)^2}$$

$$L(a,b) = \prod_{i=1}^{n} p_{Y_i}(y)$$

$$\ell(a,b) = \ln L(a,b) = \sum_{i=1}^{n} \ln p_{Y_i}(y) = -\frac{n}{2}\ln(2\pi) - n\cdot\ln(\sigma) - \frac{1}{2\sigma_2}\sum_{i=1}^{n}(y_i - ax_i - b)^2$$

$$\frac{\partial \ell}{\partial a} = 0 \qquad \frac{\partial \ell}{\partial b} = 0 \qquad \frac{\partial \ell}{\partial \sigma} = 0$$

$$\boxed{\hat{a} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \qquad \hat{b} = \bar{y} - \hat{a}\bar{x}}$$

regresijska premica $y = \hat{a}x + \hat{b}$ je enaka po MNK kot MNV

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - 2\bar{x}\overbrace{\sum_{i=1}^{n} x_i}^{=n\bar{x}} + n\bar{x}^2\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right)$$

$$S_x^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$$

Potem je $\hat{a} = \dfrac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{(n-1)\, S_x^2}$

$$\hat{b} = \bar{y} - \hat{a}\bar{x}$$

Zgled: Imamo podatke o gostoti vzorcev železove rude (X) in količini železa (Y) v njej. Na podlagi vzorca velikosti 9 poišči regresijsko premico.

Vzorec:

(g/cm^3) x_i	y_i %	x_i^2	$x_i y_i$
2,8	27	7,84	.
2,9	23	8,4	.
3,0	30	.	.
3,1	28	.	.
3,2	30	.	.
3,2	32	.	.
3,3	33	.	.
3,4	30	.	.

$$\bar{x} = \frac{28,1}{9} = 3,12$$

$$\bar{y} = 29,67$$

$$s_x^2 = \frac{1}{8}\left(88,03 - \frac{28,1^2}{9}\right) = 0,0369$$

$$S_{xy} = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right) =$$

$$= \frac{1}{8}\left(837,2 - 9 \cdot 3,12 \cdot 29,67\right)$$

$$= 0,4458$$

$$\hat{a} = 12,07$$

$$\hat{b} = \bar{y} - \hat{a}\bar{x}$$

$$\hat{b} = -7,99$$

regresijska premica je

$$y = 12,07x - 7,99$$

$$(X,Y) \sim N(\mu_x, \mu_Y, \sigma_x, \sigma_Y, \rho)$$

$$E(Y \mid X=x) \sim N\left(\mu_1 + \rho\frac{\sigma_x}{\sigma_y}(x-\mu_x)\right)$$

$$\alpha x + \beta$$

$$\alpha = \rho\frac{\sigma_x}{\sigma_y} \qquad \beta = \mu_y + \rho\frac{\sigma_x}{\sigma_y}\mu_x$$

Nova predavanja

Korelacijska analiza

$$r(X,Y) = \frac{K(X,Y)}{\sigma(X)\,\sigma(Y)}$$ korelacijski koeficient

$$-1 \leq r(X,Y) \leq 1$$

(X_i, Y_i) $\quad i = 1, 2, \ldots, n$

Cenilka za r po metodi momentov (korelacijski koeficient)

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

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

vzorčna kovarianca (popravljena)

delimo z $n-1$

$$R = \frac{S_{XY}}{S_X\, S_Y}$$

opomba: R se ne spremeni, če uporabimo transformacijo $U = kX - x_0$ in $V = lY - y_0$

Predpostavimo, da je $(X,Y) \sim N(\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$

Potem ima R gostoto

$$p_R(r) = \frac{1}{B\left(\frac{n-2}{2}, \frac{1}{2}\right)}\left(1 - r^2\right)^{\frac{n-4}{2}} \quad \text{za } -1 < r < 1$$

in $p_R(r) = 0$; sicer

Pri testiranju hipotez, oz. iskanju intervala zaupanja pa vaje uporabljamo naslednjo statistiko:

$$T = \frac{R}{\sqrt{1-R^2}}\sqrt{n-2} \sim t(n-2)$$

Zgled: V graški univerzitetni kliniki so dobili naslednje podatke o velikosti in obsegu glave novorojenčkov (x_i, y_i), $n=100$

obseg \ velikost	47	48	49	50	51	52	53	54	55	56	vsota
39							1			1	2
38							1	1			2
37					1	1	1	1			4
36				1	4	7	2				14
35	3		5	9	6	1	1				25
34	1	7	10	9	3	3					33
33	1	6	5	4			1				17
32		1	1	1							3
	5	14	21	24	14	12	7	2	0	1	100

$$\sum_{i=1}^{100} x_i = 5009 = n\bar{x}$$

$$\sum_i y_i = 3460 = n\bar{y}$$

$$\sum_i x_i^2 = 251\,215 = C_x^2$$

$$\sum_i y_i^2 = 119\,908 = C_y^2$$

$$\sum_i x_i y_i = 173\,477 = C_{xy}$$

$$R = \frac{S_{xr}}{S_x S_y} = \frac{\sum_{i=1}^{100} x_i y_i - \frac{1}{100}\sum_{i=1}^{100} x_i \cdot \sum_{i=1}^{100} y_i \quad \cdot 99\, S_{xy}}{\sqrt{\sum_{i=1}^{100} x_i^2 - \frac{1}{100}\left(\sum_{i=1}^{100} x_i\right)^2 \sqrt{99}\, S_x + \sum_{i=1}^{100} y_i^2 - \frac{1}{100}\left(\sum_{i=1}^{100} y_i\right)^2 \sqrt{99}\, S_y}}$$

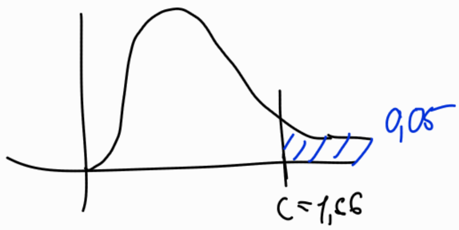
$$r = \frac{\ldots}{\ldots} = 0{,}674$$

Na danem vzorcu testiraj hipotezo, da je $\rho = 0$ proti $H_1 (\rho > 0)$ pri stopnji zaupanja $\alpha = 0{,}05$

ρ — korelacijski koeficient

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{98} \sim t(98)$$

$c = 1{,}66$



$$t \text{ na podatkih } = \frac{0{,}674}{\sqrt{1-0{,}674^2}} \sqrt{98} = 9{,}03$$

H_0 zavrnemo

če H_0 zavrnemo, zavrnemo tudi hipotezo H_0 (X, Y neodvisni)

Statistika

$$Z = \frac{1}{2} \ln \frac{1+R}{1-R} \sim N\Big(\underbrace{\frac{1}{2} \ln \frac{1+\rho}{1-\rho}}_{= \mu_0}, \underbrace{\frac{1}{\sqrt{n-3}}}_{= \sigma_0}\Big)$$

To statistiko uporabimo za iskanje intervala zaupanja za ρ. Če je c tak, da je

$$P\left(\left| \frac{Z - \mu_0}{\sigma_0} \right| < c \right) = 1 - \alpha$$

potem je interval zaupanja za μ_0

$$(z_0 - k, z_0 + k),$$

z_0 — vrednost z na vzorcu

kjer je $k = \frac{c}{\sqrt{n-3}}$

in iskani interval zaupanja za ρ je (r_1, r_2), kjer je

$$\boxed{r_1 = \tanh(z_0 - k) \quad \text{in} \quad r_2 = \tanh(z_0 + k)}$$

$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

zgled: Poišči interval zaupanja za zgled dojenčkov pri stopnji $\alpha = 0{,}05$

$$z_0 = \frac{1}{2} \ln \frac{1+r}{1-r} \doteq 0{,}8$$

$c = 1{,}645$??

$$k = \frac{c}{\sqrt{99}} = 0{,}2$$

$r_1 = \tanh 0{,}6 \doteq 0{,}550$

$r_2 = \tanh 1{,}0 \doteq 0{,}769$

95% interval zaupanja za ρ je $(0{,}55,\ 0{,}769)$

0,674

---

linearna regresija - nadaljevanje

(x_i, Y_i)

$Y = ax + b + U$ $\qquad U \sim N(0, \sigma)$

$$S_{xY} = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i Y_i - n\bar{x}\bar{Y}\right)$$

$Y \sim N(ax+b, \sigma)$

$$A = \frac{S_{xY}}{S_x^2} \qquad B = \bar{Y} - A\bar{X} = \bar{Y} - \frac{S_{xY}}{S_x^2}\bar{X}$$

$$R = \frac{S_{xY}}{S_x S_Y} = \frac{A S_x}{S_Y}$$

izračunajmo $E(A)$ in $E(B)$:

$$E(A) = \frac{1}{(n-1) S_x^2}\left(\sum_{i=1}^{n} x_i E(Y_i) - n\bar{x} E(\bar{Y})\right)$$

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$E(\bar{Y}) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) =$$

$$= \frac{1}{(n-1)\, S_x^2} \left(\sum_{i=1}^{n} x_i (a x_i - b) - n\bar{x}(a\bar{x} + b) \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} (a x_i + b)$$

$$= \frac{1}{n} \left(a \sum_{i=1}^{n} x_i + nb \right)$$

$$= a\bar{x} + b$$

$$= \frac{1}{(n-1)\, S_x^2} \left(a \sum_{i=1}^{n} (x_i^2 - \bar{x}^2) + b \underbrace{(n\bar{x} - n\bar{x})}_{=0} \right)$$

$$= \frac{1}{S_x^2} \, a \cdot S_x^2 = a$$

$$B = \bar{Y} - A\bar{x}$$

$$E(B) = E(\bar{Y}) - E(A\bar{x})$$

$$= E(\bar{Y}) - \bar{x}\, E(A)$$

$$= a\bar{x} + b - \bar{x}\, a$$

$$= b$$

podobno dobimo $\quad \text{var}(A) = \dfrac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \dfrac{\sigma^2}{(n-1)\, S_x^2}$

$$\text{var}(B) = \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n \cdot \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \right)}$$

Zadnja predavanja

Linearna regresija: $Y = ax + b + U$

$$A = \frac{n S_{xy} - S_x S_y}{n S_{xx} - S_x^2} \qquad B = \bar{Y} - A\bar{X}$$

za iskanje intervala zaupanja in testiranje hipotez za a in b

pa uporabimo $H_0(a = a_0) : H_1(a \neq a_0)$

$$T_b = \frac{B - b_0}{S}\sqrt{n - \frac{S_x^2}{S_{xx}}} \sim t(n-2)$$

$$T_a = \frac{A - a_0}{S}\sqrt{S_{xx} - \frac{S_x^2}{n}} \sim t(n-2)$$

kjer je $S^2 = \frac{1}{n-2}\sum_{k=1}^{n}(Y_k - Ax_k - B)^2$

---

interval zaupanja za A je $[a_0 - c_\alpha, a_0 + c_\alpha]$, kjer je

$$c_\alpha = t_\alpha \sqrt{\frac{(n-1)S_y^2 - a_0^2 S_x^2}{(n-2)(n-1)S_x^2}}$$, kjer t_α zadošča $P[T_a \geq t_\alpha] = \alpha$

$T_a \sim t(n-2)$

$$P[T_a > t_\alpha] = P\left[T_a \geq c_\alpha \sqrt{\frac{(n-2)(n-1)S_x^2}{(n-1)S_y^2 - a_0^2 S_x^2}}\right]$$

Analiza variance

Izvedemo več poskusov, kjer enega od pogojev (faktorjev) spreminjamo.

Ali faktor vpliva na parameter, ki ga testiramo, ali je vpliv samo slučajen?

Npr: testiramo več zdravil za isto bolezen, merimo donos pšenice pri različnih gnojilih,...

k zdravil recimo

$X_i \sim N(\mu_i, \sigma) \qquad i=1,\dots,k$

$X_{i,j}, \quad j=1,\dots,n_i$

spremenljivka za zdravilo i

$x_{ij}; \quad i=1,\dots,k, \quad j=1,\dots,n_i$ vzorci

$$n=\sum_{i=1}^{k} n_i$$

$H_0 \; (\mu_1=\mu_2=\dots=\mu_k)$

H_1 (niso vsi μ_i enaki)

$$\mu = \frac{1}{n}\sum_{i=1}^{k} n_i \mu_i$$

Pri H_0 velja $\mu_1=\dots=\mu_k$

$$\overline{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}$$

$$Q^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\overline{X})^2$$

$$Q_v^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2$$

$$Q_m = \sum_{i=1}^{k}(\overline{X}_i-\overline{X})^2$$

$$\boxed{Q^2 = Q_v^2 + Q_m^2}$$

Naj bo $S_v^2 = \frac{1}{n-k} Q_v^2$ in

$$S_m^2 = \frac{1}{k-1} Q_m^2$$

Izkaže se, da sta S_v^2 in S_m^2 nepristranski statistiki za σ^2.

$$E(S_v^2) = E(S_m^2) = \sigma^2$$

Če H_0 velja, potem je $F = \frac{S_m}{S_v}$ blizu 1.

F ima Fisher-Smedecorjevo porazdelitev $F(k-1, n-k)$

Običajno naredimo tabelo:

	VK	PS	PV	F
faktor	Q_m^2	$k-1$	S_m^2	F
slučaj	Q_v^2	$n-k$	S_v^2	

F uporabimo za testiranje hipotez.

zgled: 15 enako velikih parcel posejemo s pšenico iste sorte.

Uporabimo 3 različna gnojila, vsakega na 5 parcelah.

Povprečni pridelek pšenice na hektar je bil:

1	47	47	40	32	40
2	76	68	71	46	54
3	49	40	34	36	44

Testiramo H_0 (vsi donosi so enaki)

$n_1 = n_2 = n_3 = 5 \qquad n = 15$

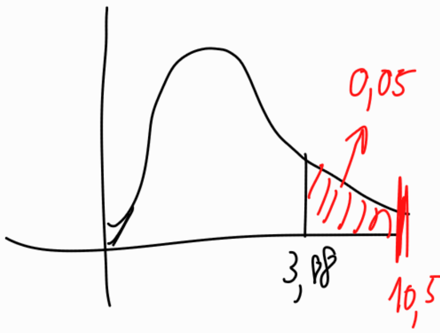
$k = 3$

	VU	PS	PV	F
faktor	1628,9	2	814,5	10,5
slučaj	930,0	12	77,5	

$$F = \frac{S_m}{S_v} \sim f(2,12)$$

$\alpha = 0{,}05$

$f_\alpha = 3{,}88$



hipotezo zavrnemo

Če H_0 zavrnemo, potem nas zanima bolj natančna analiza vpliva faktorja

Hipotezo H_0 $(\mu_p = \mu_q)$ za neka p in q $\quad 1 \leq p, q \leq k$ (različna)

lahko testiramo s t-statistiko

$$T_{pq} = \frac{\overline{X}_p - \overline{X}_q}{S_{pq}} \sqrt{\frac{n_p \cdot n_q}{n_p + n_q}} \sim t(n_p + n_q - 2)$$

Lahko uporabimo Duncanov test (Duncanove porazdelitve):

vzorčna povprečja $\overline{X}_i$ uredimo po velikosti:

$$\overline{X}_{i_1} \geq \overline{X}_{i_2} \geq \ldots \geq \overline{X}_{i_k}$$

Za testiranje uporabimo Duncanovo statistiko

$$D_{pq} = \frac{\overline{X}_p - \overline{X}_q}{S_v} \sqrt{\frac{2\, n_p \cdot n_q}{n_p + n_q}} \geq 0 \qquad (\overline{X}_p \geq \overline{X}_q)$$

Porazdelitev D_{pq} je odvisna od $n-k$ in od števila členov v zaporedju ℓ_{pq}

$$\overline{X}_p \geq \ldots \geq \overline{X}_q$$

Zgled: Uporabimo Duncanov test na zgledu donosov pšenice:

$\overline{X}_1 = 41,2$

$\overline{X}_2 = 63$

$\overline{X}_3 = 40,6$

$$\overbrace{\underbrace{\overline{X}_2 > \overline{X}_1}_{l_{12}=2} > \overline{X}_3}^{l_{23}=3} \qquad \underbrace{\overline{X}_1 > \overline{X}_3}_{l_{13}=2}$$

$n-k=12$

$S_N^2 = 77,5$

$S_N = 8,8$

$$P[D_{pq} > d^{l_{pq}}] = 0,05$$

index

$d_{0,05}^2 = 4,02$, $d_{0,05}^2 = 4,30$

$$\frac{2n_p \cdot n_q}{n_p + n_q} = 5$$

$$\frac{1}{S_N}\sqrt{\frac{2n_p \cdot n_q}{n_p+n_q}} = \frac{1}{3,5}$$

Na vzorcih poiščemo

$l=2 \begin{cases} d_{12} \\ d_{13} \end{cases}$ $H_0(\mu_1=\mu_2)$ $\overline{X}_2-\overline{X}_1 = 21,8$

$H_0(\mu_1=\mu_3)$ $\overline{X}_1-\overline{X}_3 = 0,6$

$l=3$ d_{23} $H_0(\mu_2=\mu_3)$ $\overline{X}_2-\overline{X}_3 = 22,4$

H_0 zavrnemo, če je $\overline{X}_p - \overline{X}_q \geq 3,5 \cdot d_{0,05}^{l_{pq}}$

$21,8 > 14$ //

$0,6 < 15$

$22,4 > 14$ //

Večrazsežni regresijski model (linearni model)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \underbrace{\varepsilon}_{\text{napaka}}$$

$\varepsilon \sim N(0, \sigma^2)$

$\to$ y je slučajna spremenljivka

$N(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k, \sigma^2)$

parametri modela so $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$, σ^2

Dan imamo model velikosti n:

$(y_1, x_{11}, x_{12}, \dots, x_{1k})$ $\quad n \geq k+1$

$(y_2, x_{21}, x_{22}, \dots, x_{2k})$ $\quad$ običajno $n \gg k$

$\vdots$

$(y_n, x_{n1}, x_{n2}, \dots, x_{nk})$

$$y_1 = \beta_0 + \beta_1 x_{11} + \dots + \beta_k x_{1k} + \varepsilon_1$$

$$y_2 = \beta_0 + \beta_1 x_{21} + \dots + \beta_k x_{2k} + \varepsilon_2$$

$$\vdots$$

$$y_n = \beta_0 + \beta_1 x_{n1} + \dots + \beta_k x_{nk} + \varepsilon_n$$

$\to$ $y = X \cdot \beta + \varepsilon$

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{array}{l} \text{za } \beta_0 \\ \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix} \end{array} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

privzamemo

$\varepsilon_i \sim N(0, \sigma^2)$

$\vec{\varepsilon} \sim N(0, \sigma^2 \cdot I)$

ε_i so neodvisni

$\operatorname{rang}(X) = k+1$

izrek: Cenilka za β po metodi najmanjših kvadratov (ki minimizira vsoto kvadratov napak) je

$$\beta = (\underbrace{\underset{(k+1)\times n}{X^T} \underset{n\times(k+1)}{X}}_{\substack{(k+1)\times(k+1) \\ \text{obrnljiva}}})^{-1} X^T y$$

dokaz:

$$f(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 \quad , \text{kjer je} \quad \varepsilon_i = y_i - \sum_{j=0}^{k} \beta_j x_{ij} \qquad x_{i0} = 1$$

$$= \varepsilon^T \varepsilon = (y - X\beta)^T (y - X\beta)$$

$$= (y^T - \beta^T X^T)(y - X\beta)$$

$$= y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X \beta$$

$$= y^T y - (\beta^T X^T y)^T - \beta^T X^T y + \beta^T X^T X \beta$$

$$f(\beta) = y^T y - 2\beta^T X^T y + \beta^T X^T X \beta$$

vso so števila

$št^T = št$

gledamo $\frac{\partial f}{\partial \beta} = \underbrace{\begin{bmatrix} \frac{\partial f}{\partial \beta_0} \\ \vdots \\ \frac{\partial f}{\partial \beta_k} \end{bmatrix}}_{\text{gradient}} = 0 - 2X^T y + 2X^T X \beta \underset{\text{pogoj za ekstrem}}{=} \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$

$$\cancel{2} X^T X \beta = \cancel{2} X^T y$$

$$\beta = (X^T X)^{-1} X^T y$$

drugi odvod:

$H(f) = 2\underbrace{X^T X}_{\substack{\text{je pozitivno} \\ \text{definitna}}}$ je konstanta > 0 $\Rightarrow$ imamo minimum

$$\Rightarrow \beta = (X^T X)^{-1} X^T y$$

(b)

izrek: b je nepristranska statistika β

dokaz:

$$y = X\beta + \varepsilon \qquad E(y) = X\beta$$

$$E(b) = E(\underbrace{(X^TX)^{-1}X^T}_{\text{konst.}}\underbrace{y}_{\text{spr.}}) = (X^TX)^{-1}X^T \cdot E(y)$$

$$= (X^TX)^{-1}X^T \cdot X\beta = \cancel{(X^TX)^{-1}}\,\cancel{(X^TX)}\,\beta = \beta$$

$$\Sigma = \begin{bmatrix} var(b_0) & cov(b_0,b_1) & \dots & cov(b_0,b_k) \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \\ cov(b_0,b_k) & \dots & \dots & var(b_k) \end{bmatrix}$$ (varianično) kovariančna matrika za b

izrek: Kovariančna matrika za cenilko b za β je

$$\Sigma = \sigma^2 (X^TX)^{-1}$$

$$\langle \Sigma a, a \rangle = var(a^Tb) \geq 0$$

$$a^Tb = a_0b_0 + a_1b_1 + \dots + a_kb_k$$

dokaz:

$$\Sigma = \overset{\text{kov. matr.}}{cov}((X^TX)^{-1}X^Ty)$$

$$= (X^TX)^{-1}X^T \underbrace{cov(y)}\,((X^TX)^{-1}X^T)^T$$

$$cov(y - Xb) = cov(y)$$

$$\parallel$$

$$cov(\varepsilon) = \sigma^2 I$$

$$= \sigma^2 (X^TX)^{-1}X^T X (X^TX)^{-1^T}$$

$$= \sigma^2 \big(\underbrace{(X^T X)}_{\text{simetrična}}{}^{-1} \big)^T =$$

$$= \sigma^2 (X^T X)^{-1}$$

Izrek (Gauß-Markov):

Med vsemi nepristranskimi linearnimi cenilkami za β ima cenilka b najmanjšo kovariančno matriko (a druga cenilka, Σ_a njena kovariančna matrika, potem je $\Sigma_a - \Sigma \underset{\substack{\text{pozitivno} \\ \text{semidefinitna}}}{\geq} 0$)

Za dokaz ne potrebujemo predpostavke normalnosti, potrebujemo druge momente.

$b = (X^T X)^{-1} X^T y$ je cenilka po MNK, je tudi cenilka po metodi momentov in po metodi največjega verjetja

---

testiranje hipotez o koeficientih v b:

$$T_i = \frac{b_i - \beta_i}{S_i} = \frac{b_i - \beta_i}{S\sqrt{(X^T X)^{-1}_{ii}}} \quad , \text{ kjer je } S^2 = \frac{(y - Xb)^T (y - Xb)}{n-k-1}$$

$$T_i \sim t(n-k-1)$$

Zgled: Poišči b in cenilko za σ, če je dan vzorec:

$n = 12$, $k = 2$

y	x_1	x_2
2	0	2
3	2	6
2	2	7
7	2	5
6	4	9
8	4	8
10	4	7
7	6	10
8	6	11
12	6	9
11	8	15
14	8	13

$$X^T X = \begin{array}{cc} & k+1 \\ k+1 & \begin{bmatrix} 12 & 52 & 102 \\ 52 & 395 & 536 \\ 102 & 536 & 1004 \end{bmatrix} \end{array}$$

$$b = (X^T X)^{-1} X^T y = \begin{bmatrix} 5{,}38 \\ 3{,}01 \\ -1{,}29 \end{bmatrix}$$

izrek: $\operatorname{cov}(b) = \sigma^2 (X^T X)^{-1}$

nepristranska cenilka za $\operatorname{cov}(b)$ je $s^2 (X^T X)^{-1}$