Osnovni pojmi

Naključni poskusi (naključna opazovanja)

Običajni fiz. poskus: čim bolj nadzirati potek → po večkratnih ponovitvah, enak odgovor

Naključni poskus: ne moremo nadzorovati vseh okoliščin / parametrov ⇒ vsakokrat drug izid

Zgled: Kovanec: grb / cifra
Kocka: 1, 2, 3, 4, 5, 6

Pri naključnem procesu ne moremo enolično napovedati rezultata samo na podlagi vhodnih podatkov.

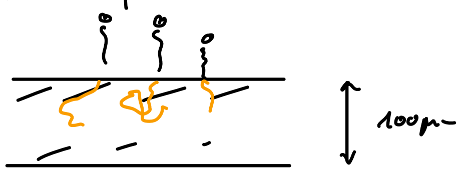
Govorili bomo lahko le o relativni pogostnosti dogodkov

Idealno bi radi imeli ∞ dogodkov / poskusov

Pri vsakem poskusu ugotovimo nek izid (outcome, elementarni dogodek

Množica vseh izidov = vzorčni prostor (sample space

Zgled: kocka: možni izidi $S = \{1, 2, \dots, 6\}$
implantacija ionov v kvarc: vzorčni prostor je interval $[0, 100]\mu m$ in je zvezen!



Elementarne dogodke lahko združujemo v sestavljene dogodke

Zgled: kocka sode / lihe pike

$\{\ \}$ = prazna množica ⇝ Nemogoč dogodek

S .. univerzalni (ali gotov) dogodek

Zgled

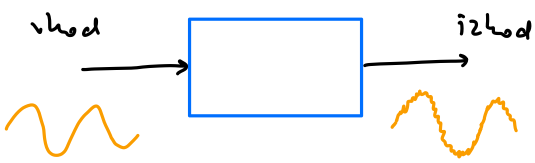
$$X(t) = \sum_{i=1}^{n} w_i \cos(\omega_i t + Y_i)$$

Očitno je to tudi naključno

znane uteži

krožna frekvenca

neodvisne naključne spremenljivke, porazdeljene enakomerno po $[0, 2\pi)$



Naključna spremenljivka je spremenljivka, ki ob vsakem poskusu zavzame vrednost samo z določeno verjetnostjo

Meritev takšnih količin imenujemo realizacijo naključne spr. oz. žrebanje

Velika črka X ... spremenljivka
mala črka x ... njena konkretna vrednost

Zgled: kovanec C = cifra, G = grb mečemo 2x

$$S = \{cc, cg, gc, gg\}$$

izberemo si dogodek A = vsaj enkrat se pojavi cifra

$$A = \{cc, cg, gc\}$$

dogodek B = pri drugem metu dobimo grb

Zgodi se A ali B ali oba: $A \cup B$
$A \cup B = \{cc, cg, gc, gg\}$

Zgodi se A in B: $A \cap B = \{cg\}$

↑
če bi prišlo $A \cap B = \{\}$
potem bi rekli da se
A in B izključujeta
(exclusive events)

$\neg A = \{gg\}$

A, vendar ne B: $A - B = A \cap \neg B = \{cc, gc\}$

$\neg A = S - A$

Vsem tem dogodkom bi radi pripisali verjetnost, da se zgodijo.

Aksiomi verjetnosti:

1. $P(A) \geq 0$ verjetnost je nenegativna
2. $P(S) = 1$ verjetnost gotovega elementa je 1 ... $P(\{\}) = 0$
3. če se dogodki izključujejo, velja $P(A \cup B) = P(A) + P(B)$
   če pa se ne izključujejo, velja $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. $P(\neg A) = 1 - P(A)$

Zgled V škatli je 10 listkov z veliko nagrado ... A
100 listkov z majhno nagrado ... B
1000 listkov brez nagrade ... C,
potegnemo 1 listek

$$P(A) = \frac{10}{1110} = 0{,}9\% \qquad P(\neg A) = 1 - P(A) = 1 - \frac{10}{1110} = 99{,}1\%$$

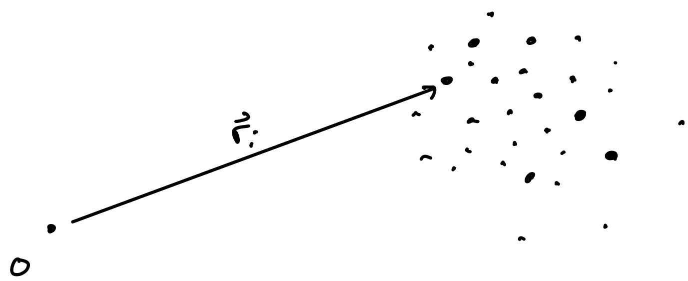
$$P(B) = \frac{100}{1110} = 9\% \qquad P(A \text{ ali } B) = P(A) + P(B) = 9{,}9\%$$

$$P(C) = \frac{1000}{1110} = 90{,}1\%$$

Diskretne in zvezne porazdelitve

Zgledi

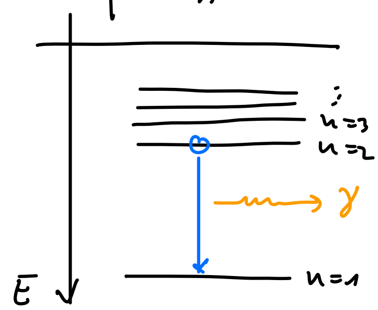
① Porazdelitev mase v zvezdnih kopicah

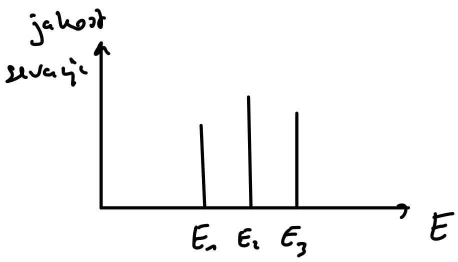


"od daleč" (iz gravitacijskega vidika) je to diskretna porazdelitev

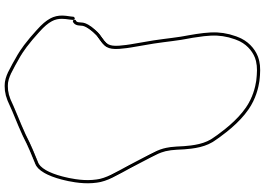
masa zvezno porazdeljena

② Sevanje H atoma





③



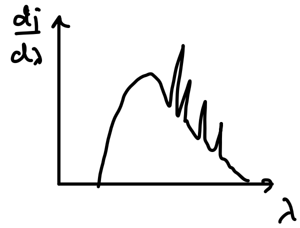
$$\rho(\vec{r}) = \frac{dm}{dV}$$

$$m = \int_{\text{telo}} \rho(\vec{r})\, dV$$

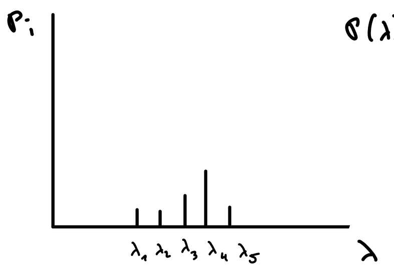
zvezna porazdelitev

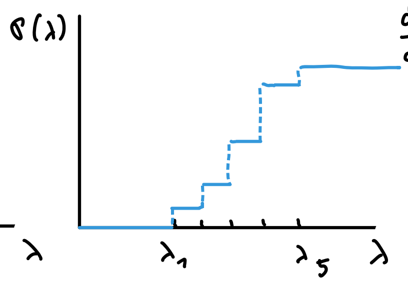
④ Mešane porazdelitve

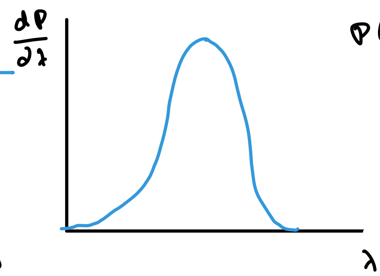
- šibre v leseni kladi
- rozine v potici
- rentgensko sevanje

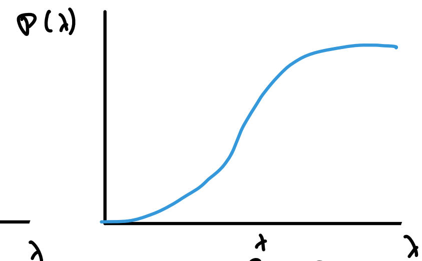


(Kumulativne) porazdelitvene funkcije







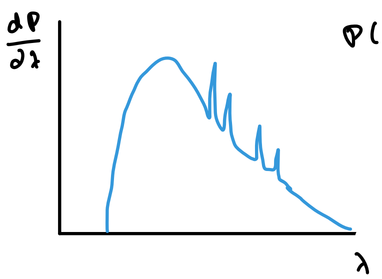


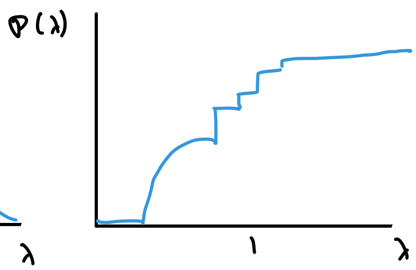
$$P(\lambda) = \sum_{\lambda_i \le \lambda} P(\lambda_i)$$

stopničasto monotona rastoča funkcija

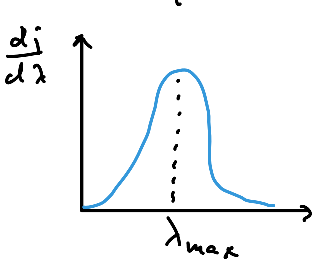
$$P(\lambda) = \int_0^{\lambda} \frac{dP}{d\lambda'}\, d\lambda'$$

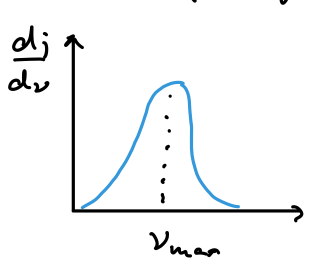
monotono rastoča





Pri vajah bomo naredili primerjavo



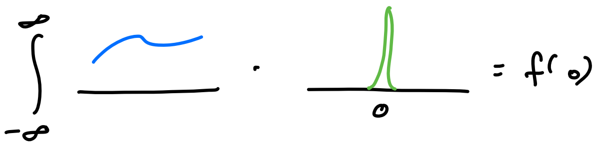


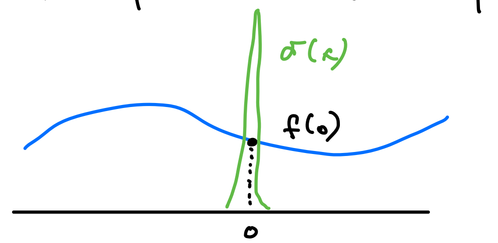
$c = \lambda \nu$

$$\lambda_{max} = \frac{c}{\nu_{max}}$$

Diracova delta funkcija

$$\int_{-\infty}^{\infty} f(x)\,\delta(x)\,dx = f(0)$$ definicija delta funkcije





Zato včasih pišejo $\delta(x) = \begin{cases} \infty; & x=0 \\ 0; & x \neq 0 \end{cases}$

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

Matematično nepravilno

Aproksimacije za delta funkcijo

$$\delta(x) = \lim_{\sigma \to 0} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-x^2/2\sigma^2}$$ Zelo ozka Gaussovka

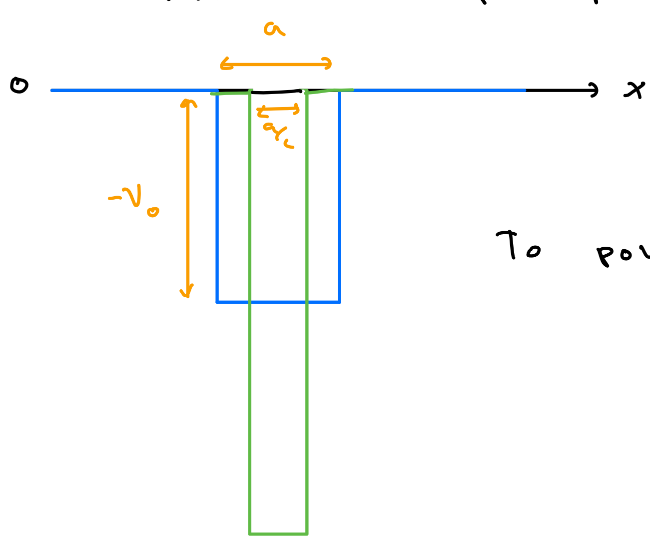
Posplošitev v 3D

$$\int_{\mathbb{R}^3} f(\vec{r})\,\delta(\vec{r})\,d\vec{r} = f(0)$$

formalno $\delta(x)\,\delta(y)\,\delta(z)$

δ funkcija ima vrednost enote $\frac{1}{[dx]}$ recipročne od diferenciala

Zgled: Približek za hipno spremembo potenciala na zelo majhni razdalji



To ponazorimo $V(x) = -c\,\delta(x)$

Lastnosti δ funkcije

- Translacija $$\int_{-\infty}^{\infty} f(x)\,\delta(x-y)\,dx = f(y)$$

- Skaliranje $$\int_{-\infty}^{\infty} \delta(\alpha x)\,dx \underset{\text{nove sprem.}}{=} \int_{-\infty}^{\infty} \delta(t)\,\frac{dt}{\alpha} = \frac{1}{|\alpha|}\,1$$

  dimenzija

  zaradi sodosti

- Sodost $$\delta(x) = \delta(-x)$$

- Kompozitum $$\int_{\mathbb{R}} f(g(x))\,\delta(g(x))\,|g'(x)|\,dx = \int_{g(\mathbb{R})} f(u)\,\delta(u)\,du$$

Ta kompozitum deluje, kot da bi imeli

$$\delta(g(x)) = \frac{\delta(x-x_0)}{|g'(x_0)|}$$

če ima g realno ničlo pri x_0

$$\int f(g(x))\,\frac{\delta(x-x_0)}{|g'(x_0)|}\,|g'(x)|\,dx =$$

δ funkcija rešili $x = x_0$

$$= f(g(x_0)) = f(0)$$

Zgled

$$\delta(x^2-a^2) = \frac{\delta(x+a)}{|g'(-a)|} + \frac{\delta(x-a)}{|g'(a)|}$$

} ker ima $g(x)$ dve ničli, eno pri $x=a$ in $x=-a$

$$g(x) = x^2 - a^2$$
$$g'(x) = 2x$$
$$|g'(a)| = |g'(-a)| = 2a$$

$$\delta(x^2-a^2) = \frac{1}{2a}\left(\delta(x-a) + \delta(x+a)\right)$$

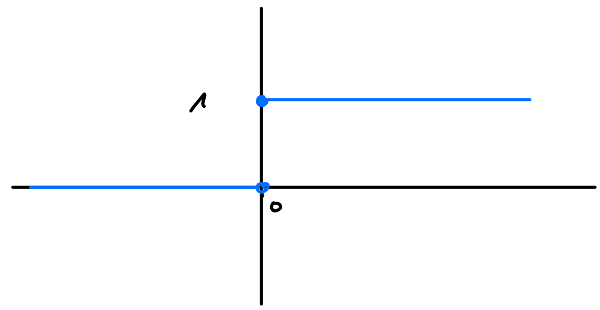
Zgled Opis porazdelitve točkastih mas

$$\underset{[\mathrm{kg}/\mathrm{m}^3]}{\rho(\vec r)} = \sum_i \underset{[\mathrm{kg}]}{m_i}\,\underset{[1/\mathrm{m}^3]}{\delta(\vec r - \vec r_i)}$$

masa zvezdne kopice $$\sum_i m_i = \int \rho(\vec r)\,d^3\vec r = \int \sum_i \delta(\vec r - \vec r_i)\,m_i\,d^3\vec r$$

Heavisidova funkcija (stopničasta, "unit step")

$$H(x) = \int_{-\infty}^{x} \delta(x)\, dx$$



Analitični aproksimaciji

- $H(x) = \lim\limits_{k\to\infty} \left(\frac{1}{2} + \frac{1}{\pi} \arctan kx \right)$

- $H(x) = \lim\limits_{k\to\infty} \dfrac{1}{1 + e^{-2kx}}$

Verjetnostne porazdelitve

Diskretne porazdelitve

X ... diskretna naključna spremenljivka, x_i $(i=1,2,\dots)$ njene vrednosti

Verjetnost da X zavzame vrednost x_i, označimo

$$P(X = x_i) = \underbrace{f_X}_{\text{verjetnostna funkcija}}(x_i)$$

Lastnosti:

- $f_X(x) \geq 0$ nenegativnost
- $\sum_i f_X(x_i) = 1$ normalizacija verj. funkcije (oz. porazdelitve)

Vpeljemo (kumulativno) porazdelitveno funkcijo (CDF = cum. dist. f)

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

Lastnosti:

- $F_X(x)$ je nepadajoča f
- $\lim\limits_{x\to-\infty} F_X(x) = 0$
- $\lim\limits_{x\to\infty} F_X(x) = 1$

Zgled: Idealno kocko vržemo 2x. Kakšno verjetnostno porazdelitev pričakujemo za vsoto pik pri teh metih?

X = naključna spem., ki meri vsoto pik

$x_1 = 1 + 1 = 2$, 36 vseh možnosti, $f_X(x_1) = 1/36$

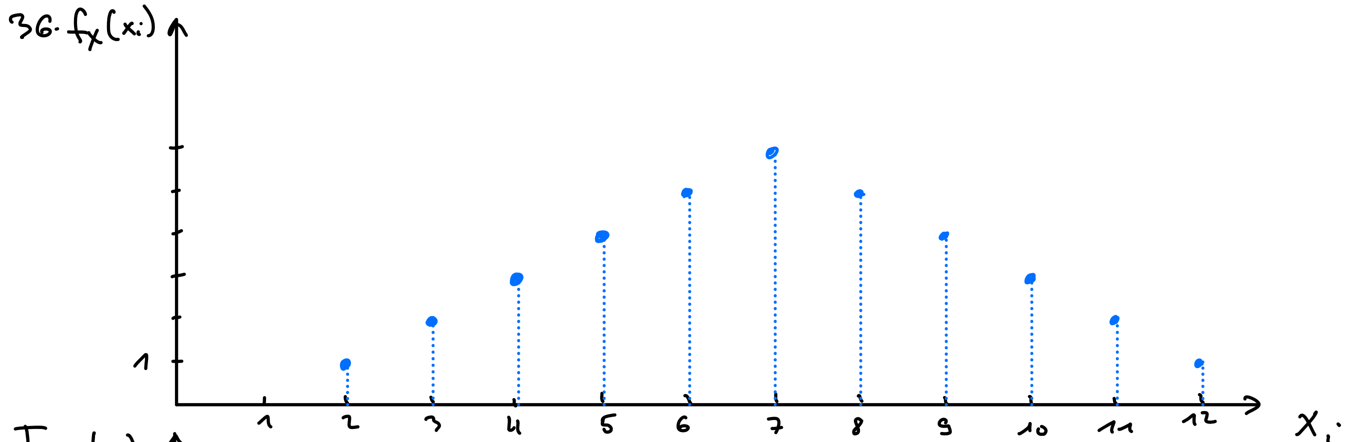
$x_2 = 1 + 2 = 2 + 1 = 3$, 36 vseh možnosti, $f_X(x_2) = 2/36$

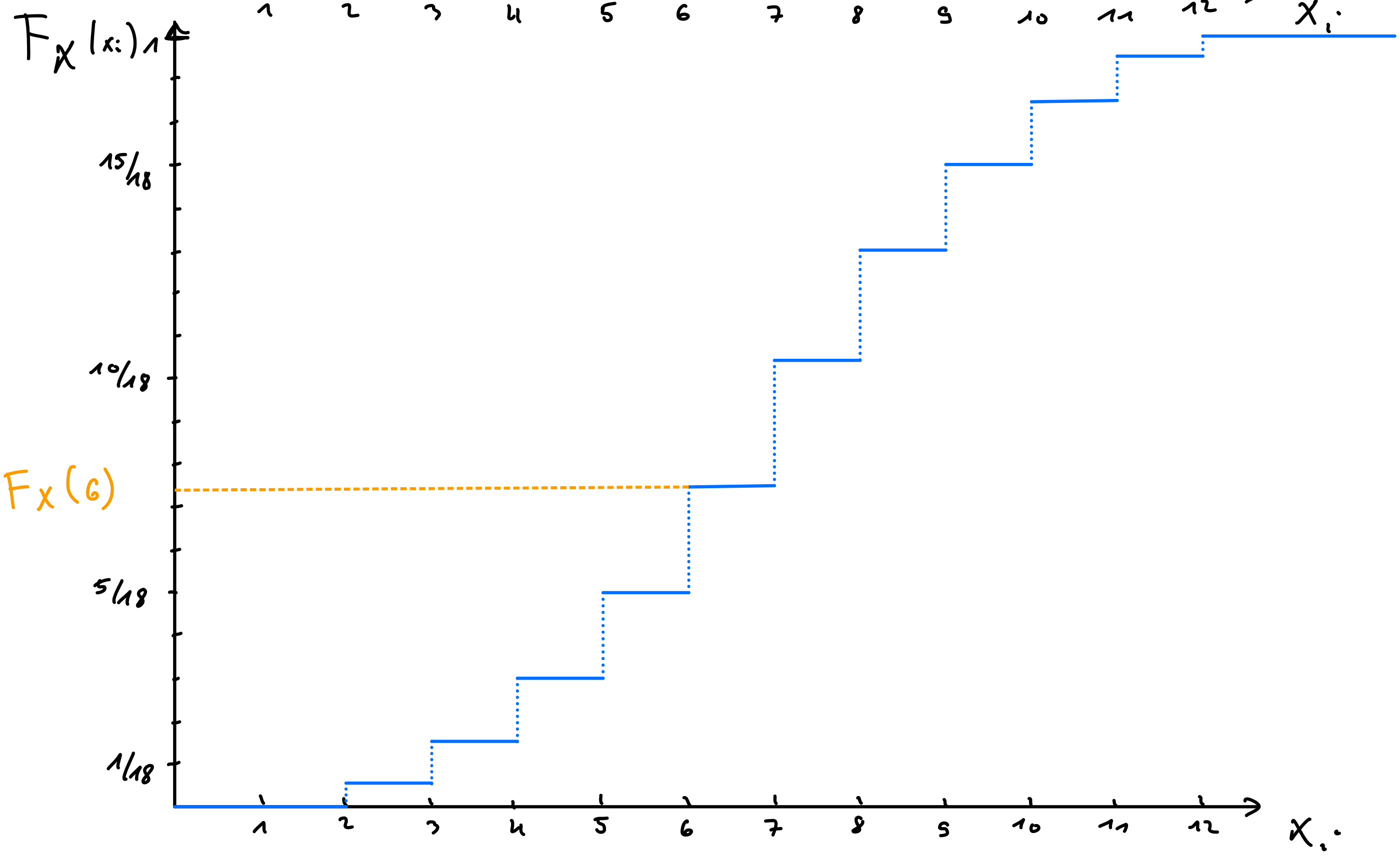
$\vdots$

$x_6 = 1+6 = 2+5 = 3+4 = 4+3 = 5+2 = 6+1 = 7$, $f_X(x_6) = 6/36$

$\vdots$

$x_{11} = 6+6 = 12$, $f_X(x_{11}) = 1/36$





Npr. če nas zanima verjetnost, da je vsota pik ≤ 6

$$P(X \leq 6) = F_X(6) = 15/36 = 5/12$$
$$= P(X=2) + \dots + P(X=6)$$

$$P(X > 6) = 1 - P(X \leq 6) = 7/12$$

$$P(X \leq 1) = 0$$
$$P(X \leq 12, 13, \dots) = 1$$
$$P(X > 12, 13, \dots) = 0$$

Zvezne porazdelitve

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

zvezna naključna spremenljivka

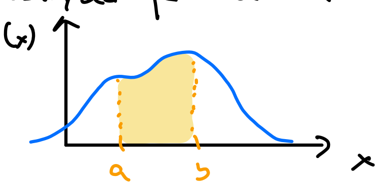
gostota verjetnostne porazdelitve / verjetnostna gostota / PDF prob. density func.

Lastnosti:

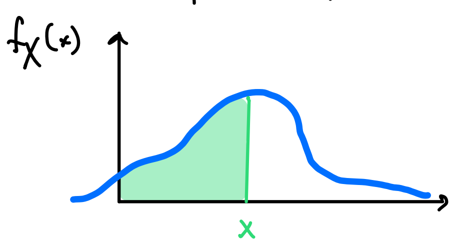
- $f_X(x) \ge 0$
- $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ normalizacija
- $f_X(x) = \dfrac{dF_X(x)}{dx}$

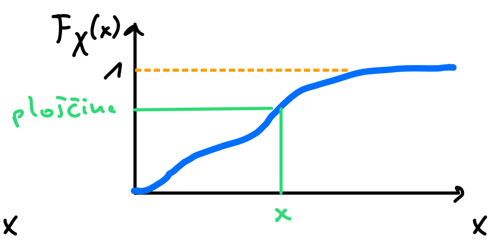
Smiselno se je vprašati po verjetnosti, da je $a \le X \le b$

$$P(a \le X \le b) = \int_a^b f_X(x)\,dx = F_X(b) - F_X(a)$$



Vprašanje: kot je $P(X=a)$ nima smisla!





Zgled Naključna sprem. X je porazdeljena v skladu z verjetnostno gostoto

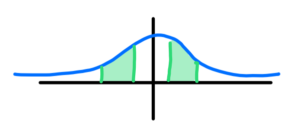
$$f_X(x) = \frac{c}{1+x^2}, \quad -\infty < x < \infty$$

Določi verjetnost, da X^2 leži med $1/3$ in 1

Porazdelitev mora biti normirana

$$\underbrace{\int_{-\infty}^{\infty}}_{\text{ali } D_f} f_X(x) = 1 = \int_{-\infty}^{\infty} \frac{c}{1+x^2}\,dx = c \arctan x \Big]_{-\infty}^{\infty} = c\left(\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right) = c\pi \Rightarrow c = \frac{1}{\pi}$$

Pravilno normirana verjetnostna gostota je $f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}$



$$1/3 \le X^2 \le 1 \Rightarrow \frac{1}{\sqrt{3}} \le X \le 1 \text{ ali } -1 \le X \le -\frac{1}{\sqrt{3}}$$

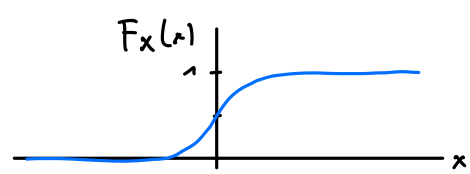
} to sta izključujoči dogodke

$$P(1/3 \le X^2 \le 1) = P\left(\frac{1}{\sqrt{3}} \le X \le 1\right) + P\left(-1 \le X \le -\frac{1}{\sqrt{3}}\right)$$

$$= \frac{1}{\pi}\int_{1/\sqrt{3}}^{1} \frac{dx}{1+x^2} + \frac{1}{\pi}\int_{-1}^{-1/\sqrt{3}} \frac{dx}{1+x^2} = \frac{1}{\pi}\left(\arctan 1 - \arctan\frac{1}{\sqrt{3}} + \arctan -\frac{1}{\sqrt{3}} - \arctan -1\right) = \frac{1}{6}$$

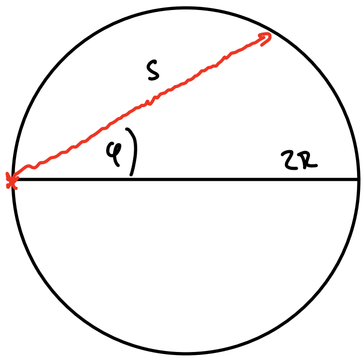
Kakšna je (kumulativna) porazdelitvena funkcija?

$$F_X(x) = \frac{1}{\pi}\int_{-\infty}^{x} \frac{dx}{1+x^2} = \frac{1}{\pi}\left(\arctan x + \frac{\pi}{2}\right) = \frac{1}{\pi}\arctan x + \frac{1}{2}$$



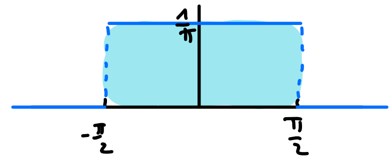
Transformacija neodvisne spremenljivke

Zgled



Z oboda kroga naključno vlečemo tetive (s). Zanima nas porazdelitev po dolžini tetiv

$$\frac{dF_\phi}{d\varphi} = f_\phi(\varphi) = \frac{1}{\pi}$$



$$\frac{dF_s}{ds} = \frac{dF_\phi}{d\varphi}\left|\frac{d\varphi}{ds}\right|$$

$$= \frac{1}{\pi}\frac{1}{\sqrt{(2R)^2 - s^2}}$$

Zgornji in spodnji del sta enaka zato smo se zmotili za 2 !

Preslikava je enolična le na 0 do $\pi/2$

$$S(\varphi) = 2R\cos\varphi$$

$$\varphi(s) = \arccos\left(s/(2R)\right)$$

$$\frac{d\varphi}{ds} = -\frac{1}{\sqrt{(2R)^2 - s^2}}$$

$$\int_{-\pi/2}^{\pi/2} f_s(x)\,dx = \frac{1}{2}$$

Kaj je šlo narobe?

Bolj formalno: Naj ima naključna spremenljivka X verj. gostoto $f_X(x)$. Zanima nas verjetnostna gostota spremenljivke $Y = h(X)$, kjer je h bijektivna.

naključna ⇐ naključna

$$F_Y(y) = P(Y \le y) = P(h(X) \le y) \overset{!}{=} P(X \le h^{-1}(y)) = \int_{-\infty}^{h^{-1}(y)} f_X(x)\,dx =$$

bijektivnost

$z = h(x)$ $\quad x = h^{-1}(z)$

$dx = \left|\frac{d}{dz}h^{-1}(z)\right| dz$

$$= \int_{-\infty}^{y} f_X(h^{-1}(z))\left|\frac{d}{dz}h^{-1}(z)\right| dz$$

$$\boxed{f_Y(y) = \frac{dF_Y(y)}{dy} = f_X(h^{-1}(y))\left|\frac{d}{dz}h^{-1}(y)\right|}$$

V prejšnjih oznakah

$$f_s(s) = f_\varphi(\arccos(s/2R))\left|\frac{d}{ds}\varphi(s)\right|$$

$\frac{1}{\pi}$

$s = 2R\cos\varphi = h(\varphi)$ $\qquad \varphi = h^{-1}(s)$

$\varphi = \arccos\frac{s}{2R}$

$$f_s = \underbrace{\frac{1}{\pi}}_{[0,\pi/2]}\left|\frac{d}{ds}\underbrace{\left(\arccos\frac{s}{2R}\right)}_{\text{inverz na } [0,\pi/2]}\right| + \frac{1}{\pi}\left|\frac{d}{ds}\underbrace{\left(-\arccos\frac{s}{2R}\right)}_{\text{inverz na } [-\pi/2,0]}\right|$$

$$= \frac{2}{\pi}\frac{1}{\sqrt{(2R)^2 - s^2}}$$ ✓✓

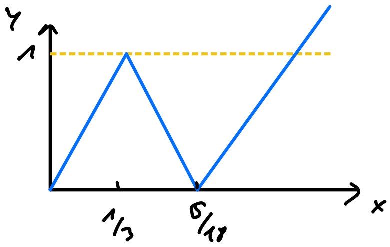
Splošneje, če inverz h^{-1} ni enoličen, velja

$$f_Y(y) = \sum_{x \in h^{-1}(\{y\})} f_X(g(y))\,|g'(y)|$$

$g(y)$ je prejšnji $h^{-1}(y)$, tod tak, da je vedno bijektiven

Zgled Naj ima X eksponentno porazdelitev: $f_X(x)=e^{-x}$ $(x\ge 0)$ in naj bo $Y=h(x)$ kjer je

$$h(x)=\begin{cases}3x & 0<x\le 1/3\\ 1-5(x-\frac{1}{3}) & \frac{1}{3}<x\le 8/15\\ 2(x-8/15) & x>8/15\end{cases}$$



Izračunaj $f_Y(y)$

1. segment $y=3x \rightarrow x=\frac{1}{3}y=g_1(y)$ inverz h^{-1} na 1. segmentu

2. segment $y=1-5(x-\frac{1}{3}) \rightarrow x=-\frac{1}{5}y+\frac{8}{15}=g_2(y)$

3. segment $x=\frac{1}{2}y+\frac{8}{15}=g_3(y)$

$$f_Y(y)=f_X\left(\frac{1}{3}y\right)\left|\frac{d}{dy}\left(\frac{1}{3}y\right)\right|+f_X\left(-\frac{1}{5}y+\frac{8}{15}\right)\left|\frac{d}{dy}\left(-\frac{1}{5}y+\frac{8}{15}\right)\right|+f_X\left(1/2\,y+8/15\right)\left|\frac{d}{dy}\left(1/2\,y+8/15\right)\right|$$

$$f_Y(y)=\frac{1}{3}e^{-\frac{1}{3}y}+\frac{1}{5}e^{\frac{1}{5}y-8/15}+\frac{1}{2}e^{-1/2\,y-8/15}$$

Pogojna verjetnost

A in B naj bosta naključna dogodka, naj bo $P(A)>0$. Zanima nas verjetnost, da se zgodi B, če vemo, da se je že zgodil A ("pogoj"). To označimo kot $P(B\mid A)$

$$\boxed{P(A\cap B)=P(B\mid A)\,P(A) \qquad P(B\mid A)=\frac{P(A\cap B)}{P(A)}}$$

Zgled Vržemo kocko (1x), kolikšna je verjetnost, da bo št. pik <4 (dog. B)?

a) $P(B)=P(1)+P(2)+P(3)=3\cdot\frac{1}{6}=1/2$

Kolikšna pa je verjetnost, da bo št. pik <4 ob pogoju, da bo št. pik liho?

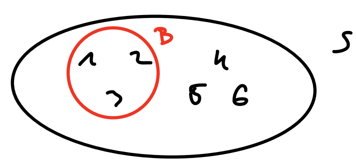
"pogoj" = dogodek A = # pik liho

$P(A)=1/2$

b) $P(B\mid A)=\frac{P(A\cap B)}{P(A)}=\frac{2\cdot 1/6}{1/2}=\frac{2}{3}$ ← dodatna informacija (dogodek A) poviša verjetnost iz $1/2$ na $2/3$

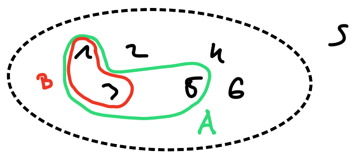
a) Vzorčni prostor $S=\{1,2,3,4,5,6\}$

B: # pik <4



b) A: # pik liho

to skrči vzorčni prostor na $\{1,3,5\}$



Neodvisni dogodki (independent events)

B je neodvisen od A, če velja $P(B|A) = P(B)$ oziroma $P(B \cap A) = P(A)P(B)$

Trije dogodki so neodvisni, če so neodvisni po parih: za A_1, A_2, A_3 mora veljati

$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$
$$P(A_1 \cap A_3) = P(A_1)P(A_3)$$
$$P(A_2 \cap A_3) = P(A_2)P(A_3)$$

in $$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

in ustrezne posplošitve za več dogodkov

To lahko uporabimo pri definiciji verjetnostnih porazdelitev več spremenljivk $\longrightarrow$ skupne (oz. kombinirane / ang. joint) verjetnostne porazdelitve.

Diskretne

X, Y ... diskretni naključni spremenljivki

$P(X=x, Y=y) = f_{X,Y}(x,y)$ skupna verjetnostna funkcija

z lastnostmi
$$f_{X,Y}(x,y) \geq 0$$
$$\sum_{x,y} f_{X,Y}(x,y) = 1$$

Naj X zavzame vrednosti $\{x_1, x_2, \dots, x_m\}$
Y $\{y_1, y_2, \dots, y_n\}$

$$P(X=x_i, Y=y_j) = f_{X,Y}(x_i, y_j)$$

$$P(X=x_i) = f_X(x_i) = \sum_{j=1}^{n} f_{X,Y}(x_i, y_j)$$

$$P(Y=y_j) = \sum_{i=1}^{m} f_{X,Y}(x_i, y_j)$$

če vse pravilno naredimo $\sum_{i,j} f_{X,Y}(x_i, y_j) = \sum_i f_X(x_i) = \sum_j f_Y(y_j) = 1$

Skupna porazdelitvena funkcija

$$P(X \leq x, Y \leq y) = F_{X,Y}(x,y) = \sum_{u \leq x} \sum_{v \leq y} f_{X,Y}(u,v)$$

Če sta dogodka $X=x$ in $Y=y$ neodvisna za $\forall x$ in $\forall y$, sta X in Y neodvisni naključni spremenljivki, in velja

$$P(X=x, Y=y) = P(X=x)\,P(Y=y)$$

oz. $$f_{X,Y}(x,y) = f_X(x)\,f_Y(y)$$

Zvezne

po analogiji: vpeljemo skupno verjetnostno gostoto (joint prob. density) ... $f_{X,Y}(x,y)$

z lastnostmi

$$f_{X,Y}(x,y) \ge 0 \quad \text{in} \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$$

Verjetnost, da je $a \le X \le b$ in $c \le Y \le d$

$$P(a \le X \le b,\ c \le Y \le d) = \int_a^b \int_c^d f_{X,Y}(x,y)\,dy\,dx$$

Skupna porazdelitvena funkcija je

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

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

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

izpolnjene morajo biti vse normalizacije

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,v)\,dv \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(u,y)$$

Včasih jima rečemo marginalni (robni) verjetnostni gostoti

$$\int f_X(x)\,dx = \int f_Y(y)\,dy = \iint f_{X,Y}(x,y)\,dx\,dy = 1$$

Zvezni naključni spremenljivki X in Y sta neodvisni, kadar so dogodki $X \le x$ in $Y \le y$ neodvisni za vse x in y

$$P(X \le x,\ Y \le y) = P(X \le x)\,P(Y \le y)$$

in torej tudi za gostote velja $f_{X,Y}(x,y) = f_X(x)\,f_Y(y)$

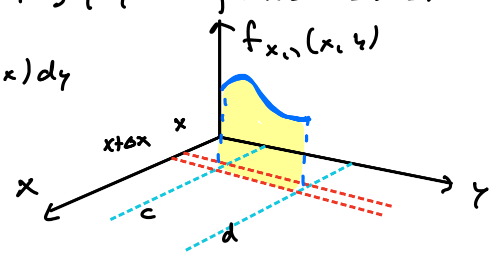
Dodajmo še pojem pogojne verjetnosti za skupne porazdelitve
X, Y ... naključni, def. pogojno verjetnostno gostoto, da se je zgodilo "$Y=y$" pri čemer obstaja pogoj / dodatna informacija, da je tudi "$X=x$" (narekovaji zato, ker je vse mišljeno v smislu gostot)

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \qquad P(B|A) = \frac{P(A\cap B)}{P(A)} \ \text{za diskretne}$$

in analogno
$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Uporaba:
- ko iščemo verjetnost, da je $c \le Y \le d$ ob pogoju, da je $x \le X \le x+\Delta x$

$$P(c \le Y \le d \mid x \le X \le x+\Delta x) = \int_c^d f_{Y|X}(y|x)\,dy$$



Transformacija spremenljivk "2D"

$X, Y \to U, V$

1D $\quad X \dots f_X(x) \qquad U = \phi(X) \qquad X = \psi(U) = \phi^{-1}(U)$ bijektivna

$$\Rightarrow \quad f_U(u) = f_X(\psi(u))\,|\psi'(u)|$$

2D $\quad X, Y \dots f_{X,Y}(x,y), \quad U = \phi_1(x,y) \quad V = \phi_2(x,y) \quad X = \psi_1(U,V) \quad Y = \psi_2(U,V)$ bijektivni

$$\Rightarrow \quad f_{U,V}(u,v) = f_{X,Y}(\psi_1(u,v), \psi_2(u,v))\,|\det J|$$

$$|\det J| = \left\| \begin{matrix} \frac{\partial \psi_1(u,v)}{\partial u} & \frac{\partial \psi_1(u,v)}{\partial v} \\ \frac{\partial \psi_2(u,v)}{\partial u} & \frac{\partial \psi_2(u,v)}{\partial v} \end{matrix} \right\|$$

Zgled $\quad X, Y$ (zvezni) s skupno ver. gostoto

$$f_{X,Y}(x,y) = \begin{cases} xy/96 \ ; & 0<x<4,\ 1<y<5 \\ 0 \ ; & \text{sicer} \end{cases}$$

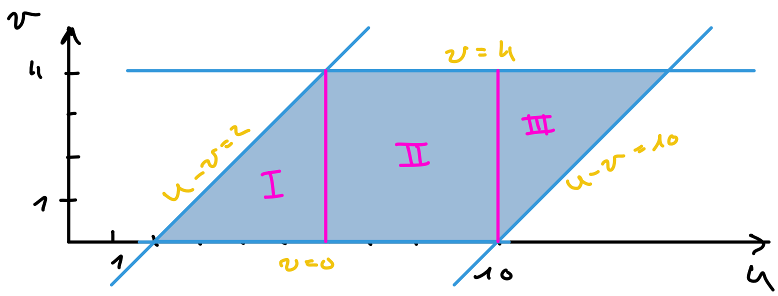
Poišči verjetnostno gostoto, ki ustreza $U = X + 2Y = \phi_1(X,Y)$

Izmislimo si še $V = X$ (npr.) ne sme biti konst ker je $\Rightarrow \det J = 0$

$\qquad\qquad\qquad \text{"} \phi_2(X,Y)$

Inverzi $\quad X = V = \psi_2(U,V)$

$\qquad\qquad Y = \frac{1}{2}(U - X) = \frac{1}{2}(U - V) = \psi_1(U,V)$



Ref. območje

$0 < x < 4, \ 1 < y < 5$

$\downarrow$

$0 < v < 4, \ 2 < u - v < 10 \quad ?$

$$f_{U,V}(u,v) = f_{X,Y}(\psi_1(u,v), \psi_2(u,v))\,|\det J|$$
$$= f_{X,Y}\left(v, \tfrac{1}{2}(u-v)\right) \tfrac{1}{2} \qquad |\det J| = \begin{vmatrix} 0 & 1 \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}$$

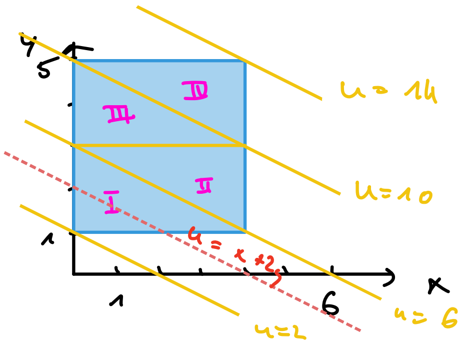
$$= \frac{v(u-v)}{2} \, \frac{1}{96} \cdot \frac{1}{2} = \begin{cases} v(u-v)/384 & 0<v<4,\ 2<u-v<10 \\ 0 & ;\ \text{sicer} \end{cases}$$

$f_U(u) = \int f_{U,V}(u,v)\,dv \quad \dots$ jasno, ampak težave z mejami

$$f_U(u) \begin{cases} \int_0^{u-2} f_{U,V}(u,v)\,dv = \frac{(u-2)^2(u+4)}{2304} & ;\ 2<u<6 \quad \mathrm{I} \\ \int_0^{4} f_{U,V}(u,v)\,dv = \frac{3u-8}{144} & ;\ 6<u<10 \quad \mathrm{II} \\ \int_{u-10}^{4} f_{U,V}(u,v)\,dv = \frac{348u - u^3 - 2128}{2304} & ;\ 10<u<14 \quad \mathrm{III} \end{cases}$$

Alternativni pristop: Lahko bi se takoj lotili porazdelitvene funkcije za u, in jo odvajam

$$P(X+2Y \le u) = F_U(u) = \iint\limits_{x+2y\le u} f_{X,Y}(x,y)\,dx\,dy = \iint\limits_{\substack{0<x<4\\ 1<y<5\\ x+2y\le u}} \frac{xy}{96}\,dx\,dy$$



$$\mathrm{I} \quad \int_{x=0}^{u-2}\int_{y=1}^{\frac{u}{2}-\frac{x}{2}} \frac{xy}{96}\,dy\,dx = \int_0^{u-2} \frac{x(u-x)^2}{768} - \frac{x}{192}\,dx = I(u)$$

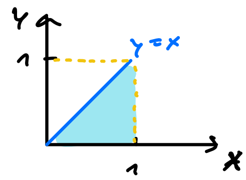
$$\frac{d\,I(u)}{du} = \frac{(u-2)^2(u+4)}{2304}$$

Zgled Skupna verjetnostna gostota X in Y naj bo

$$f_{X,Y}(x,y) = \begin{cases} 8xy & : 0\le x\le 1,\ 0\le y\le x \\ 0 & : \text{sicer} \end{cases}$$

← že iz omejitve območja vemo da sta sprem. odvisni.

Zanimajo nas f_X, f_Y, $f_{X|Y}$, $f_{Y|X}$



$$f_X(x) = \int_{y=0}^{x} f_{X,Y}(x,y)\,dy = \int_0^x 8xy\,dy = 4x^3 \quad \text{za } 0<x<1,\ 0 \text{ sicer}$$

$$f_Y(y) = \int_{x=y}^{1} f_{X,Y}(x,y)\,dx = \int_y^1 8xy\,dx = 4y(1-y^2) \quad \text{za } 0<y<1,\ 0 \text{ sicer}$$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{8xy}{4y(1-y^2)} = \frac{2x}{1-y^2} \quad \text{za } 0\le y\le x\le 1,\ 0 \text{ sicer}$$

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{8xy}{4x^3} = \frac{2y}{x^2} \quad \text{za } 0\le y\le x\le 1,\ 0 \text{ sicer}$$

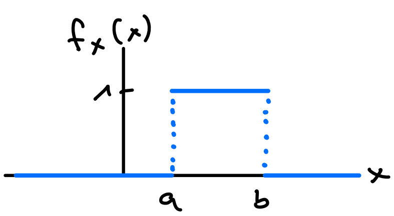
Ali sta slučajni spremenljivki neodvisni?

V (senčenem) def. območju imamo $f_{X,Y}(x,y) = 8xy$ toda: $f_X(x) = 4x^3$ $f_Y(y) = 4y(1-y^2)$

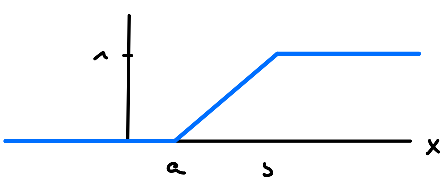
$$f_{X,Y}(x,y) \neq f_X(x)\cdot f_Y(y) \Rightarrow X \text{ in } Y \text{ nista neodvisni} \ !$$

Posebne zvezne verjetnostne porazdelitve

Enakomerna



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



$$P(X\le x) = F_X(x) = \begin{cases} 0 & x<a \\ \frac{x-a}{b-a} & a\le x\le b \\ 1 & x\ge b \end{cases}$$

uniform

Oznaka $X \sim U(a,b)$ ali $\mathcal{I}(a,b)$

porazdeljena po

Eksponentna

- npr. razpad radioaktivnih jeder : naključen proces
- verjetnost za razpad v nekem časovnem intervalu je odvisna samo od dolžine tega intervala ne pa od starosti jedra "Jedra se ne starajo"
- če je Δt dovolj majhen, je verjetnost za razpad sorazmerna z Δt.

$$P(\text{jedro razpadlo}) = \lambda \Delta t$$

razpadna konstanta ($\lambda = \frac{1}{\tau}$ τ ... raz. čas)

$$P(\text{jedro ni razpadlo}) = 1 - \lambda \Delta t$$

$$P(\text{jedro ni razpadlo v 2}) = (1 - \lambda \Delta t)^2$$

$$\vdots$$

Verjetnost, da jedro ni razpadlo po dolgem času $t = k \Delta t$ je

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

ko čas teče je vedno manj verjetno, da jedro še ni razpadlo

Namesto τ lahko uporabljamo $t_{1/2}$ ... razpolovni čas

$$\frac{N(t_{1/2})}{N_0} = \frac{\frac{1}{2} N_0}{N_0} = \frac{1}{2} = e^{-\frac{t_{1/2}}{\tau}} \qquad t_{1/2} = \tau \log 2$$

Eksponentna porazdelitev

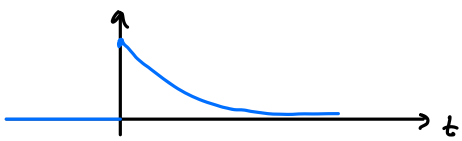
$P(t) = e^{-\lambda t}$ $\quad \lambda = 1/\tau$ $\quad$ verjetnost, da do t še ni razpadlo

Kako iz tega dobimo verjetnostno gostoto? (verjetnost na $[t, t+\Delta t)$)

Začetno vrednost poznamo $f_T(0) = \lambda$ = verjetnost za razpad / časovno enoto

Ker se jedra ne starajo, se verjetnost, da do t preživi in potem razpademo, ker faktorizira:

$$f_T(t) = P(t)\lambda = \lambda e^{-\lambda t} \qquad t \geq 0$$



Verjetnost, da jedro do časa t še ni razpadlo

$$P(t) = \int_t^\infty f_T(t)dt = \int_t^\infty \lambda e^{-\lambda t} dt = e^{-\lambda t}$$

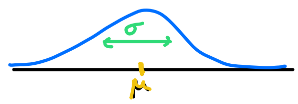
Kako preverimo, da eksp. porazdelitev "nima spomina"

$$P(T > t + \Delta t \mid T > t) = \frac{P(T > t + \Delta t \cap T > t)}{P(T > t)} = \frac{P(T > t + \Delta t)}{P(T > t)} = \frac{e^{-\lambda(t+\Delta t)}}{e^{-\lambda t}} = e^{-\lambda \Delta t}$$

rezultat neodvisen od t

Normalna porazdelitev

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad x \in (-\infty, \infty)$$



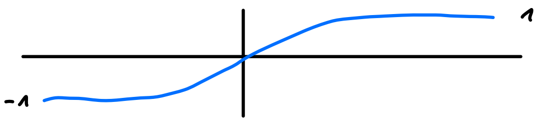
oznaka $X \sim N(\mu, \sigma^2)$, tudi $\sim N(0,1)$ standardna normalna porazdelitev

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(t)\,dt = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x-\mu}{\sqrt{2}\,\sigma}\right)\right)$$

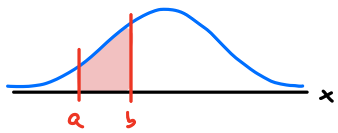
$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\,dt$$

error function

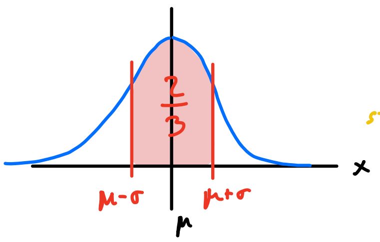
erf (z)



$$\mathrm{erfc}(z) = 1 - \mathrm{erf}(z)$$



$$P(a \le X \le b) = \frac{1}{2}\left(\mathrm{erf}\left(\frac{b-\mu}{\sqrt{2}\,\sigma}\right) - \mathrm{erf}\left(\frac{a-\mu}{\sqrt{2}\,\sigma}\right)\right)$$



$$z = \frac{X - \mu}{\sigma}$$

običajno

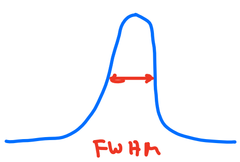
standardizirano

$$P(\mu - \sigma \le X \le \mu + \sigma) = P(-1 \le z \le 1) = \mathrm{erf}\left(\tfrac{1}{\sqrt{2}}\right) \doteq 0{,}683 = \frac{2}{3}$$

$$P(\mu - 2\sigma \le X \le \mu + 2\sigma) = \mathrm{erf}\left(\tfrac{2}{\sqrt{2}}\right) \doteq 0{,}955$$

$$P(\mu - 3\sigma \le X \le \mu + 3\sigma) = \mathrm{erf}\left(\tfrac{3}{\sqrt{2}}\right) \doteq 0{,}997$$

FWHM (full width at half maximum)



$$\frac{\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi}} \cdot 1} = \frac{1}{2}$$

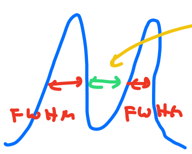
vrh

$$e^{-\frac{z^2}{2\sigma^2}} = \frac{1}{2}$$

$$-\frac{z^2}{2\sigma^2} = -\ln 2$$

$$z = \sigma\sqrt{2\ln 2}$$

$$\mathrm{FWHM} = 2\sqrt{2\ln 2}\,\sigma \doteq 2{,}35\,\sigma$$



če je tok ≥ FWHM lahko hribčka ločimo

Maxwellova porazdelitev

- dobimo jo iz normalnih porazdelitev za tri naključno neodvisne komponente vektorja v 3d, npr. hitrosti plinskih molekul $\vec{v} = (v_x, v_y, v_z)$.
- v termodinamičnem ravnovesju je

$$f_{\vec{v}}(\vec{v}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{3}{2}} e^{-\frac{v_x^2+v_y^2+v_z^2}{2\sigma^2}} = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{3}{2}} e^{-\frac{v^2}{2\sigma^2}}$$

$$-\frac{v^2}{2\sigma^2} = -\frac{mv^2}{2k_BT} = -\frac{W_k}{W_T}$$

kinetična energija

termična energija

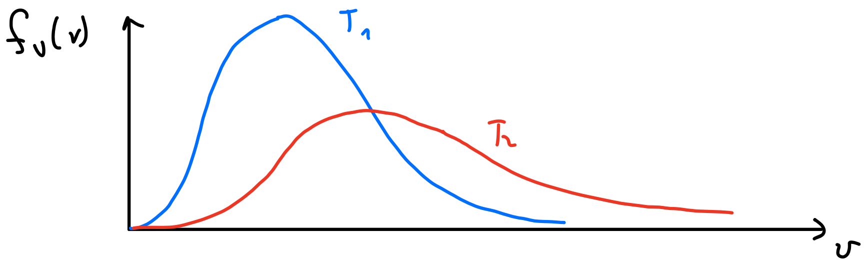
$\sigma^2 = \frac{k_BT}{m}$

- ker imamo krogelno simetrijo, je diferencial verjetnosti po velikosti vektorja hitrosti

$$dP = \left(\frac{m}{2\pi k_BT}\right)^{\frac{3}{2}} e^{-\frac{mv^2}{2k_BT}} 4\pi v^2 \, dv$$

$$\boxed{\frac{dP}{dv} = f_v(v) = \left(\frac{m}{2\pi k_BT}\right)^{\frac{3}{2}} 4\pi v^2 e^{-\frac{mv^2}{2k_BT}}}$$

(opomba $\bar{\vec{v}} = \vec{0}$)



Pričakovane vrednosti (porazdelitve oz. spremenljivke)

Diskretna naključna spremenljivka X:

naključna

$$\bar{X} = E[X] = \sum_{i=1}^{n} x_i \underbrace{P(X=x_i)}_{f_X(x_i)}$$ Povprečje

"expected"

Če so vsi izidi enako verjetni (enak $P(X=x_i)$ za $\forall x_i$), je $P(X=x_i) = \frac{1}{n}$

$$\bar{X} = E[X] = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Zvezne spremenljivke:

$$E[X] = \int_{-\infty}^{\infty} x \, f_X(x) \, dx$$

Ti dve pričakovani vrednosti ste povprečji.

Zgled Kazino, igre s kocko,
2 = dobimo 10 €
4 = dobimo 30 €
6 = plačamo 20 €
1,3,5 = 0

$$E[X] = \frac{1}{6} \cdot 0 + \frac{1}{6} \cdot 10 + \frac{1}{6} \cdot 0 + \frac{1}{6} \cdot 30 + \frac{1}{6} \cdot 0 + \frac{1}{6}(-20) = 3{,}67 \text{ €}$$

Funkcije naključnih spremenljivk

Diskretno

$$E[g(x)] = \sum_{i=1}^{n} g(x_i)\, f_X(x_i)$$

Zvezno

$$E[g(x)] = \int_{-\infty}^{\infty} g(x_i)\, f_X(x_i)\, dx = \overline{g(x)}$$

Pri Mod Fiz 1

operator (deluje to na ψ_i)

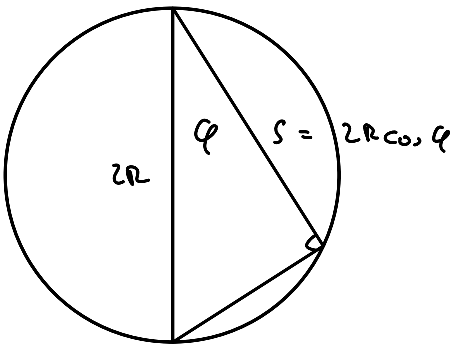
$$\bar{\sigma} = \int \psi_f^* \,\hat{\sigma}\, \psi_i \, dV$$

zač. stanje

končno stanje

$$\bar{\sigma} = \int \underbrace{|\psi(\vec{r})|^2}_{\varphi(\vec{r})\ \ldots\ f_X(x)} \sigma(\vec{r})\, dV$$

Zgled



$$f_\Phi(\varphi) = \frac{dF_\Phi}{d\varphi} = \frac{1}{\pi}$$

$$E[s] = \bar{s} = \int_{-\pi/2}^{\pi/2} s(\varphi)\, dF_\Phi = \int_{-\pi/2}^{\pi/2} s(\varphi)\, f_\Phi(\varphi)\, d\varphi$$

$$= \int_{-\frac{\pi}{2}}^{\pi/2} 2R\cos\varphi \,\frac{1}{\pi}\, d\varphi = \frac{4R}{\pi}$$ povprečna dolžina tetive

Če vzamemo kot neodvisno spremenljivko kar s namesto φ bi morali dobiti enak rezultat.

$$E[s] = \int_0^{2R} s\, f_S(s)\, ds = \int_0^{2R} s\, \frac{2}{\pi\sqrt{(2R)^2 - s^2}}\, ds = \ldots = \frac{4R}{\pi}$$

Variance in standardna deviacija

$$\mathrm{var}[X] = E\left[(X - E[X])^2\right] = \overline{(X - \bar{X})^2}$$ varianca

$$\sigma_X = \sqrt{\mathrm{var}[X]}$$ standardna deviacija, std. odklon, efektivni odmik, ...

Za diskretno X:

$$\mathrm{var}[x] = \sum_{i=1}^{n} (x_i - \bar{x})^2 f_X(x_i)$$

Poseben primer $f(x_i) = f(x_j)$ za vse i, j

$$\mathrm{var}[x] = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

moment je vse tipa $\int (\ )^n dx$

odštejemo centralno vrednost $\bar{x}$

Za zvezno X:

$$\mathrm{var}[x] = \int_{-\infty}^{\infty} (x - \bar{x})^2 dx$$

Centralni moment porazdelitve

Vse to nekakšna mera za "raztresenost", "disperzijo"

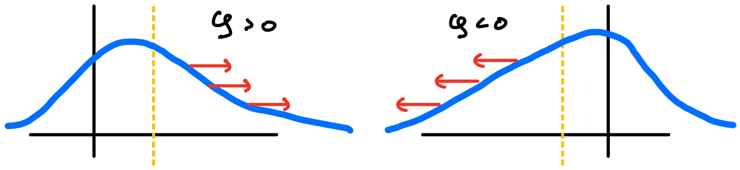
Splošen n-ti moment $M_n = \overline{(x - \bar{x})^n} = \int_{-\infty}^{\infty} (x - \bar{x})^n f_X(x)\, dx$

$\Rightarrow$ $\bar{X} = M_1$

$\mathrm{var}[X] = M_2$

uporabljajo se še M_3 in M_4 v praksi

$\boxed{\gamma = \dfrac{M_3}{\sigma_x^3}}$ poševnost (skewness)

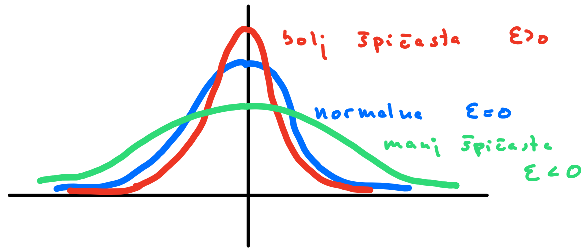


$\boxed{\varepsilon = \dfrac{M_4}{\sigma_x^4} - 3}$ eksces (presežek, excess)

↓

kurtosis

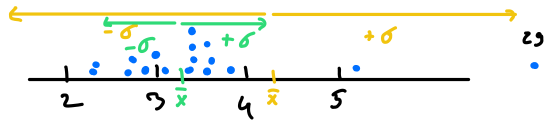
↳ za normalno por $\dfrac{M^4}{\sigma_x^4} = 3$



Zgled 24 izmerkov vsebnosti Cu v moki:

$x_i = \{2.20,\ 2.20,\ 2.40,\ 2.40,\ \dots,\ \underbrace{5.28,\ 28.95}_{\text{zelo odstopata}}\}$

to sta "ubežnika" (outlier) oz "osamelca"



za $n = 24$	$n = 23$ (brez zadnjega)
$\bar{x} = 4,28$	$\bar{x} = 3,21$
$\sigma_x = 5,30$	$\sigma_x = 0,69$

$\Rightarrow$ en sam ubežnik popolnoma spremeni prič. končne vrednosti

$\Rightarrow$ potrebujemo robustnejše mere za povprečje in disperzijo

Mediana

Robustnejša mera za "težišče" vzorca je MEDIANA

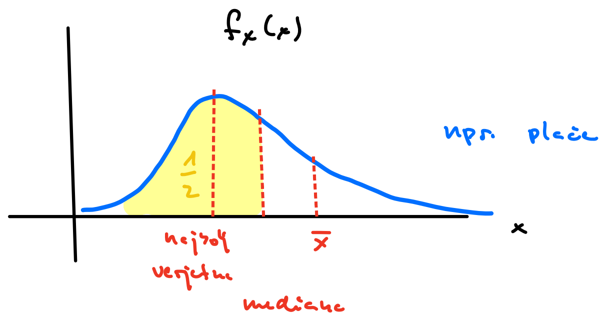
$X = \{x_i\}_{i=1}^{n}$, urejeni $x_i \le x_{i+1}$

$$\mathrm{med}[x] = \begin{cases} X_{\frac{n+1}{2}} & ; \ n \text{ lih} \\ \frac{1}{2}\left(X_{\frac{n}{2}} + X_{\frac{n}{2}+1}\right) & ; \ n \text{ sod} \end{cases}$$

Za zvezne porazdelitve

mediana je tisti x, za katerega

$$P(X < x) = P(X > x) = 1/2$$



Tudi za razpršenost obstaja robustnejša mera: za diskretno porazdelitev

$$\mathrm{MAD}[x] = \mathrm{med}\left[\,\left|X - \mathbb{1}_n \mathrm{med}[x]\right|\,\right]$$

median absolute deviation

$$\mathrm{MAD}[x] = \mathrm{med}\left[\,\left|\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} - \begin{bmatrix} \mathrm{med} \\ \mathrm{med} \\ \vdots \\ \mathrm{med} \end{bmatrix}\right|\,\right]$$

Na vzorcu moke $n = 24$

$\mathrm{med}[x] = 3,385$

$\mathrm{MAD}[x] = 0,355$

z izločeno 24. meritvijo

$\mathrm{med}[x] = 3,37$

$\mathrm{MAD}[x] = 0,340$

Običajno podajamo $MADN[X] \doteq 1{,}4826\, MAD[X]$

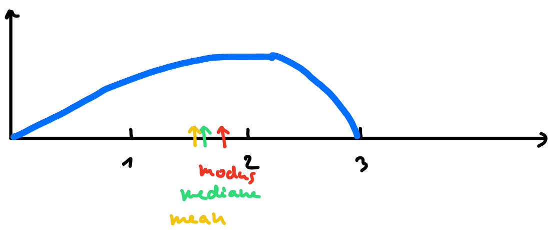
Izbran tako, da za normalno porazdelitev $MADN[X] = \sigma_x$

def. MAD za zvezno porazdelitev $P(|X-\bar{X}| \leq MAD) = \frac{1}{2}$

Zgled zvezna sprem. X, ustrezna verjet. gostota

$$f_X(x) = \begin{cases} \frac{4x(9-x^2)}{81} & ; \ 0 \leq x \leq 3 \\ 0 & ; \ \text{sicer} \end{cases}$$

Poišči "mode" (= maximum) modus, mediano, povprečje.



① Modus $\frac{df_x}{dx} = 0 = \frac{36-12x^2}{81} \Rightarrow x = \sqrt{3} = 1{,}73$

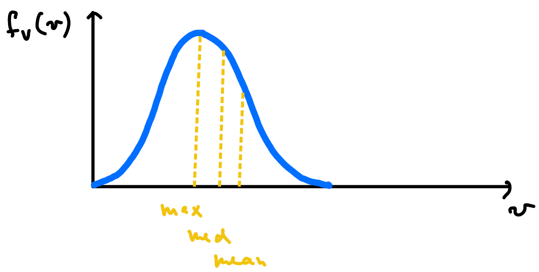
② Mediana $P(X \leq \overset{a}{\overset{\parallel}{med}}) = \frac{1}{2}$

$$\frac{1}{2} = \frac{4}{81}\int_0^a x(9-x^2)\,dx = \frac{4}{81}\left(\frac{9a^2}{2} - \frac{a^4}{4}\right) \Rightarrow a^2 = 9 \pm \frac{9}{2}\sqrt{2} \qquad a = 1{,}62$$

③ "Mean" (povprečje)

$$\bar{X} = \frac{4}{81}\int_0^3 x \cdot x(9-x^2)\,dx = \frac{4}{81}\left(\frac{9\,3^3}{3} - \frac{3^5}{5}\right) \doteq 1{,}60$$

Zgled Maxwell – Boltzmannova porazdelitev



$$f_V(v) = \left(\frac{m}{2\pi k_0 T}\right)^{\frac{3}{2}} 4\pi v^2 e^{-mv^2/2k_0T} \qquad m = \frac{M}{N_A}$$

$$a = \sqrt{\frac{k_0 T}{m}} \Rightarrow \frac{mv^2}{2k_0T} = \frac{x^2}{2a^2}$$

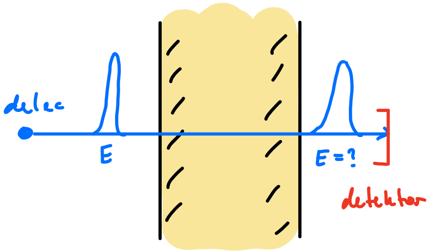
$$f_X(x) = \sqrt{\frac{2}{\pi}}\,\frac{x^2 e^{-x^2/2a^2}}{a^3}$$

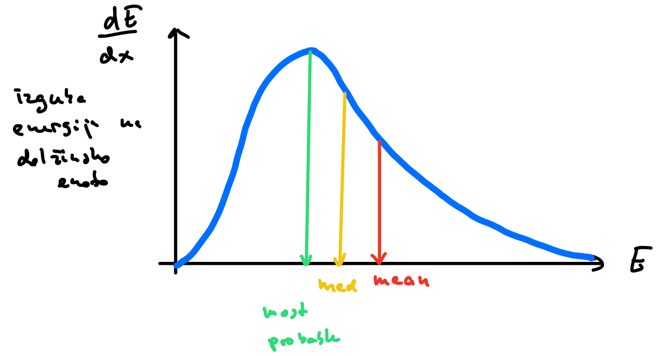
"RMS" root mean square

$$v_{max} = \sqrt{\frac{2k_0T}{m}} < \bar{v} = \sqrt{\frac{8k_0T}{\pi m}} < \sqrt{\overline{v^2}} = \sqrt{\frac{3k_0T}{m}}$$

N_2 pri $T = 300\,K$ $\quad = 423\,\frac{m}{s} \quad = 478\,\frac{m}{s} \quad = 519\,\frac{m}{s}$

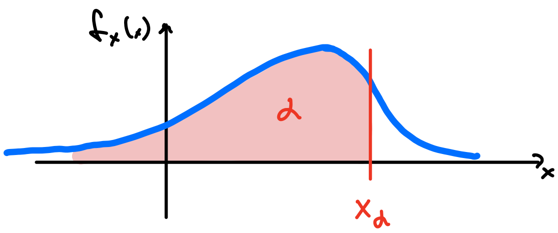
Zgled





da je tako "razmetano" (in porazdelitev tako nesimetrična) vidimo pri tankih absorberjih.

Kvantili / percentili



$$\int_{-\infty}^{x_\alpha} f_x(x)\,dx = \alpha$$

$\alpha = 0{,}1$ $\quad x_{0,1}$ = 10. percentil = 1. decil

$= 0{,}5$ $\quad x_{0,5}$ = 50. percentil = 5. decil = mediana

$x_{0,25}$ = 1. kvartil

$x_{0,75}$ = 3. kvartil

$x_{0,5}$ = 0,5-ti kvantil = 50. percentil = 5. decil = 2. kvartil = mediana

Povprečja (pričakovane vrednosti) in momenti v 2D

X, Y ... skupna verjetnostna gostota $f_{x,y}(x,y)$

$$\bar{X} = \mu_x = E[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f_{x,y}(x,y)\,dx\,dy$$

$$\bar{Y} = \mu_y = E[Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f_{x,y}(x,y)\,dx\,dy$$

Variance:

$$\sigma_x^2 = \mathrm{var}[X] = E[(X-\mu_x)^2] = \iint_{-\infty}^{\infty} (x-\mu_x)^2 f_{x,y}(x,y)\,dx\,dy$$

$$\sigma_y^2 = \mathrm{var}[Y] = E[(Y-\mu_y)^2] = \iint_{-\infty}^{\infty} (y-\mu_y)^2 f_{x,y}(x,y)\,dx\,dy$$

Novost v 2D: <u>Kovarianca</u>

$$\sigma_{x,y} = \mathrm{cov}[X,Y] = E[(X-\mu_x)(Y-\mu_y)] = \iint_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y) f_{x,y}(x,y)\,dx\,dy$$

Lastnosti kovariance:

$$\sigma_{x,y} = E[(X-\mu_x)(Y-\mu_y)] = E[XY - \mu_x Y - \mu_y X + \mu_x\mu_y] =$$

$$= E[XY] - \mu_x \underbrace{E[Y]}_{\mu_y} - \mu_y \underbrace{E[X]}_{\mu_x} + \mu_x \mu_y$$

$$= E[XY] - \mu_x \mu_y \qquad (\overline{XY} - \bar{x} \cdot \bar{y})$$

Če sta X in Y neodvisni naključni sprem. je $\sigma_{xy} = 0$, kajti tedaj imamo faktorizacijo

$$E[XY] = E[X]E[Y] = \mu_x \mu_y$$

Relevantno za merske napake

$$\mathrm{var}[X \pm Y] = \iint \left((x-\mu_x) \pm (y-\mu_y) \right)^2 f_{x,y}(x,y)\, dx\, dy$$

$$= \iint (x-\mu_x)^2 f_{x,y}(x,y)\, dx\, dy \, +$$

$$+ \iint (y-\mu_y)^2 f_{x,y}(x,y)\, dx\, dy$$

$$\pm\ 2 \iint (x-\mu_x)(y-\mu_y) f_{x,y}(x,y)\, dx\, dy =$$

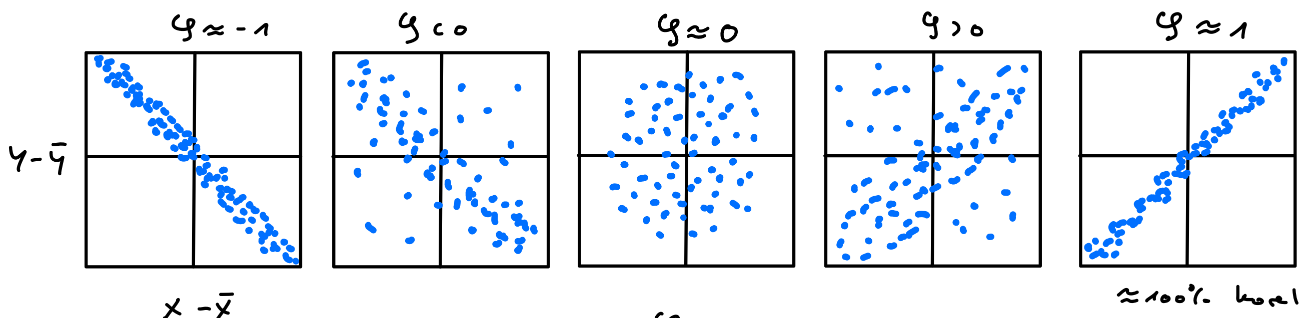
$$= \mathrm{var}[X] + \mathrm{var}[Y] \pm 2\, \mathrm{cov}[X,Y]$$

$$\boxed{\sigma^2_{x \pm y} = \sigma_x^2 + \sigma_y^2 \pm 2\sigma_{xy}}$$

Linearni korelacijski koeficient

$$\boxed{\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}}$$

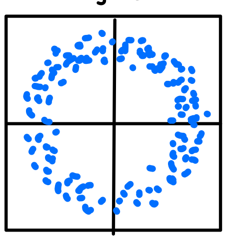
$$|\rho| \le 1$$



$\approx 100\%$ antikorelacija

$\approx 100\%$ korelacija

$\rho \approx 0$



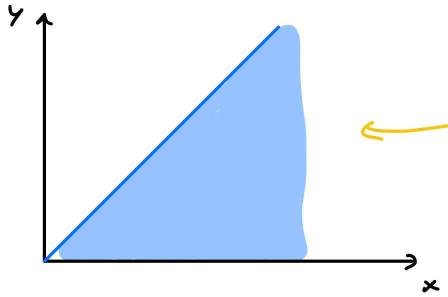
takšni podatki so sicer nekorelirani ($\rho \approx 0$) a so očitno stohastično odvisni.

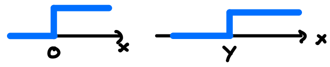
stohastično neodvisni X, Y $\Rightarrow$ $\rho = 0$

$\Leftarrow$ ni nujno res

Zgled $X, Y \dots f_{X,Y}(x,y) = 2e^{-x}e^{-y}\,\Theta(y)\,\Theta(x-y)$

Θ … Heavisideova f.





$y > 0$ $\qquad (x-y) > 0$

oz. $x > y$

Izračunaj $E[\dots]$ in ρ

$$E[X] = \iint_{-\infty}^{\infty} x\, f_{X,Y}(x,y)\,dx\,dy = 2\int_{-\infty}^{\infty} x e^{-x} \int_0^x e^{-y}\,dy\,dx = \frac{3}{2}$$

$E[Y] = \frac{1}{2}$

$E[X^2] = \frac{7}{2}$

$E[Y^2] = \frac{1}{2}$

$$\sigma_x = \sqrt{\mathrm{var}[X]} = \sqrt{E[X^2] - E[X]^2} \doteq 1{,}118$$

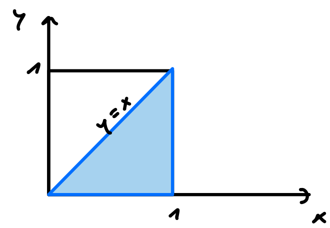
$\sigma_y = 1/2$

$$E[XY] = \iint xy\, f_{X,Y}\,dx\,dy = 1$$

$$\mathrm{cov}[X,Y] = \sigma_{xy} = E[XY] - E[X]E[Y] = 1 - \frac{3}{2}\cdot\frac{1}{2} = \frac{1}{4}$$

$$\rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y} \doteq 0{,}447$$

Zgled Ver. gostota $f_{X,Y}(x,y) = \begin{cases} 8xy & 0 \le x \le 1 \quad 0 \le y \le x \\ 0 & \text{sicer} \end{cases}$



Poišči pogojni pričakovani vrednosti:

ⓐ Y ob pogoju $X = x$

ⓑ X ob pogoju $Y = y$

ⓐ $$E[Y \mid X = x] = \int_{-\infty}^{\infty} y\, \underbrace{f_{Y|X}(y|x)}_{\text{smo nekoč že izračunali}}\,dy = \int_0^x y\,\frac{2y}{x^2}\,dy = \frac{2}{3}x$$

$$f_{Y|X}(y|x) = \begin{cases} \frac{2y}{x^2} & 0 \le y \le x \\ 0 & \text{sicer} \end{cases}$$

ⓑ $$E[X \mid Y = y] = \int_{-\infty}^{\infty} x\, \underbrace{f_{X|Y}(x|y)}\,dx = \int_y^1 \frac{2x^2}{1-y^2}\,dx = \frac{2}{3}\,\frac{1-y^3}{1-y^2} = \frac{2(1+y+y^2)}{3(1+y)}$$

$$f_{X|Y}(x|y) = \begin{cases} \frac{2x}{1-y^2} & y \le x \le 1 \\ 0 & \text{sicer} \end{cases}$$

Zgled 2D normalna porazdelitev (binormala)

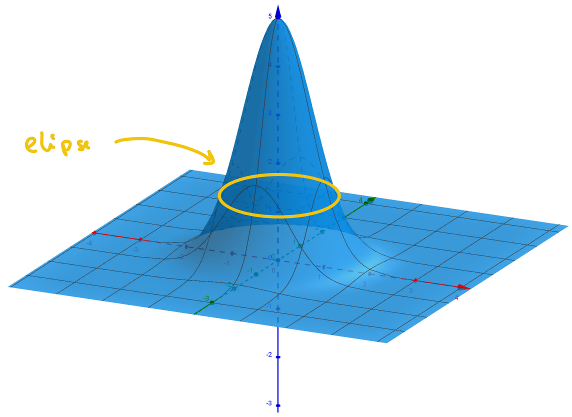
1. Neodvisni X, Y

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y) =$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-(x-\mu_x)^2/2\sigma_x^2} \frac{1}{\sqrt{2\pi}\,\sigma_y} e^{-(y-\mu_y)^2/2\sigma_y^2}$$

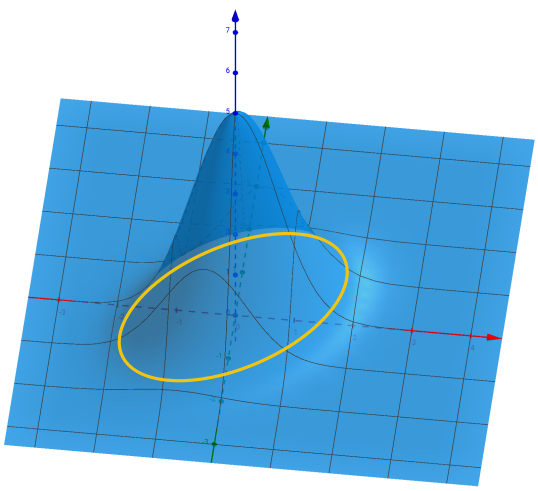
V splošnem elipsa $\sigma_x \neq \sigma_y$ toda niso nagnjene

elipsa



2. Odvisni X, Y (s parametri $\mu_x, \mu_y, \sigma_x, \sigma_y, \rho$)

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



Postrani elipsa

3. Pogojna verjetnostna gostota

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} =$$

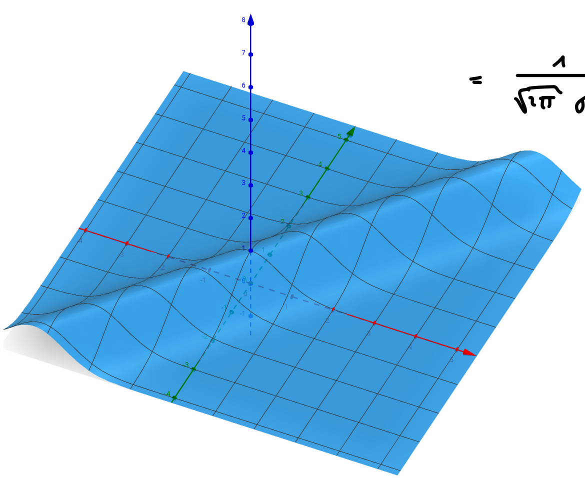
binormala

Gaussovka

$$= \frac{\sqrt{2\pi}\,\sigma_y}{2\pi\,\sigma_x\sigma_y\sqrt{1-\rho^2}} \frac{\exp\left(-\frac{1}{1-\rho^2}\frac{(y-\mu_y)^2}{2\sigma_y^2}\right)}{\exp\left(-(y-\mu_y)^2/2\sigma_y^2\right)} \left(\;\ldots\; \right)$$

ker ostane od binormale

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



Za $f_{Y|X}(x|y)$ samo
$\sigma_x \leftrightarrow \sigma_y$
$\mu_x \leftrightarrow \mu_y$

Entropija (informacijska = Shannonova entropija)

Diskretna naklj. spremen. X, vrednosti $\{x_1, x_2, \dots x_n\}$ z verjetnostmi $\{p_1, p_2, \dots p_n\}$, kjer $P(X = x_i) = p_i$ in $\sum_{i=1}^{n} p_i = 1$.

Z vsakim izidom dobimo tudi informacijo $I(p)$.

Kako razumeti "količino informacije"

Očitno mora imeti ta mera logaritemski značaj

Razmislek: "sneži" (1. izid) p_1, "danes je sreda" (2. izid) p_2

Dogodka neodvisna $\Rightarrow$ verjetnost $p_1 \cdot p_2$

$$I(p_1 p_2) = I(p_1) + I(p_2)$$

Poleg tega želimo nenegativnost $I(p) \geq 0$, in monotonost:

$$p_1 < p_2 \Rightarrow I(p_1) > I(p_2)$$

Manj verjetni dogodki nosi več informacije

$$\rightsquigarrow \quad I(p) = -C \log_b p$$

baza, običajno $2 \Rightarrow I$ v bitih

običajno 1

Od tod le še korak do entropije, s katero bi radi podali povprečno količino dobljene informacije

$$H(p_1, p_2, \dots p_n) = \sum_{i=1}^{n} p_i I(p_i) = -\sum_{i=1}^{n} p_i \log p_i$$

Zgled: poskus kovanca: grb in cifra enako verjetna $\Rightarrow p_1 = p_2 = \frac{1}{2}$

Entropija naklj. spremen. s takšno porazd. je

$$H = -\frac{1}{2}\log_2 \frac{1}{2} - \frac{1}{2}\log_2 \frac{1}{2} = 1$$

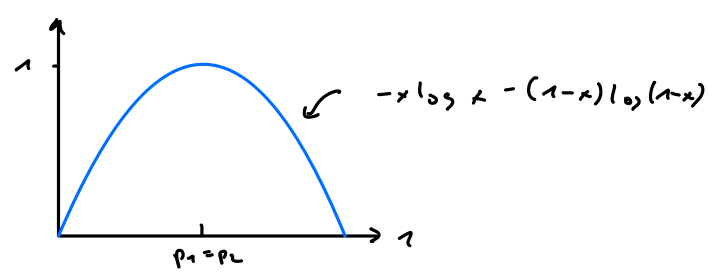
z metom kovanca dobimo 1 bit informacije

Če kovanec ni pošten npr. $p_1 = \frac{9}{20}$ $p_2 = \frac{11}{20}$

$$H = -\frac{9}{20}\log_2 \frac{9}{20} - \frac{11}{20}\log_2 \frac{11}{20} \approx 0{,}9928$$

Večja kot je entropija manj smo gotovi v končno stanje tega sistema

Splošneje $H(p_1, p_2) = H(p, 1-p)$



Zgled: kocka, poštena: $p_i = \frac{1}{6}$, $1 \le i \le 6$

$$H = -6 \cdot \frac{1}{6} \log_2 \frac{1}{6} = \log_2 6 = 2{,}585$$

pri metu kocke v povprečju dobimo 2,5 bit informacij

Kaj pa če nam nekdo razkrije, da je # pik sodo ⇒ zmanjša se vzorčni prostor
⇒ $p_2 = p_4 = p_6 = 1/3$

$$H = -3 \, \frac{1}{3} \log_2 \frac{1}{3} = \log_2 3 = 1{,}585$$

dodatna informacija je vredna 1 bit.

Očitna posplošitev na zvezni primer $\{ X$ zvez. nakl. spr. $\}$

$$H = -\overline{\log f_X} + C = -\int \log f_X(x)\, f_X(x)\, dx + C$$

! Pokažimo, da ima normalna porazdelitev, med vsemi zveznimi porazd. z enakim povprečjem μ in varianco σ^2 maksimalno entropijo.

Lagrangeova funkcija:

lagr. multiplikator

$$\mathcal{L} = -\int_{-\infty}^{\infty} f_X(x) \log f_X(x)\, dx + \lambda \Big(\overbrace{1 - \int f_X(x)\, dx}^{\text{vez = constraint}} \Big) + \nu \Big(\overbrace{\sigma^2 - \int (x-\mu)^2 f_X(x)\, dx}^{\text{vez = constraint}} \Big)$$

lagr. mult.

Poskusimo z variacijo δf_X

$$0 = \delta \mathcal{L} = \int (-\delta f_X(x)) \Big(\underbrace{\log f_X(x) + 1 + \lambda + \nu (x-\mu)^2}_{=0} \Big)\, dx$$

$$\delta \left(-\int f \log f\, dx \right) = -\int \left(\delta f \log f + f \frac{1}{f} \delta f \right) dx = \int (-\delta f)(\log f + 1)\, dx$$

Ker mora to veljati za poljubno variacijo $\delta f_X(x)$ mora $\sqcup = 0$

$$\log f_X(x) + 1 + \lambda + \nu (x-\mu)^2 = 0$$

$$f_X(x) = e^{-1-\lambda-\nu(x-\mu)^2}$$

Uporabimo enačbi vezi za normo in σ^2:

$$\int_{-\infty}^{\infty} f_X(x)\, dx = e^{-\lambda-1} \int_{-\infty}^{\infty} e^{-\nu(x-\mu)^2} dx = e^{-\lambda-1} \frac{\sqrt{\pi}}{\sqrt{\nu}} = 1$$

$$\int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\, dx = e^{-\lambda-1} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\nu(x-\mu)^2} dx = e^{-\lambda-1} \frac{\sqrt{\pi}/2}{\sqrt{\nu}^3} = \sigma^2$$

$$\Rightarrow \quad 1 = 2\nu\sigma^2 \qquad \nu = \frac{1}{2\sigma^2}$$

$$\text{in} \quad e^{-\lambda-1} = \frac{1}{\sqrt{2\pi}\,\sigma}$$

$$\Rightarrow \quad f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

Posebne diskretne porazdelitve

Dve najpomembnejši: binomska, Poissonova

Binomska porazdelitev : samo dve vrsti izidov, da/ne, 0/1, ...
= Bernoullijeva

verjetnost za "uspešen" izid ... p

"neuspešen" ... $1-p$

Verjetnost, da v N poskusih dobimo n "uspešnih" (in $N-n$ "neuspešnih") podaja dvoparametrična porazdelitev

$$\boxed{P(X=n;N,p) = \binom{N}{n} p^n (1-p)^{N-n}}$$

vse kombinacije, verjetnost uspešnih, verjetnost neuspešnih

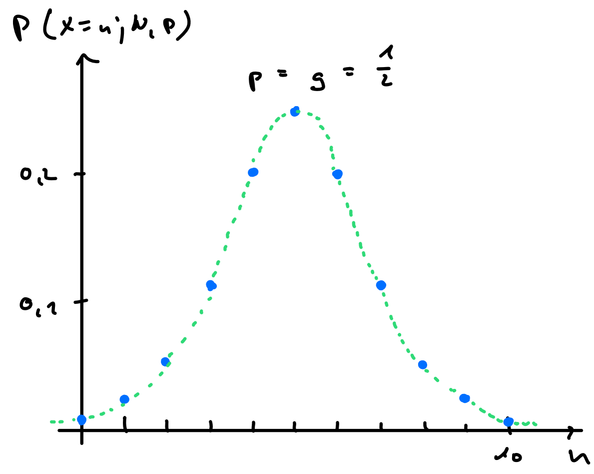
Včasih $q = 1-p$

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

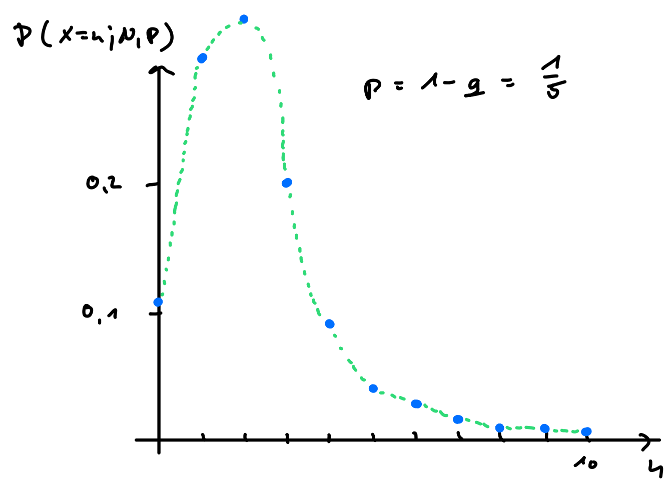
Porazdelitev je normirana: $\sum_{n=0}^{N} P(X=n;N,p) = \sum_{n=0}^{N} \binom{N}{n} p^n q^{N-n} = (p+q)^N = 1^N = 1$

Zgled

$N = 10$



št uspešnih izidov



Zgled 5× zaporedoma vržemo pošteno kocko

? Verjetnost, da je število pik 7 nastopilo natanko 2×?

$P(X=2; N=5; p=\frac{1}{6}) = \binom{5}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{5-2} = 0{,}161$

? Verjetnost, da se 7 pik pojavijo največ enkrat $(X \le 1)$?

$P(X \le 1) = P(X=0) + P(X=1) = \binom{5}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^5 + \binom{5}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^4 = \dots$

? Verjetnost, da se 3 pike pojavijo najmanj 2× $(X \ge 2)$?

Naporno: $P(X \ge 2) = P(X=2) + \dots + P(X=5) = \dots = 0{,}196$

Enostavneje: $P(X \ge 2) = 1 - P(X \le 1)$

Povprečje in varianca binomske porazdelitve

izmislimo si novo [illegible]

$$\sum_{n=0}^{N} \binom{N}{n} (\lambda p)^n q^{N-n} = (\lambda p + q)^N$$

odvajamo po λ

$$\underbrace{\sum_{n=0}^{N} \binom{N}{n} n \lambda^{n-1} p^n q^{N-n}}_{=\bar{X}} = N(\lambda p + q)^{N-1} p$$

še enkrat

$$\underbrace{\sum_{n=0}^{N} \binom{N}{n} n(n-1) \lambda^{n-2} p^n q^{N-n}}_{=\overline{X(X-1)}} = N(N-1)(\lambda p + q)^{N-1} p^2$$

- $\lambda \to 1$: uvidimo $\bar{X} = N(p+q)^{N-1} p = Np$ $\qquad \boxed{\bar{X} = Np}$
- $\overline{X(X-1)} = N(N-1)p^2$

$$\mathrm{var}[X] = \overline{X^2} - \bar{X}^2 = \overline{X(X-1)} + \bar{X} - \bar{X}^2 \qquad \boxed{\mathrm{var}[X] = Npq}$$

$$= N(N-1)p^2 + Np - N^2p^2 = Np - Np^2 = Np(1-p) = Npq$$

Zgled 2000 družin, vsaka 4 otroke. Koliko družin pričakujemo

a) vsaj 1 deček

b) 2 dečka $\qquad p = \frac{1}{2} = q$ (verjetnost za deklico, dečka)

c) 1 ali 2 deklici

d) nobene deklice

$N = 4$ (ena sama družina)

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

$P(X=2) = \ldots = \frac{3}{8}$

$P(X=3) = \ldots = \frac{1}{4}$

$P(X=4) = \ldots = \frac{1}{16}$

$P(X=0) = \ldots = \frac{1}{16}$

a) $P(X \geq 1) = \frac{15}{16} \Rightarrow 1875$ družin

b) $P(X=2) = 3/8 \Rightarrow \frac{3}{8} \cdot 2000 = 750$ družin

Aproksimacija binomske porazdelitve z normalno

- deluje, kadar je N velik in niti p niti $q = 1-p$ nista preblizu nič.
- Laplaceov limitni izrek:

$$\lim_{N\to\infty} P\left(a \leq \frac{X - \overbrace{Np}^{\bar{X}}}{\underbrace{\sqrt{Npq}}_{\sigma_X}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx$$

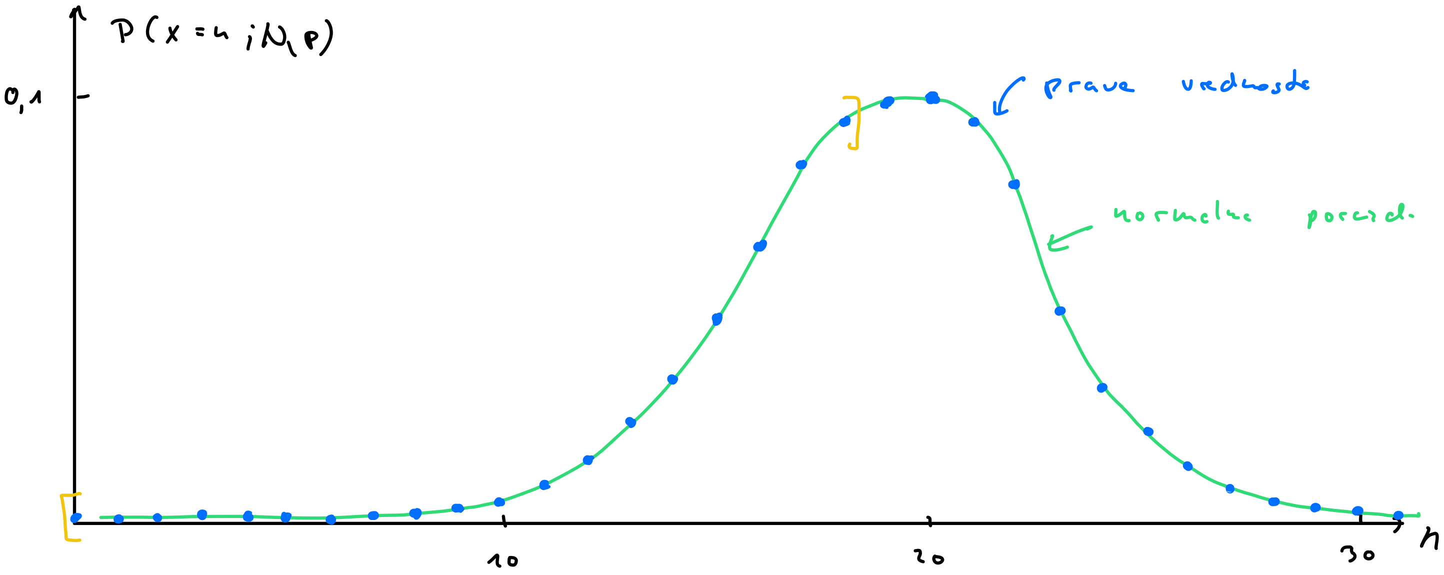
- spremenljivka $\frac{X - \bar{X}}{\sigma_X} = \frac{X - Np}{\sqrt{Npq}}$ ≈ asimptotsko normalno porazdeljena
- dobro deluje že pri $Np, Nq \gtrsim 5$

Zgled Kocko vržemo 120-krat ($N=120$). Kolikšna je verjetnost, da se bodo 4 pike prikazale največ 18-krat.

binomska $N=120$ $\quad P(4 \text{ pike}) = p = \frac{1}{6}$, $\quad q = \frac{5}{6}$

Točen odgovor $\displaystyle\sum_{n=0}^{18} P(X=n; N, p) = \sum_{n=0}^{18} \binom{120}{n}\left(\frac{1}{6}\right)^n\left(\frac{5}{6}\right)^{120-n} \doteq 0{,}3657$

Aproksimacija $\bar{X} = Np = 120 \cdot \frac{1}{6} = 20 \qquad \sigma_X = \sqrt{Npq} \doteq 4{,}08$



Meje pri 0 in 18 v stand. sprem. $z_1 = \frac{0-20}{4{,}08}$ $z_2 = \frac{18-20}{4{,}08}$

bolje $\rightarrow$ $z_1 = \frac{-0{,}5-20}{4{,}08}$ $z_2 = 18{,}5-20$

Odgovor z norm. aproks.: $P(X \le 18) \approx F_X(-0{,}37) - F_X(-5{,}02)$

Poissonova porazdelitev

Poissonova por. je limita binomske, ko $p \to 0$ (posamezni uspešni izidi zelo malo verjetni) in istočasno $N \to \infty$, vendar tako, da ostane $\bar{X} = Np$ enako.

$$p = \frac{\bar{X}}{N} \qquad q = 1 - \frac{\bar{X}}{N}$$

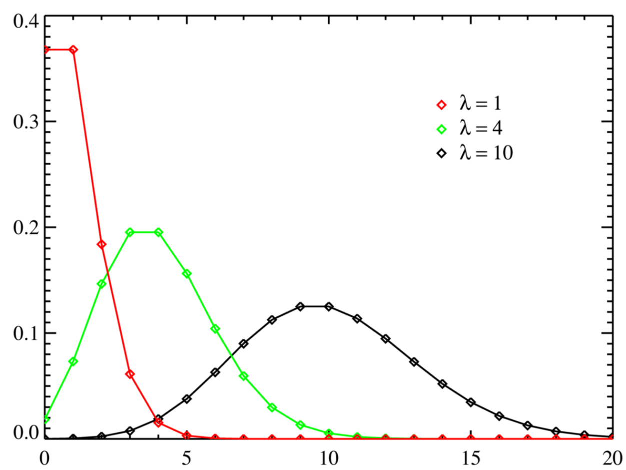
Dano v eno samo binomsko verjetnost

$$\binom{N}{n}\left(\frac{\bar{X}}{N}\right)^n \left(1-\frac{\bar{X}}{N}\right)^{N-n} = \frac{N\cdot(N-1)(N-2)\cdots(N-n+1)}{n!\,N^n}\,\bar{X}^n \left(1-\frac{\bar{X}}{N}\right)^{N-n} =$$

$$= \frac{1}{n!}\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)\bar{X}^n\left(1-\frac{\bar{X}}{N}\right)^{N-n}$$

Ko $N \to \infty$, je zadnji faktor $\lim_{N\to\infty}\left(1-\frac{\bar{X}}{N}\right)^{N-n} = \lim_{N\to\infty}\left(1-\frac{\bar{X}}{N}\right)^{N}\underbrace{\left(1-\frac{\bar{X}}{N}\right)^{-n}}_{\to 1} = e^{-\bar{X}}$

Ostane le $P(X=n;\bar{X}) = \frac{1}{n!}\bar{X}^n e^{-\bar{X}}$



Običajno označimo

$$\boxed{\bar{X} = \lambda = Np \qquad P(X=n, \lambda) = \frac{\lambda^n e^{-\lambda}}{n!} \qquad n = 0, 1, 2, \ldots}$$

Od binomske "podedujemo" tudi povprečje in varianco

$$E[X] = \bar{X} = Np$$
$$\mathrm{var}[X] = \sigma_X^2 = \underbrace{Np}_{\lambda}\underbrace{q}_{1} = Np = \lambda$$

Razmislek:

$$X_{izmerjena} = \bar{X} \pm \overset{\pm\sigma_x}{\sqrt{\bar{X}}}$$

Npr. pri razpadih jeder v času t zabeležimo X razpadov ($X \gg 1$)
Ocena za aktivnost vzorca (= # razpadov na čas. enoto) je $\hat{a} = \frac{X}{t}$, toda prava aktivnost vzorca je $a \neq \hat{a}$
Izmerjeni X je negotov, zato je raztresen okrog $\bar{X} = a t$ za $\pm\sigma_x = \pm\sqrt{\bar{X}}$

tega ne poznamo
zato $\sigma_x \approx \pm\sqrt{X}$

$$\underset{\text{izmerjen}}{X} = \bar{X} \pm \sqrt{\bar{X}} \approx \underset{\text{pravi}}{\bar{x}} \pm \sqrt{\underset{\text{izmerjen}}{x}}$$

$$\bar{X} \doteq X \pm \sqrt{X}$$

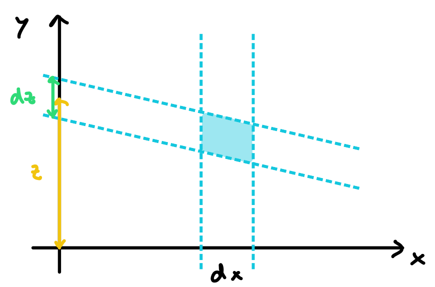
$$a \doteq \hat{a}\left(1 \pm \frac{1}{\sqrt{X}}\right) \quad \text{prava aktivnost}$$

Če hočemo določiti a na 1% natančno moramo izmeriti $X = 10^4$ razpadov

Razpad delcev = Poissonova porazd., ker je jalen izid (razpad, ne razpad) in p je majhen.

Vsota stohastično neodvisnih spremenljivk (~ konvolucija)

- neodvisni X in Y, porazdeljeni po $f(x)$ in $g(y)$
- zanima nas porazdelitev vsote spremenljivke $Z = X + Y$



$z = x + y$

$y = z - x$

$$dP = f(x)\, g(y)\, dx\, dz$$

Konvolucija

$$\boxed{h(z) = \int f(x)\, g(z-x)\, dx = f * g}$$

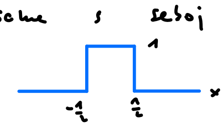
verj. gostota za $Z = X + Y$

Posplošitev na več dimenzij

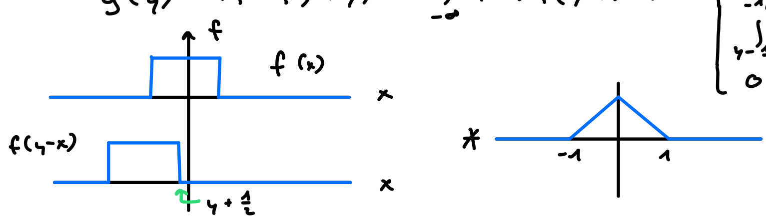
$$(f_1 * f_2 * f_3)(z) = \iint f_1(x_1)\, f_2(x_2)\, f_3(z - x_1 - x_2)\, dx_1\, dx_2$$

Zgled: Zaporedna konvolucija enak. porazdelitve same s seboj

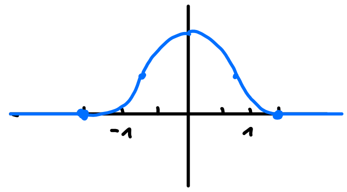
$$f(x) = \begin{cases} 1 ; & -\frac{1}{2} \le x \le \frac{1}{2} \\ 0 ; & \text{sicer} \end{cases}$$



$$g(y) = (f * f)(y) = \int_{-\infty}^{\infty} f(x)\, f(y-x)\, dx = \begin{cases} \int_{-1/2}^{y+\frac{1}{2}} 1 \cdot 1\, dx = y + 1 & (-1 \le y \le 0) \\ \int_{y-\frac{1}{2}}^{1/2} 1 \cdot 1\, dx = 1 - y & (0 \le y \le 1) \\ 0 & y < -1,\ y > 1 \end{cases}$$



DN: $h(y) = (f*g)(y) = (f*f*f)(y)$



Kaj se pri konvoluciji dogaja z momenti porazdelitev?

$Z = X + Y$, $X \sim f$, $Y \sim g$, $Z \sim h = f*g$

$$\bar{Z} = \int z\, h(z)\, dz = \int z \left(\int f(x)\, g(z-x)\, dx \right) dz \overset{\text{zamenjava vrst. reda integracije}}{=} \int \left(\int z\, g(z-x)\, dz \right) f(x)\, dx =$$

$$= \int \left(\int (x+y)\, g(y)\, dy \right) f(x)\, dx = \int x \Big(\overbrace{\int g(y)\, dy}^{1}\Big) f(x)\, dx + \overbrace{\int f(x)\, dx}^{1} \int y\, g(y)\, dy =$$

$$= \bar{X} + \bar{Y}$$

To ni presenečenje: $\underline{\bar{Z} = \bar{X} + \bar{Y}}$

Varianca

∅ zaradi predpostavke, da sta X in Y neodvisni.

$$\overline{(Z - \bar{Z})^2} = \overline{(X + Y - \bar{X} - \bar{Y})^2} = \overline{(X - \bar{X})^2} + 2\,\overline{(X-\bar{X})(Y-\bar{Y})} + \overline{(Y - \bar{Y})^2} =$$

Torej $\underline{\mathrm{var}[X+Y] = \mathrm{var}[X] + \mathrm{var}[Y]}$

oz. $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$

Za 3. in 4. centralni moment (poševnost, eksces) velja:

$$\varphi_{X+Y}\, \sigma_{X+Y}^3 = \varphi_X\, \sigma_X^3 + \varphi_Y\, \sigma_Y^3$$

$$\varepsilon_{X+Y}\, \sigma_{X+Y}^4 = \varepsilon_X\, \sigma_X^4 + \varepsilon_Y\, \sigma_Y^4$$

Bolj zanimivo po N-kratni konvoluciji:

$$\bar{Z} = N \bar{X}$$
$$\sigma_Z^2 = N \sigma_X^2$$
$$\varphi_Z = \frac{\varphi_X}{N} \longrightarrow 0$$
$$\varepsilon_Z = \frac{\varepsilon_X}{N} \longrightarrow 0$$

ko $N \to \infty$

Centralni limitni izrek CLT (central limit theorem)

$$h^{(n)} = (f * f * \dots * f)(z) = \iint \cdots \int f(x_1)\, f(x_2) \dots f(z - x_1 - x_2 - \dots - x_n)\, dx_1\, dx_2 \dots dx_{n-1}$$

n-kratna konvolucija

- po n-konvolucijah povprečje "leze": $\bar{z}^{(n)} = \sum_{i=1}^{n} \bar{x}_i$
- podobno lahko "leze" tudi varianca vsote sprem.: $(\sigma^{(n)})^2 = \sum_{i=1}^{n} \sigma_i^2$

Teh težav se rešimo s sprotnim popravljanjem

$$\zeta = \frac{z - \bar{z}^{(n)}}{\sigma^{(n)}}$$

Sproti prilagajamo $\zeta \leftarrow \sigma^{(n)} \zeta + \bar{z}^{(n)}$

Slutimo, da bo veljalo

$$\boxed{\lim_{n\to\infty} \sigma^{(n)} h^{(n)}\left(\sigma^{(n)}\zeta + \bar z^{(n)}\right) = \frac{1}{\sqrt{2\pi}} e^{-\zeta^2/2}}$$

Ni za izpit

Dokaz

- poenostavimo: vsi f_i enaki, $f = f_i$ $\forall i$, dokazati se da tudi brez te omejitve
- privzetek: vse povprečja nič, $\bar x_i = 0$, vsi σ_i enaki, $\sigma = \sigma_i$ $\forall i$
- Uporabimo Fourierovo transformacijo $F(k) = \int_{-\infty}^{\infty} f(x) e^{-ikx}\, dx$

transformiran h

$$H^{(n)}(k) = \int_{-\infty}^{\infty} h^{(n)}(x) e^{-ikx}\, dx$$

Iz analize vemo: $H^{(n)}(k) = (F(k))^n$

Kaj vemo o $F(k)$? e^{-ikx} razvijemo po Taylorju

$$e^{-ikx} = 1 - ikx - \frac{k^2x^2}{2} + i\frac{k^3}{6}x^3 + O((kx)^4)$$

$$F(k) = \underbrace{\int f(x)\,dx}_{=1} - ik \underbrace{\int x f(x)\, dx}_{=\bar x = 0 \text{ po privzetku}} - \frac{k^2}{2} \underbrace{\int x^2 f(x)\, dx}_{=\sigma^2 \text{ ker } \bar x = 0} + i\frac{k^3}{6}\int x^3 f(x)\,dx + \dots$$

$$= 1 - \frac{1}{2}\sigma^2 k^2 + \underset{\text{residual (ostanek)}}{R} \qquad |R| \leq \frac{1}{6}|k|^3\, \overline{|x|^3}$$

Torej
$$H^{(n)}(k) = (F(k))^n = \left(1 - \frac{1}{2}\sigma^2k^2 + R\right)^n = \left(1 - \frac{1}{2}\frac{\varkappa^2}{n} + \underbrace{O\left(\frac{\varkappa^3\, \overline{|x|^3}}{6 n^{3/2}\sigma^3}\right)}_{\text{če tole pozabimo}}\right)^n$$

$$\varkappa = \sigma^{(n)} k = \sqrt{n}\,\sigma k$$

V limiti
$$\lim_{n\to\infty} H^{(n)}(k) = \left(1 + \frac{1}{n\left(-\frac{2}{\varkappa^2}\right)}\right)^n = e^{-\frac{\varkappa^2}{2}}$$

to pa je ravno Four. transformiranka normalne gostote.

Nazadnje, po obratni Fourierovi transformaciji dobimo

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\varkappa^2/2} e^{+i\varkappa x}\, d\varkappa = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

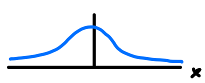
$\square$

Edini problem, če $\overline{|x|^3}$ ne bi bil omejen

$$\overline{|x|^3} = \int_{-\infty}^{\infty} |x|^3 f(x)\, dx < \infty$$

če velja CLT, če ta moment ni omejen, CLT ne velja in $h^{(n)}$ ne gre proti Gaussovki

Npr. Cauchy: $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$



$$\nexists \int_{-\infty}^{\infty} x^n f(x)\, dx \quad (= \infty) \quad \Rightarrow \quad \overline{|x|^3} = \infty$$

$$h^{(2)}(z) = \frac{1}{\pi}\frac{2}{4+z^2} \qquad h^{(3)}(z) = \frac{1}{\pi}\frac{3}{9+z^2} \quad \dots \quad h^{(n)}(z) = \frac{1}{\pi}\frac{n}{n^2+z^2}$$

Cauchyjeva porazdelitev je stabilna na konvolucijo.

Statistike

Osnovna naloga: empirično določanje verjetnostnih porazdelitev oz. njihovih značilnih parametrov (npr μ, σ pri gaussovki)

Ključni pojmi: populacija in vzorec

Vedno vzamemo (zajamemo) vzorec iz populacije, običajno:
$\dim(\text{vzorec}) \ll \dim(\text{populacija})$

Zgled: populacija $N = 2 \cdot 10^6$ slovencev, vzorec $n = 1000$ ljudi, izmerimo njihovo višino → na podlagi vzorca bi radi sklepali na pravo povprečje in pravo razpršenost

Populacija je lahko tudi ∞ (∞ metov kovanca)

Vzorčno povprečje (sample mean, sample average)

Pričakovana vrednost (povprečje) diskretne naklj. spre. smo def. kot

$$E[X] = \sum_{i=1}^{n} x_i P(x_i) , \quad \text{če} \quad P = P(x_i) = \frac{1}{n} \ \forall i \ \text{dobimo} \quad E[x] = \frac{1}{n} \sum_i x_i$$

Po analogiji def. vzorčno povprečje

tudi naklj. sprem.

$$\boxed{\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i} \qquad \text{(ali tudi z malimi črkami)}$$

Podobno za varianco

$$\sigma_x^2 = \mathrm{var}[x] = E[(x - \bar{x})^2] = \sum_{i=1}^{n} (x_i - \mu)^2 P(x = x_i)$$

$$\rightsquigarrow \text{oz.} \quad \frac{1}{n} \sum_{i=1}^{n} (x - \mu)^2 \quad \text{če} \quad P(x_i) = \frac{1}{n}$$

Torej def. vzorčno varianco

$$S_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Vzorec: $\bar{x}$, S_x^2 (vsakokrat drugačni za vsak vzorec)
Populacija: μ, σ_x^2 (pravi vrednosti)

Količinam, ki jih tvorimo iz naklj. sprem $X_1, X_2 \ldots X_n$, ki lahko zavzamejo vrednosti $x_1, x_2, \ldots, x_n$, so tudi same naklj. sprem. in jih imenujemo STATISTIKE (oz. vzorčne statistike.

statistika (veda) = statistics
statistika (kons. naklj. sprem.) = statistic

npr. $\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$

Ker $\bar{X}$ fluktuira se lahko vprašamo po povprečju povprečja $E[\bar{X}]$

$E[\bar{X}] = \mu$ ⇒ pravimo, da je vzorčno povprečje nepristranska cenilka (unbiased estimator) populacijskega povprečja

Kaj pa varianca? Za neodvisni X in Y smo že pokazali:

$var[X \pm Y] = var[X] + var[Y]$

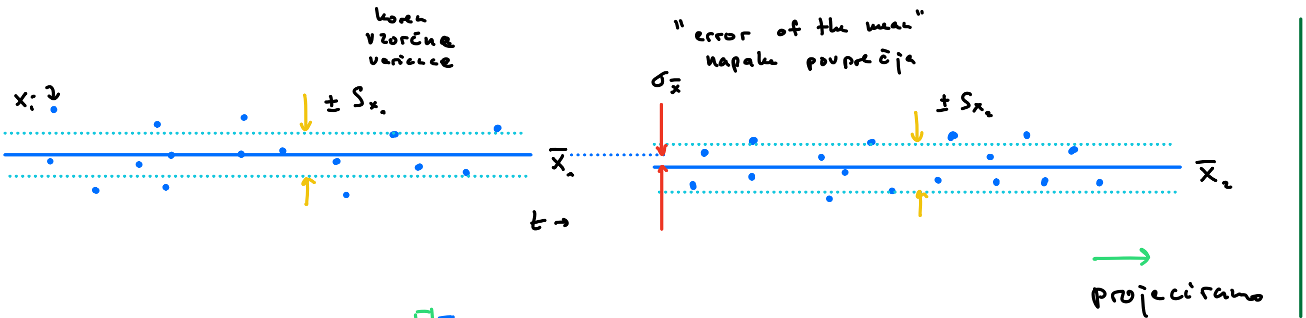
in $var[cX] = c^2 \, var[X]$ c=konst.

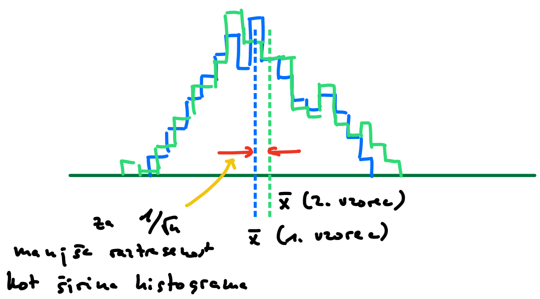
$$var[\bar{x}] = \frac{1}{n^2} var[x_1] + \frac{1}{n^2} var[x_2] + \dots + \frac{1}{n^2} var[x_n]$$

$$= n \frac{1}{n^2} var[x]$$

↑ brez indeksa, vsi X_i tak iste porazdelitve

$\boxed{\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n}}$ oz. $E[(\bar{x}-\mu)^2] = \sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n}$





Kaj pa velja za negotovost določitve variance (na podlagi vzorca)? – varianca variance

$$E[(S_x^2 - \sigma^2)^2] = \overline{\left(\frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \dots + (x_n-\bar{x})^2}{n} - \sigma^2\right)^2} = \dots$$

Kaj dobimo pri kvadriranju $(\dots)^2$

- kvartične člene $(x_i-\bar{x})^4$, ki jih je n
- kvadratične člene $2(x_i-\bar{x})^2(x_j-\bar{x})^2$ za $i<j$, ki jih je $\frac{n(n-1)}{2}$
- mešani člen $-2S_x^2\sigma^2 \approx -2\sigma^4$
- in σ^4

→ približek $\overline{(x_i-\bar{x})^2(x_j-\bar{x})^2} \approx \overline{(x_i-\bar{x})^2}\;\overline{(x_j-\bar{x})^2} = S_x^2 S_x^2 \approx \sigma^4$

→ približek $\overline{(x_i-\bar{x})^4}$ = 4. centralni moment $= M_4 = 3\sigma^4$

↑ to velja za normalno porazdelitev

$$\dots = \frac{1}{n^2}\left(3n\sigma^4 + 2\frac{n(n-1)}{2}\sigma^4 - 2\sigma^4 n^2\right) + \sigma^4 = \frac{2\sigma^4}{n}$$

→ $\boxed{S_x^2 = \sigma^2\left(1 \pm \sqrt{\frac{2}{n}}\right)}$ oz. $\boxed{S_x = \sigma\left(1 \pm \frac{1}{\sqrt{2n}}\right)}$ Raztresenost variance in σ

Taylor

Zadnjič smo pokazali $E[\bar{X}] = \mu$

vzorčno povprečje, pravo populacijsko povprečje

Ali velja tudi $E[S_x^2] = \sigma^2$? Ne!

Namreč $X_1 - \bar{X} = X_1 - \frac{1}{n}(X_1 + X_2 + \dots + X_n) = \frac{1}{n}\left((n-1)X_1 - X_2 - \dots - X_n\right)$

$$= \frac{1}{n}\left((n-1)(X_1-\mu) - (X_2-\mu) - \dots - (X_n-\mu)\right)$$

Kvadriram $(X_1 - \bar{X})^2 = \frac{1}{n^2}\left((n-1)^2(X_1-\mu)^2 + (X_2-\mu)^2 + \dots + (X_n-\mu)^2 + \text{mešani členi}\right)$

odpadejo, ker kovarianca $\sim \overline{(X_i-\mu)(X_j-\mu)} = 0$ če $i \neq j$

$$E[(X_1-\bar{X})^2] = \frac{1}{n^2}\Big((n-1)^2 \underbrace{E[(X_1-\mu)^2]}_{\sigma^2} + \underbrace{E[(X_2-\mu)^2]}_{\sigma^2} + \dots\Big)$$

$$= \frac{1}{n^2}\left((n-1)^2\sigma^2 + (n-1)\sigma^2\right) = \frac{n-1}{n}\sigma^2$$

To ponovimo še za vsak $E[(X_k-\bar{X})^2]$ za $k = 2, \dots, n$

$$E[S_x^2] = \frac{1}{n}\left(E[(X_1-\bar{X})^2] + \dots + E[(X_n-\bar{X})^2]\right) = \frac{1}{n}\, n\, \frac{n-1}{n}\sigma^2$$

$$\boxed{E[S_x^2] = \frac{n-1}{n}\sigma^2}$$

Vzorčna varianca je pristranska cenilka (biased estimator) populacijske variance

Nepristranska bi bila: $\hat{\sigma}^2 = \frac{n}{n-1} S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ Nepristranska

$S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ Pristranska

(za $n \gg 1$ nerelevantno)

Podrobnost: doslej smo vzorec (n) zajemali iz populacije (N, $N > n$ ali $N \gg n$) z nadomeščanjem (with replacement)

Če vzorčimo brez nadomeščanja i.e. istega elementa populacije ne moremo vzorčiti večkrat, velja:

$$E[\bar{X}] = \mu$$

$$\operatorname{var}[\bar{X}] = \underbrace{\frac{N-n}{N-1}}_{\text{Novo}} \frac{\operatorname{var}[X]}{n} \quad \operatorname{var}[X] = \sigma^2$$

Zgled Celotna populacija $= \{2, 3, 6, 8, 11\} = \{x_i\}_{i=1}^{5}$ $N = 5$

Pravo povprečje in varianca

$\mu = \frac{1}{N}\sum_i x_i = \underline{6}$

$\sigma^2 = \frac{1}{N}\sum_i (x_i - \mu) = 10,8$

$\sigma \doteq \underline{3,29}$

Iz te populacije izberemo vse možne vzorce velikosti $n=2$ z nadomeščanjem. Takih je $\mathcal{N} = N^2 = 25$

$$\begin{matrix} \{2,2\}, & \{2,3\}, & \{2,6\}, & \{2,8\}, & \{2,11\}, \\ \{3,2\}, & \{3,3\} & \dots & & \{3,11\}, \\ \vdots & & \ddots & & \vdots \\ \{11,2\} & & \dots & & \{11,11\} \end{matrix}$$

Lahko tvorimo $\mathcal{N}$ vzorčnih povprečij
$\{2.0,\ 2.5,\ 4.0,\ 5.0,\ \dots,\ 11.0\}$ označimo jih z $\bar{X}_k$ $k=1,2,\dots,\mathcal{N}$

povprečje vseh teh je
$$E[\bar{X}] = \frac{1}{\mathcal{N}} \sum_{k=1}^{\mathcal{N}} \bar{X}_k = \underline{6} \quad (=\mu)$$

Razpršenost povprečja
$$\mathrm{var}[\bar{X}] = \frac{1}{\mathcal{N}} \sum_{k=1}^{\mathcal{N}} (\bar{X}_k - \mu)^2 = 5{,}40 \qquad \sqrt{\mathrm{var}[\bar{X}]} = \underline{2{,}32}$$

Res: $\mathrm{var}[\bar{X}] = \frac{\mathrm{var}[X]}{n} = \frac{10{,}8}{2} = 5{,}4$

Če vzorčimo brez vračanja v populacijo. Lahko naredimo $\mathcal{N} = \binom{5}{2} = 10$:

$\{2,3\}, \{2,6\}, \{2,8\}, \{2,11\}, \{3,6\}, \{3,8\}, \{3,11\}, \{6,8\}, \{6,11\}$
$\{8,11\}$

Vzorčna povprečja $= 2.5,\ 4.0,\ \dots,\ 9.5 = \{\bar{X}_k\}_{k=1}^{\mathcal{N}}$

$$E[\bar{X}] = \frac{1}{\mathcal{N}} \sum_{1}^{\mathcal{N}} \bar{X}_k = 6 \quad (=\mu)$$
$$\mathrm{var}[\bar{X}] = \frac{1}{\mathcal{N}} \sum_{1}^{\mathcal{N}} (\bar{X} - \mu)^2 = 4{,}05 \qquad = \frac{N-n}{N-1} \frac{\mathrm{var}[X]}{n} = \frac{5-2}{4-1} \frac{10{,}8}{2}$$

Vzorčne porazdelitve vsot in razlik (sampling distribution of ...)

- Imamo 2 populaciji; iz ene izberemo vzorce velikosti n_1, iz druge pa n_2.
- Izračunamo vzorčni povprečji: $\bar{X}_1$ in $\bar{X}_2$

Če vemo
$$E[\bar{X}_1 \pm \bar{X}_2] = E[\bar{X}_1] \pm E[\bar{X}_2] = \mu_1 \pm \mu_2$$
$$\mathrm{var}[\bar{X}_1 \pm \bar{X}_2] = \mathrm{var}[\bar{X}_1] + \mathrm{var}[\bar{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Velja, da je standardizirana spremenljivka

$$z = \frac{\bar{X}_1 \pm \bar{X}_2 - (\mu_1 \pm \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1}}}$$

v zelo dobrem in približnu normalno porazdeljena (velja za $n_1, n_2 \geq 30$).

Zgled Žarnice proizvajalca A: povp. živ. doba 1400 h, std. dev. 200 h
Žarnice proizvajalca B: povp. živ. doba 1200 h, std. dev. 100 h

Testiramo $n_A = 125$ in $n_B = 125$ žarnic iz vsake tovarne.

ⓐ ? verjetnost, da bodo imele žarnice A povprečno dobo vsaj 160 h več kot B?
ⓑ ? 250 h več kot B?

$$E[\bar{X}_A - \bar{X}_B] = 1400\,h - 1200\,h = 200\,h$$
$$\mathrm{var}[\bar{X}_A - \bar{X}_B] = \sigma_A^2/n_A + \sigma_B^2/n_B = \frac{1}{125}\left(200^2 + 100^2\right) = 400\,h^2$$

Std. spremenljivka

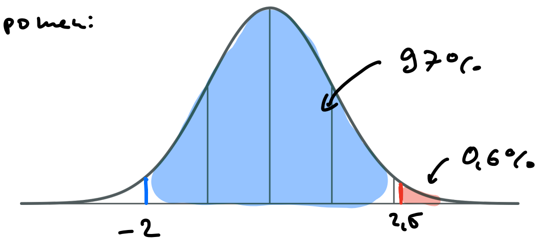
$$z = \frac{\bar{X}_A - \bar{X}_B - (\mu_A - \mu_B)}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} = \frac{\bar{X}_A - \bar{X}_B - 200\,h}{20\,h}$$

in ta je normalno porazdeljena po izreku ⇒

ⓐ razlika 160 h med $\bar{X}_A$ in $\bar{X}_B$ v "enotah z" pomeni:

$$z = \frac{160 - 200}{20} = -2$$

ⓑ razlika > 250 h $\quad z > \frac{250 - 200}{20} = 2{,}5$



Vzorčna porazdelitev variance

Iz populacije zajamemo vse možne vzorce velikosti n in izračunamo varianco za vsak tak (naključni) vzorec. Izkaže se:

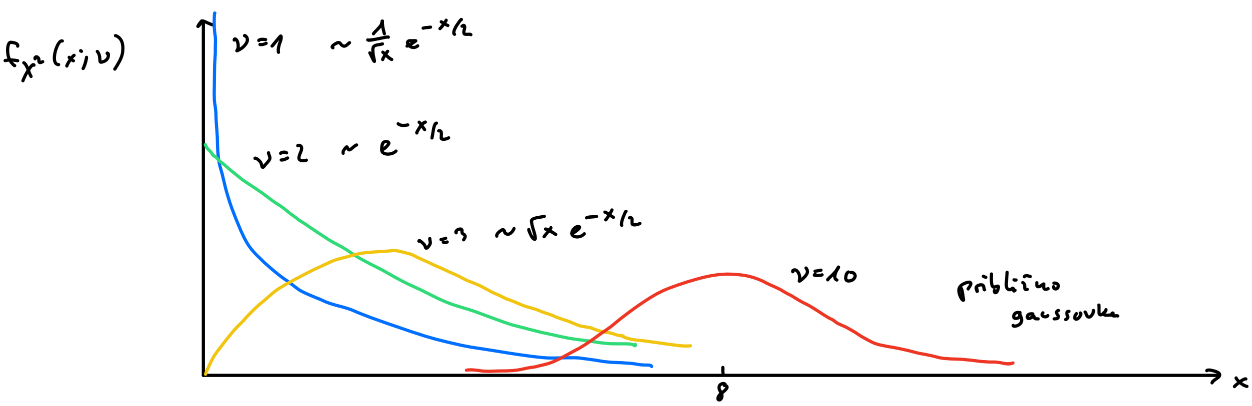
$$\chi^2 = \frac{n\,S_x^2}{\sigma^2} = n\,\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2/\sigma^2 \qquad (*)$$

hi kvadrat — χ^2; vzorčna varianca — S_x^2; prava varianca — σ^2

Izrek: če naključne vzorce zajemamo iz populacije, ki je porazdeljena normalno je ta vzorčna spremenljivka (*) porazdeljena v skladu s porazdelitvijo χ^2, z $n-1$ prostostnimi stopnjami.

$$f_{\chi^2}(x;\nu) = \begin{cases} \frac{1}{2^{\nu/2}}\frac{1}{\Gamma(\nu/2)}\, x^{\nu/2-1} e^{-x/2} & ; x > 0 \\ 0 & ; x \le 0 \end{cases}$$

gostota porazdelitve χ^2

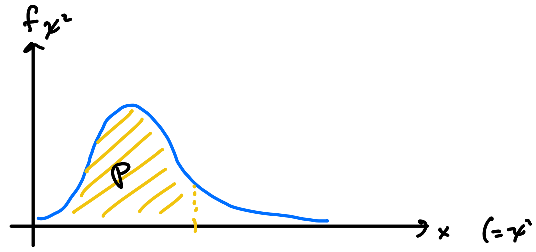


S takšno gostoto izračunamo kumulat. verjetnost:

$$P(\chi^2 \le x) = \frac{1}{2^{\nu/2}\,\Gamma(\nu/2)} \int_0^x u^{\nu/2-1} e^{-u/2}\, du$$

običajno tole predpišemo

tole je tabelirano



Vrnimo se k zgledu z znano populacijo $\{2, 3, 6, 8, 11\}$ iz katere smo nabrali $N = 25$ vzorcev velikosti $n = 2$: $\{2,2\}, \{2,3\}, \ldots$ To pot računamo vzorčne variance teh 25 vzorcev:

$$S_x^2 = \frac{1}{2}\sum_1^2 (x_i - \bar{x})^2 : \{0,\ 0.25,\ 4,\ \ldots\}$$

Povprečje teh varianc je

$$E[S_x^2] = \frac{1}{N}\sum_1^N S_x^2 = \frac{135}{25} = 5{,}40 \qquad E[S_x^2] = \frac{n-1}{n}\,\sigma^2$$

10.8

Koliko so te variance raztresene:

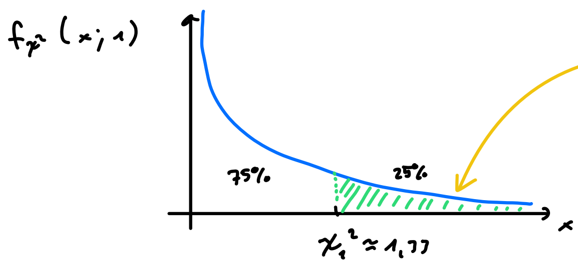
$$\mathrm{var}[S_x^2] = \frac{1}{N}\sum_{k=1}^N \left(S_{x,k}^2 - E(S_x^2)\right)^2 = 23{,}03$$

Zanima nas pričakovano število teh $N = 25$ vzorcev, pri katerih bo vzorčna varianca večja kot $S_1^2 = 7{,}2$ (ker $\nu = n - 1 = 1$)

Imamo $n = 2$, $\sigma^2 = 10{,}8$

Tvorimo spremenljivko $\chi_1^2 = \dfrac{n S_1^2}{\sigma^2} = \dfrac{2 \cdot 7{,}2}{10{,}8} \doteq 1{,}33$

Spremenljivka χ^2 je porazdeljena po $f_{\chi^2}(x; 1)$



Izjava $P(\text{vzorčna var.} \ge S_1^2)$ je enaka izjavi $P(\chi^2 \ge \chi_1^2)$

⇒ Pričakujemo ≈ 25% vzorcev z varianco ki bo več kot $S_1^2 = 7{,}2$ (torej 25% $N \doteq 6$) (dejansko jih je 6, lahko pa bi jih bilo 4, 5, 7, 8, ...)

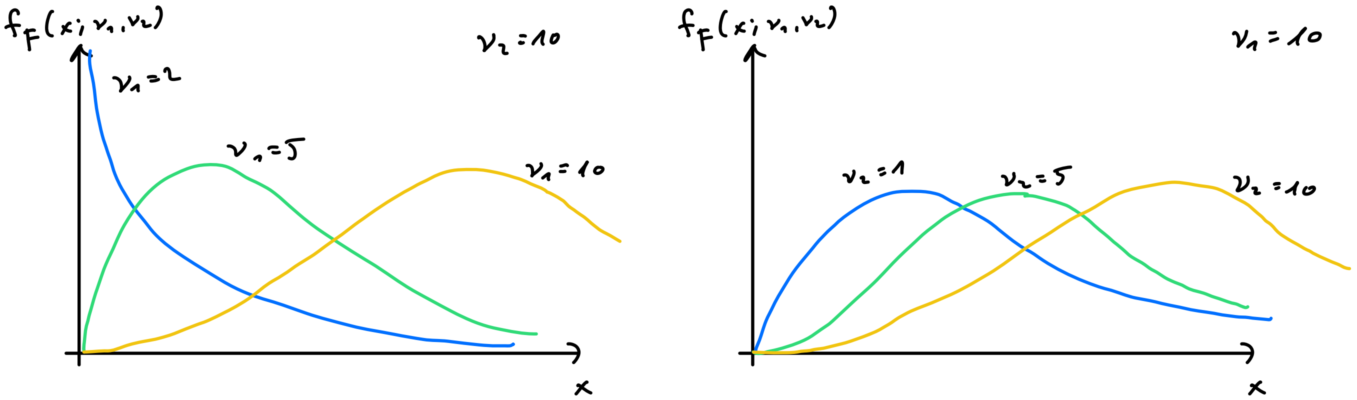
Vzorčna porazdelitev razmerja varianc

Denimo, da iz dveh normalno porazdeljenih populacij (z variancama σ_1^2, σ_2^2) vzamemo dva neodvisna vzorca velikosti n_1, n_2. Izračunamo vzorčni varianci S_1^2, S_2^2.
Tedaj je:

$$F = \frac{\frac{n_1}{n_1-1} \cdot \frac{S_1^2}{\sigma_1^2}}{\frac{n_2}{n_2-1} \cdot \frac{S_2^2}{\sigma_2^2}}$$

porazdeljena po porazdelitvi F z (n_1-1, n_2-1) prostostnimi stopnjami.
Gostota porazdelitve F:

$$f_F(x; \nu_1, \nu_2) = \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \frac{\Gamma((\nu_1+\nu_2)/2)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \, x^{\nu_1/2-1} \left(1 + \frac{\nu_1}{\nu_2} x\right)^{-\frac{\nu_1+\nu_2}{2}}$$



Zgled iz dveh populacij ($\sigma_1^2 = 20$, $\sigma_2^2 = 36$), ki sta normalno porazdeljeni, izberemo dva vzorca ($n_1 = 8$, $n_2 = 10$).

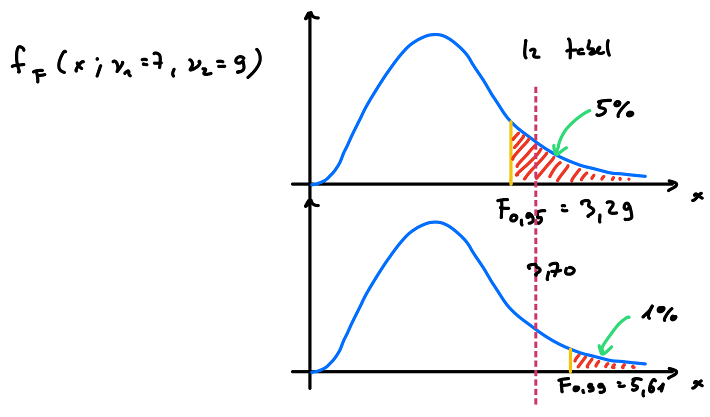
Kolikšna je verjetnost, da je varianca prvega vzorca vsaj dvakrat večja od variance drugega vzorca? (pričakujemo zelo majhno vrednost, ker je $\sigma_2^2 \approx 2\sigma_1^2$ (in ne obratno)).

Tvorimo:

$$F = \frac{\frac{n_1}{n_1-1}}{\frac{n_2}{n_2-1}} \; \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} = \frac{\frac{8}{7}}{\frac{10}{9}} \; \frac{\frac{S_1^2}{20}}{\frac{S_2^2}{36}} = 1{,}85 \, \frac{S_1^2}{S_2^2}$$

F je porazdeljen po porazdelitvi F s št. prost. st. (degree of freedom)
$\nu_1 = n_1 - 1 = 7$ $\qquad$ $\nu_2 = n_2 - 1 = 9$

Če naj bo $S_1^2 > 2 S_2^2$ (to zahteva naloga), mora biti $F > 3{,}70$



Mi pa rešimo $F_x > 3{,}70$

V resnici nekje vmes

$$P(S_1^2 > 2S_2^2) = P(F > 3{,}70) \quad \begin{matrix} >1\% \\ <5\% \end{matrix}$$

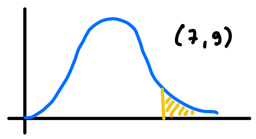
Če želimo točno številko, pa integriramo.

$$\int_0^{3{,}70} f_F(x;\, \nu_1=7,\, \nu_2=9)\, dx = 0{,}96385$$

odgovor $\approx 3{,}6\%$

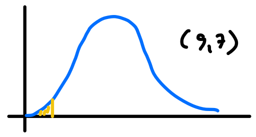
(pričakovano majhna vrednost)

Komentar: če zamenjamo vlogi "1" in "2"
(Pokljuka vs Snežnik – drevesa desno)



$n_1 = 10 \quad \sigma_1^2 = 36 \quad n_2 = 8 \quad \sigma_2^2 = 20$

Iščemo verjetnost $S_2^2 > 2S_1^2$,



dobimo $F < \frac{1}{3{,}70} \approx 0{,}270$ oz.

$$P(S_2^2 > 2S_1^2) = P\left(F < \frac{1}{3{,}70}\right) = \int_0^{0{,}27} f_F(x;\, \nu_1=9,\, \nu_2=7)\, dx = 0{,}036$$

Intervali zaupanja (confidence intervals)

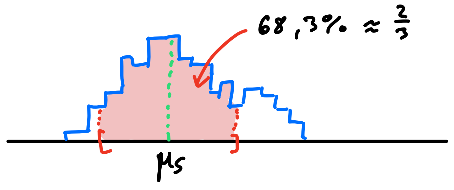
Naj bosta μ_S in σ_S^2 povprečje in varianca vzorčne porazdelitve za neko statistiko S, npr.

$$S = X \quad \text{ali} \quad S = \frac{1}{n}\sum_{i=1}^{n} X_i \,, \ \dots$$

Če je vzorčna statistika za to S približno normalna (kar velja za mnoge statistike, če je velikost vzorca $\geq$ vsaj 10), potem pričakujemo, da bo vrednost S ležala na intervalu

$$[\mu_S - \sigma_S \,,\, \mu_S + \sigma_S] \quad \text{približno} \quad 68{,}3\% \ \text{"časa"}$$

interval zaupanja
(z verjetnostjo = stopnjo zaupanja 68,3% verjamemo / zaupamo (pričakujemo, da bomo μ_S našli na intervalu $[\mu_S - \sigma_S, \mu_S + \sigma_S]$))



2σ ... 95,5%
3σ ... 99,7%

Bolj precizno izjavo za [...] ko dajemo, manj zaupamo vanjo.

Interval zaupanja za povprečje

Imamo vzorec $\{x_i\}_{i=1}^{N}$, iz smo že dobili vzorčno povprečje $\bar{x}$ in vzorčno varianco S_x^2.

Kako dobra je ocena $\bar{x}$ za pravo (populacijsko) povprečje μ..

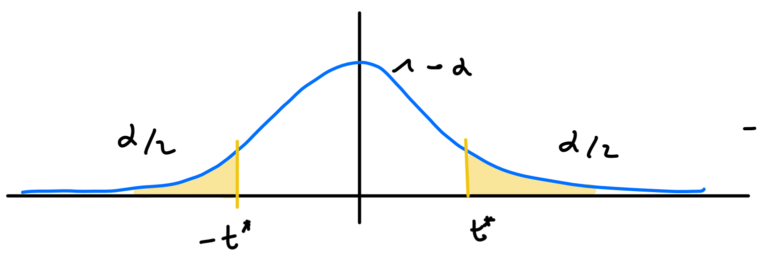
Tvorimo statistiko
$$T = \frac{\bar{x}-\mu}{\sqrt{S_x^2}}\sqrt{n}$$

Izrek: Če so $\{x_i\}$ porazdeljeni normalno po $N(\mu,\sigma^2)$, je statistika T porazdeljena po porazdelitvi t (oz. T) = po Studentovi porazdelitvi, z $n-1$ prostostnimi stopnjami. Gostota te porazdelitve

$$f_T(x,\nu) = \frac{1}{\sqrt{\nu}\,B(\nu/2, 1/2)}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2} \qquad -\infty < x < \infty$$

ν ni nujno $\mathbb{Z}$

Kvantili so tabelirani:



$$\int_{-t^*}^{t^*} f_T(x;\nu)\,dx = 1-\alpha$$

$1-\alpha$ = stopnja zaupanja
α = stopnja tveganja

Interval zaupanja dobimo iz:
$$-t_* \le \frac{\bar{x}-\mu}{S_x}\sqrt{n} \le t_*$$
hočemo izraz za μ

- Pravo povprečje populacije, iz katere smo zajeli vzorec $\{x_i\}_{i=1}^{N}$, ocenimo z

$$\mu = \bar{x}$$

in ta μ je z verjetnostjo ("zaupanjem, gotovostjo") $1-\alpha$ na intervalu zaupanja

$$\left[\bar{x} - \frac{t_* S_x}{\sqrt{n}},\ \bar{x} + \frac{t_* S_x}{\sqrt{n}}\right]$$

iz tabel

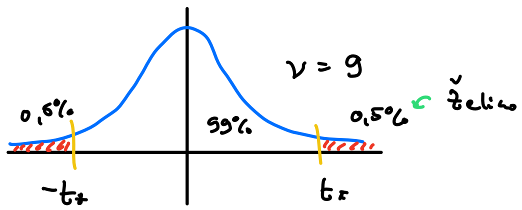
t^* dobimo iz tabel, toda za velike n ($n \gtrsim$ nekaj deset) je Studentova ~~St~~ porazdelitev že skoraj normalna

$CL = 1-\alpha$	50%	68,3%	90%	95,5%	99%	99,7%
z_* normalne porazdelitve	0,675	1 "1σ"				

Zgled Pri desetkratnem merjenju ($n=10$) premera v krogel dobimo vzorčno povprečje $\bar{X} = 4,38$ cm in std. dev. $s = 0,06$ cm. Določi interval zaupanja, na katerem je z 99% verjetnostjo dejanski premer krogel (oz. popul. povprečje).

Vzorec je razmeroma majhen $\Rightarrow t_* \neq z_* = 2,58$ (pri 99% CL) confidence level

$\nu = n-1 = 9$



Iz tabel $t_* = t_{0,995} = 3,25$

$P(-t_* < t < t_*) = 1-\alpha$

$P(t < t_*) = 1 - \frac{\alpha}{2}$

CL = 99% interval zaupanja za μ je torej

$$\left[\underset{\bar{x}}{4,38} - \underset{t_*}{3,25}\,\frac{0,06}{\sqrt{9}},\ 4,38 + 3,25\,\frac{0,06}{\sqrt{9}}\right] \text{cm}$$

če je s izračunan po pristranski formuli, je spodaj $\sqrt{n-1}$, sicer pa $\sqrt{n}$

oz. $(4,38 \pm 0,065)$ cm CL = 99%

ali npr. $(4,38 \pm 0,045)$ cm CL = 95%

Zgled Psiholog meri reakc. čas, oceni $s = \sigma = 0,05$ s. Kako velik vzorec mora narediti da so lahko 99% "gotovi", da napake v njegovi oceni za reakc. čas ne bo presegala $\Delta t = 0,01$ s

Pri 99% stopnji zaupanja sta meji intervala tako, da je $\mu = \bar{x} \pm 2,58\,\frac{\sigma}{\sqrt{n}}$ (z_* za 99%)

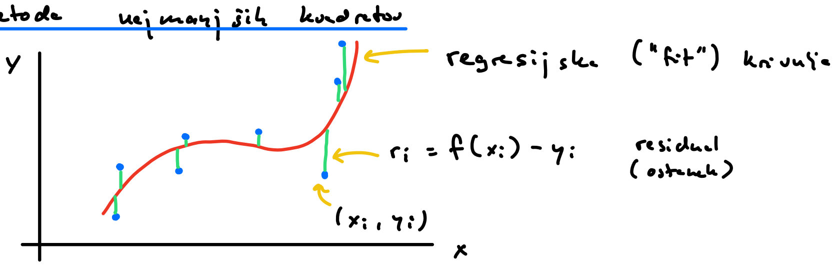
Hočem imeti $2,58\,\frac{\sigma}{\sqrt{n}} \leq \Delta t$

Prilagajanje krivulj (= regresija, fitanje)

Množica izmerkov $\{(x_i, y_i)\}_{i=1}^{n}$: v kakšni zvezi so y in x?

Oceno vrednosti y iz vrednosti x se imenuje regresija.

Metoda najmanjših kvadratov



Kako dobro se krivulja prilagaja podatkom, merimo z vsoto kvadratov odstopanj

$$\sum_{i=1}^{n} r_i^2 = \text{minimalno}$$

Zgled: prilagajanje premice

$$f(x) = ax + b \,, \quad r_i = a x_i + b - y_i$$

$$\sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (a x_i + b - y_i)^2 = S(a,b)$$

Potreben pogoj za minimum: $\frac{\partial S}{\partial a} = \frac{\partial S}{\partial b} = 0$

$$\frac{\partial S}{\partial a} = 2 \sum_i (a x_i + b - y_i)\, x_i = 0$$

$$\frac{\partial S}{\partial b} = 2 \sum_i (a x_i + b - y_i)\, 1 = 0$$

$$\Rightarrow \quad a = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}$$

$$b = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n \sum x^2 - (\sum x)^2}$$

To lahko malo predelamo

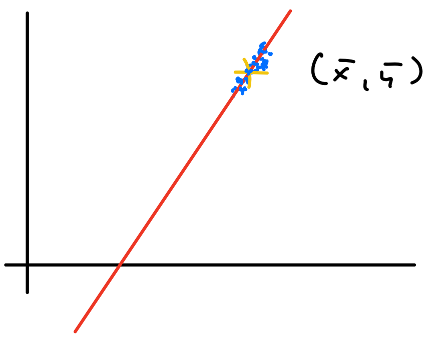
$$a = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2}$$

Smiselno je navesti $\bar{y} = a\bar{x} + b$

iz podatkov

Ni treba računati kompleksnih izrazov za a in b, ampak samo za a in potem $b = \bar{y} - a\bar{x}$

$\Rightarrow$ regresijska premica gre skozi težišče podatkov $(\bar{x}, \bar{y})$!



prestavimo $(\bar{x}, \bar{y})$ v $(0,0)$

Splošnejša linearna regresija

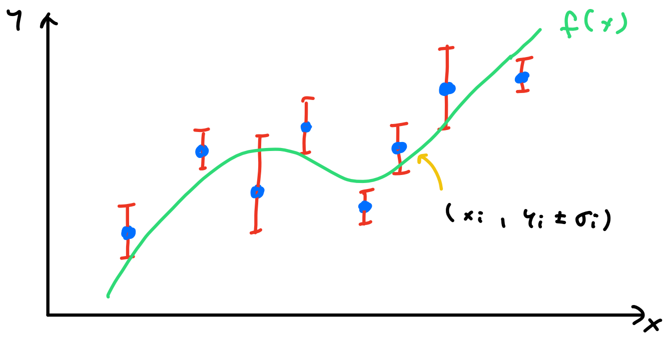
modelska funkcija $\quad f(x) = a_n x^n + \ldots + a_1 x + a_0$

še splošneje bi bilo $\quad f(x) = a_n \phi_n(x) + a_{n-1} \phi_{n-1}(x) + \ldots$

Linearna pomeni linearna v parametrih, ne v funkcijah

$a \sin x + b e^x$ $\quad$ lin. regresija

$a \sin x + b e^{cx}$ $\quad$ ni lin. regresija



Minimiziramo mero odstopanja

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - f(x_i))^2}{\sigma_i^2} = \min$$

Minimum $\quad \dfrac{\partial \chi^2}{\partial a_j} = 0 \quad \forall j \ (j = 0, 1, \ldots, n)$

Najprej trivialen primer:

Prilagajanje konstante: $\quad f(x) = a_0 = c$

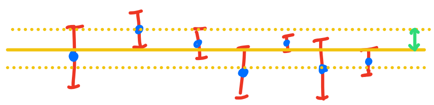
$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - c)^2}{\sigma_i^2}$$

$$\frac{\partial \chi^2}{\partial c} = 0 = -2 \sum_{i=1}^{n} \frac{x_i - c}{\sigma_i^2} \Rightarrow \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} = c \sum_{i=1}^{n} \frac{1}{\sigma_i^2}$$

$$\boxed{c = \frac{\sum w_i x_i}{\sum w_i} \qquad w_i = \frac{1}{\sigma_i^2}}$$

Uteženo povprečje

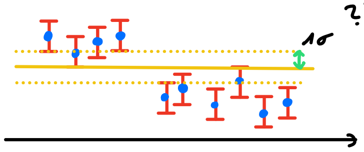
Npr.



1σ sumljivo - znotraj več kot $\frac{2}{3}$ vzorcev

?! znotraj ni $\frac{2}{3}$ vzorcev

Npr.



tehnično pravilno, a nesmiselno. Skoraj nemogoče, da bi šlo za skok statistične fluktuacije → verjetno sistematična napaka

Nazaj na splošnejši primer

$$\frac{\partial \chi^2}{\partial a_j} = 0 \quad \forall j \quad j = 0, 1, \dots m$$

Od tod sledi sistem enačb za parametre $a_m, a_{m-1}, \dots, a_1, a_0$:

$$S\vec{a} = \vec{b}$$

$\vec{a} = (a_0, a_1, \dots, a_m)^T$ ← Te iščemo

$\vec{b} = (b_0, b_1, \dots, b_m)^T$ ← Dobimo iz podatkov

$$b_k = \sum_{i=1}^{n} \frac{x_i^k \, y_i^k}{\sigma_i^2} \qquad S_{kj} = \sum_{i=1}^{n} \frac{x_i^{k+j}}{\sigma_i^2}$$

⇒ fit parametri = $\boxed{\vec{a} = S^{-1}\vec{b}}$

variance parametrov $\boxed{\mathrm{var}[a_j] = (S^{-1})_{jj}}$

$\boxed{\mathrm{cov}[a_j, a_k] = (S^{-1})_{jk}}$

DN Izpelji vse to za $f(x) = \sum_{j=1}^{m} a_m \phi_m(x)$ ← poljubne bazne funkcije (namesto x^m)

npr. $x_i^{j+k} \to$ bo postal $\phi_k(x_i)\,\phi_j(x_i)$

Prilagajanje premice $f(x) = a_1 x + a_0$

$$S = \sum_i \frac{1}{\sigma_i^2} \qquad S_x = \sum_i \frac{x_i}{\sigma_i^2} \qquad S_y = \sum_i \frac{y_i}{\sigma_i^2}$$

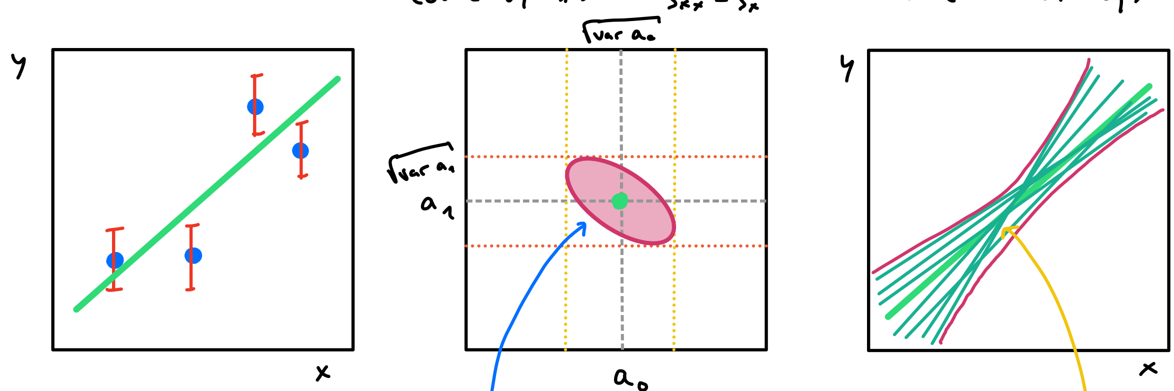
$$S_{xx} = \sum_i \frac{x_i^2}{\sigma_i^2} \qquad S_{xy} = \sum_i \frac{x_i y_i}{\sigma_i^2}$$

Presečišče z osjo y: $\quad a_0 = \dfrac{S_{xx} S_y - S_x S_{xy}}{S_{xx} S - S_x^2}$

Strmina $\quad a_1 = \dfrac{S S_{xy} - S_x S_y}{S_{xx} S - S_x^2}$

Var $\quad \mathrm{var}[a_0] = \dfrac{S_{xx}}{S_{xx} S - S_x^2} \qquad \mathrm{var}[a_1] = \dfrac{S}{S_{xx} S - S_x^2}$

$$\mathrm{cov}[a_0, a_1] = \frac{-S_x}{S_{xx} - S_x^2} \longrightarrow \rho \text{ (lin. korelacija)}$$



kovariančna krivulja

(vse točke na tej krivulji so 1σ stran od optimuma)

tole je v sistemu fit s polno vsebino