

Literatura:

Simon Širca: Verjetnost v fiziki (cena 20 eur, večino poglavij od 1 do 13)
(to je glavni vir, ki mu večinoma sledi tudi pričujoča skripta)

Ivan Kuščer, Alojz Kodre: Matematika v fiziki in tehniki (poglavje 11)

1. Temeljni pojmi pri verjetnosti

1.1 Naključni poskusi in dogodki

Fizika je empirična veda. Poskus si lahko zamislimo kot proces, ki začetno stanje spremeni v končno stanje. Seveda si želimo, da bi bil ta poskus nenaključen: ob meritvi si prizadevamo nadzorovati vse okoliščine — vhodne podatke, merski proces in analizo izhodnih podatkov — zato pričakujemo, da bomo ob vsaki ponovitvi poskusa z enakim začetnim stanjem in v enakih okoliščinah prišli do enakega rezultata.

V naključnem poskusu (angl. random experiment) se, nasprotno, lahko zgodi, da pri mnogokratnih ponovitvah poskusa z enakim začetnim stanjem in v enakih okoliščinah dobimo različne rezultate. Za naključni poskus je torej značilno, da ne moremo enolično napovedati natančnega izhodnega stanja na podlagi vhodnih podatkov. To velja za večino poskusov na atomski skali že zaradi kvantne narave: na primer, koliko jeder bo v danem času razpadlo, kakšna je gibalna količina elektrona v vodikovem atomu Velja pa tudi za številne poskuse na klasičnem nivoju: na primer, koliko plinskih molekul z dano hitrostjo se bo v določenem času zaletelo v steno posode (tu je naključnost posledica termičnega gibanja), ali število pik pri metu igralne kocke (kar je odvisno od podrobnega giba roke pri metu).

Raje se vprašamo po pogostnosti določenega končnega stanja glede na število opravljenih poskusov. Zato naj bi bilo število opravljenih poskusov čim večje: privzeli bomo, da je mogoče naključni poskus neomejeno velikokrat ponoviti.

Definicije nekaterih pojmov

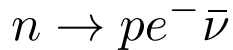
Odnose med dogodki obravnavamo v jeziku množic:

Med množicami lahko napravimo operacije unije, preseka,...

Zgled: na začetku imamo 7 nevtronov (Z),

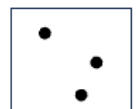
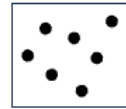
izmerimo število preostalih nevtronov po pol ure (K)

v vmesnem času so nevtroni radioaktivno razpadali



Z: 7

K: ?



vzorčni prostor (sample space) končnega stanja

$$S = \{0, 1, 2, 3, 4, 5, 6, 7\} \text{ nevtronov}$$

Nabor vseh možnih izidov (outcome) naključnega poskusa

elementaren **dogodek (event)** = posamezen izid (outcome) poskusa

primer: dogodek A: $A = \{3\}$ nevtroni

elementarne dogodke lahko združimo v sestavljen dogodek:

primer: dogodek B: manj kot 4 nevtroni $B = \{0, 1, 2, 3\}$

Dogodek A je podmnožica vzorčnega prostora S: $A \subset S$

Nemogoč dogodek $N = \{\}$ se ne more zgoditi v nobeni ponovitvi poskusa

primer N: 8 nevtronov

G je gotov dogodek se zgodi v vseh ponovitvah poskusa

primer: v našem primeru je G sestavljen dogodek $G = \{0, 1, 2, \dots, 7\}$

nasproten dogodek $\underline{B} = \{4, 5, 6, 7\}$ nevtronov : $B \cap \underline{B} = N$, $B \cup \underline{B} = G$

nezdružljiva dogodka: njun presek je prazen

poln sistem dogodkov:

primeri:

$$\{B, \underline{B}\} = \{\{0, 1, 2, 3\}, \{4, 5, 6, 7\}\},$$

{sodo nevtronov, liho nevtronov},

$$\{\{0, 4\}, \{1, 2, 3\}, \{5, 6, 7\}\}$$

$$\{\{0\}, \{1\}, \{2\}, \dots\} = S$$

Sistem dogodkov $\{A_1, \dots, A_n\}$ je poln, če se pri vsaki ponovitvi poskusa zgodi natanko eden izmed teh dogodkov. V polnem sistemu dogodkov so ti dogodki **nezdružljivi** (izključujoči) in zapolnijo vse možnosti. Izidi naključnega poskusa sestavljajo poln sistem dogodkov in vsota verjetnosti za te dogodke je enaka ena [(2.14)].

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

Operacije med množicami dogodkov: $\equiv$ označuje definicijo

produkt: $AB \equiv A \cap B$

A in B se zgodita hkrati

vsota: $A \cup B$

zgodí se vsaj eden izmed A in B

Če sta A in B nezdružljiva, se v knjigi Verjetnost uporablja posebno prikladen zapis + :
 $A+B \equiv A \cup B$ če sta A in B nerazdružljiva, kar pomeni da je njun presek nič.

Pozor: v knjigi se + uporabi tudi včasih ko sta dogodka združljiva,
takrat je na mestu previdnost, pri predavanjih se bomo tega poskusili izogniti

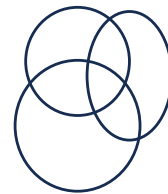
razlika: $A-B \equiv A \cap \underline{B}$

dogodek kjer se zgodi A, vendar se ne zgodi B

Zgled: izrazimo $A \cup B$ ter $A \cup B \cup C$ kot vsoto nezdružljivih členov,
saj bo to prikladno za izračun verjetnosti

$A \cup B = A + \underline{A \cap B}$

$A \cup B \cup C = A + \underline{B \cap A} + \underline{C \cap \underline{A \cup B}}$



(*) : potrebovali bomo v naslednjem koraku

Zgled Dvakrat vržemo kovanec, ki pade na glavo (g) ali cifro (c).

Vzorčni prostor tega naključnega poskusa je $S = \{cc, cg, gc, gg\}$.

Dogodek A naj bo, da se v dveh metih vsaj enkrat pojavi cifra, torej $A = \{cc, cg, gc\}$,
dogodek B pa, da pri drugem metu dobimo glavo, torej $B = \{cg, gg\}$.

Dogodek, da se obenem zgodita A in B, je $A \cap B = AB = \{cg\}$,
od koder vidimo, da se A in B ne izključujeta, sicer bi dobili $AB = \{\} = \emptyset$.

Dogodek, da se zgodi A, vendar se ne zgodi B, je
 $A - B = A \cap \underline{B} = \{cc, cg, gc\} \cap \{cc, gc\} = \{cc, gc\}$.

Dogodku A nasprotni dogodek je $\underline{A} = S - A = \{gg\}$.



1.2 Osnove kombinatorike

boste ponovili pri vajah

1.3 Verjetnost in njene lastnosti

Pri naključnem poskusu smo v dvomu, ali se bo neki dogodek zgodil ali ne. Kot mero za verjetnost, s katero smemo pričakovati določen dogodek, vzamemo kar njegovo relativno pogostnost.

$$P(A) \equiv \lim_{n \rightarrow \infty} \frac{n_A}{n} = \lim_{n \rightarrow \infty} \frac{\text{število dogodkov vrste } A}{\text{število vseh dogodkov}}$$

Izračunamo jo lahko

- “po pameti”: pri metu igralne kocke $P(A=\text{lih})=3/6=1/2$
- pragmatično: kocko vržemo tisočkrat in ugotovimo, da smo dobili liho število pik v 513 metih. Tako smo empirično določili relativno pogostnost za liho število pik $P(A=\text{lih})=513/1000=0.513$. Ta vrednost se bo seveda spremenila, če kocko vržemo še tisočkrat in še tisočkrat, na primer na 0.505, 0.477, 0.498 in tako dalje, pričakujemo pa, da se bo po velikem številu teh poskusov ustalila pri prej izračunani vrednosti 0.5.

Verjetnost $P(A)$ dogodka A v naključnem poskusu definiramo torej kot število, pri katerem se ustali relativna pogostnost dogodka A po velikem številu ponovitev poskusa.

Aksiomi verjetnosti (Kolmogorov 1933):

- $P(A)$ je realen in $0 \leq P(A) \leq 1$ za vsak dogodek A v prostoru dogodkov
- $P(G)=1$ za gotovi dogodek
- $P(A \cup B)=P(A)+P(B)$ za nezdružljiva A in B oz
 $P(A+B)=P(A)+P(B)$; zapis $A+B$ pomeni da sta dogodka nezdružljiva

Od tod sledijo seveda številne izpeljane lastnosti verjetnosti:
bistveno: verjetnosti nezdružljivih dogodkov bomo sesteli

- $P(A)+P(\underline{A})=1$ ker sta A in $\underline{A}$ nezdružljiva
- $P(A-B)=P(A)-P(B)$ ker $P(A-B)+P(B)=P(A-B+B)$ saj sta $A-B$ in B nezdružljiva in $A-B+B=A$
- $P(A \cup B) = P(A) + P(B \setminus A) = P(A) + P(B \cap \underline{A}) = P(A) + P(B) - P(AB)$
- $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$ [*]

izpeljava: $A \cup B \cup C = A + B \setminus A + C \setminus (A \cup B)$ verjetnost nezdružljivih dogodkov seštejemo

$$P(A \cup B \cup C) = P(A) + P(B \setminus A) + P(C \setminus (A \cup B))$$

$$P(C \setminus (A \cup B)) = P(C) - P(C \cap A) - P(C \cap B) + P(C \cap A \cap B) \text{ iz Vennovega diagrama}$$

Zgled (str 7):

Merlina naprava za zaznavanje kozmičnih žarkov je sestavljena iz devetih manjših, neodvisnih detektorjev, ki so vsi usmerjeni v isto smer neba.

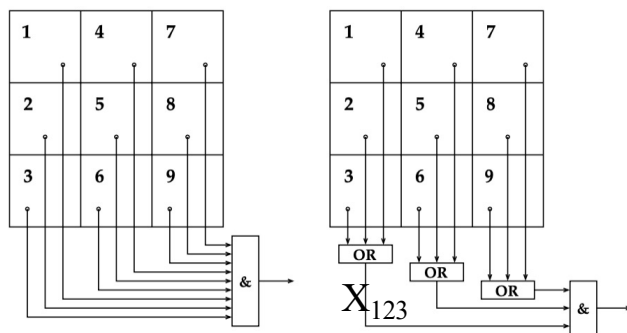
Na napravo pade pljuska; definiramo naslednje dogodke:

dogodek E: zaznava pljuska v posameznem detektorju, $P(E) = \varepsilon = 90\%$.

dogodek X: zaznava pljuska na izhodu naprave

levo: Če zahtevamo, da pljuska zaznajo vsi detektorji hkrati (vezava v devetkratno koincidenco), je celotna verjetnost za zaznavo pljuska enaka
 $P(X) = P(E)^9 \approx 0.387$

desno: Detektorje lahko zvežemo tudi v tri trojice, pri čemer ugoden izid v okviru posamezne trojice pomeni, da žarke zazna vsaj en detektor v trojici. Šele iz nastalih treh signalov nato tvorimo trojno koincidenco. V tem primeru je verjetnost $X_{1,2,3}$ da na izhodu $X_{1,2,3}$ zaznamo pljuska enaka po zvezi $[*]$ iz prejsnje strani enaka
 $P(X_{123}) = P(E1 \cup E2 \cup E3) = (3\varepsilon - 3\varepsilon^2 + \varepsilon^3) \approx 0.999$, koincidenca treh pa vodi do
 $P(X) = 0.999^3 = 0.997$



Slika 1.1 — Detektor kozmičnih žarkov. [LEVO] Vezava detektorjev v devetkratno koincidenco. [DESNO] Vezava trojic detektorjev v trojno koincidenco.

1.4 Pogojna verjetnost

Če verjetnost za dogodek A tolmačimo pri določenem naboru pogojev B, temu pravimo pogojna verjetnost $P(A|B)$. B je lahko dogodek v istem ali drugem poskusu.

$P(A|B)$ je pogojna verjetnost za A ob pogoju B

$P(A|B)$ is conditional probability of A **given condition** B

Pri n ponovitvah poskusa s povečanim naborom pogojev se dogodek B zgodi n_B -krat, dogodek $AB = A \cap B$ pa n_{AB} -krat. Relativno pogostost zdaj merimo glede na n_B , zato je pogojna verjetnost

$$P(A|B) \equiv \lim_{n \rightarrow \infty} \frac{n_{AB}}{n_B} \rightarrow P(A|B) = \lim_{n \rightarrow \infty} \frac{n_{AB}/n}{n_B/n} = \frac{P(AB)}{P(B)}$$

KomentarČ Običajno se obravnava zglede kjer sta A in B dogodka v istem poskusu. Če sta A in B dogodka v različnih poskusih, vzorčni prostor v mislih razširimo, da obravnava oba poskusa obenem in po analognem vzorcu štejemo n_{AB} ter n_B .

Ker velja tudi obratno $P(B|A) = \frac{P(AB)}{P(A)}$

obe zvezi združimo v izrek o hkratni verjetnosti dogodkov AB

$$P(AB)=P(B|A)P(A)=P(A|B)P(B)$$

Prvi enačaj ubesedimo takole: verjetnost, da se zgodita A in B hkrati, je enaka produktu verjetnosti, da se najprej zgodi A, in verjetnosti, da se zgodi B, če vemo, da se je že zgodil A. Drugi enačaj preberemo analogno.

Od tod sledi **Bayesova formula**, ki izrazi pogojno verjetnost za B ob pogoju A s tremi verjetnostmi. O pomenu in uporabi Bayesovega teorema spregovorimo v naslednjem podglavju.

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

Iz istega razloga velja:

$$P(A_n \dots A_2 A_1) = P(A_n | A_{n-1} \dots A_2 A_1) \cdot \dots \cdot P(A_3 | A_2 A_1) P(A_2 | A_1) P(A_1)$$

$$P(A_1 A_2 A_3) = P(A_3 | A_1 A_2) P(A_2 | A_1) P(A_1) = P(A_1) P(A_2 | A_1) P(A_3 | A_1 A_2)$$

Zgled (str 9, prilagojen)

V zabojniku v vinski kleti je 10 steklenic, od katerih so 4 pokvarjene.

Iz zaboja vzamemo na slepo 3 steklenice. Kolikšna je verjetnost, da so vse tri izbrane steklenice dobre ?

Označimo z A_i dogodek, da je i-ta izbrana steklenica dobra, in z A dogodek, da so dobre vse tri. Vsi dogodki so tu del istega poskusa. Verjetnost, da je dobra prva steklenica, je $P(A_1)=6/10$. Ker je bila prva steklenica dobra, je zdaj v zaboju še 9 steklenic in od teh je 5 dobrih. Verjetnost, da bo dobra tudi druga steklenica, je torej $P(A_2|A_1)=5/9$. Verjetnost da bo dobra tretja steklenica, če sta bili prvi dve dobri je $P(A_3|A_2 A_1)=4/8$.

$$P(A)=P(A_1) P(A_2|A_1) P(A_3|A_1 A_2) =(6/10) (5/9) (4/8)= 1/6$$

$$P(A_1)=6/10$$

○ ○ ○ ○ ○	dogodek A1
○ x x x x	

$$P(A_2|A_1)=5/9$$

○ ○ ○ ○ ○	dogodek A2
x x x x	

$$P(A_3|A_2 A_1)=4/8$$

○ ○ ○ ○	dogodek A3
x x x x	

1.4.1 Neodvisni dogodki

Če sta dogodka A in B **neodvisna**, je verjetnost, da se A zgodi, neodvisna od tega, ali imamo kakršno koli informacijo o B torej je $P(A|B)=P(A)$. Verjetnost, da se takšna dogodka zgodita hkrati, je torej po $P(AB)=P(A|B)P(B)$ enaka produktu verjetnosti, da se pripetita posamič: **za neodvisna dogodka torej velja $P(AB)=P(A)P(B)$.**

Zgled (str. 12, spini, dogodka A in B)

n dogodkov $A=\{A_1, A_2, \dots, A_n\}$ je **v celoti neodvisnih** ce

$$\begin{aligned} P(A_n \dots A_2 A_1) &= P(A_n | A_{n-1} \dots A_2 A_1) \cdot \dots \cdot P(A_3 | A_2 A_1) P(A_2 | A_1) P(A_1) \\ &= P(A_n) \quad \dots \quad P(A_3) \quad P(A_2) \quad P(A_1) \end{aligned}$$

in podobno velja tudi za vsako kombinacijo $(i_1, i_2, \dots, i_k)$ k-tega ($k=2, 3, \dots, n$)

Da bi bilo n dogodkov v celoti neodvisnih v splošnem ni dovolj da so paroma neodvisni.

Zgled (str 12, prilagojen)

V dveh oddaljenih eksperimentih (1 in 2) fiziki merijo projekcijo spina na danem kvantnem sistemu. Ta je pri obeh eksperimentih pred meritvijo v istem stanju, kjer ima spin z enako verjetnostjo projekcijo $+1/2$ (spin “gor”, $\uparrow$) ter $-1/2$ (spin “dol”, $\downarrow$)

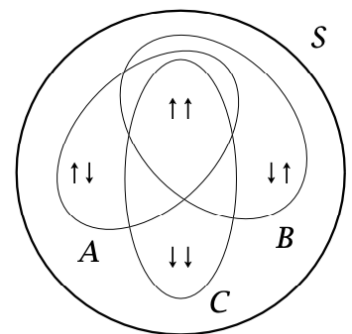
Dogodek A: “spin $\uparrow$ pri meritvi 1”

dogodek B: “spin $\uparrow$ pri meritvi 2”

dogodek C: “obe meritvi pokažeta enako projekcijo spina”.

Vpr: Ali so trije dogodki paroma neodvisni?

Vpr: Ali so trije dogodki v celoti neodvisni?



Vzorčni prostor za izmerjene pare projekcij je $S = \{\uparrow\uparrow, \uparrow\downarrow, \downarrow\uparrow, \downarrow\downarrow\}$,

izbranim trem dogodkom pa ustrezajo njegove podmnožice

$A = \{\uparrow\uparrow, \uparrow\downarrow\}$, $B = \{\uparrow\uparrow, \downarrow\uparrow\}$ in $C = \{\uparrow\uparrow, \downarrow\downarrow\}$. Od tod takoj razberemo ustrezne verjetnosti

$$P(A) = P(B) = P(C) = \frac{2}{4} = \frac{1}{2},$$

kot tudi

$$P(AB) = P(AC) = P(BC) = \frac{1}{4} \quad \text{in} \quad P(ABC) = \frac{1}{4}.$$

Ker velja

$$P(AB) = P(A)P(B) = P(AC) = P(A)P(C) = P(BC) = P(B)P(C),$$

so dogodki A, B in C paroma neodvisni. Vendar

$$P(ABC) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C),$$

torej dogodki *niso* v celoti neodvisni.

1.4.2 Več o Bayesovi formuli

Zgoraj smo izpeljali, da pogojno verjetnost $P(B|A)$ lahko izrazimo z $P(A|B)$, $P(B)$ in $P(A)$ prek Bayesove formule

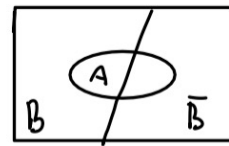
$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Brezpogojno verjetnost $P(A)$ želimo izraziti s pogojnimi verjetnostmi za dogodek A .

Kadar se neki dogodek A zgodi ob dveh različnih, med seboj izključujočih se pogojih B in $\bar{B}$, in poznamo pogojne verjetnosti za A ob obeh pogojih, lahko izračunamo tudi brezpogojno verjetnost za A , to je verjetnost dogodka samega, $P(A)$.

$$A = (A \cap B) \cup (A \cap \bar{B})$$

$$P(A) = P(AB) + P(A\bar{B}) \rightarrow$$



$$P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$$

Zgled:

Obravnavajmo najprej primer dveh pogojev ob klasičnem zavarovalniškem zgledu. Zavarovalnica razvršča voznika v dva razreda:

- boljši vozniki (B), ki jih je 85 %,
 Za te je verjetnost da povzročijo nesrečo v prvem letu 1/10
- slabši vozniki ($\bar{B}$), ki jih je 15 %.
 za te je verjetnost da povzročijo nesrečo v prvem letu 1/5.

Določimo verjetnost, da bo sklenitelj povzročil nesrečo v že prvem letu:

$$P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$$

$$P(A) = 0.1 \cdot 0.85 + 0.2 \cdot 0.15 = 0.115$$

V prvem letu se zgodi nesreča. Določimo verjetnost, da jo je povzročil "boljši" voznik:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{0.1 \cdot 0.85}{0.115} = 0.74$$

Relativno velika verjetnost ni presenečenje, saj je boljših voznikov več.

Razmislimo še o primeru, ko izključujoča pogoja nista le dva, temveč jih je več. Kadar se neki dogodek A zgodi ob različnih, med seboj izključujočih se pogojih, in poznamo pogojne verjetnosti za A ob vseh teh pogojih, lahko izračunamo tudi brezpogojno verjetnost za A , to je verjetnost dogodka samega, $P(A)$.

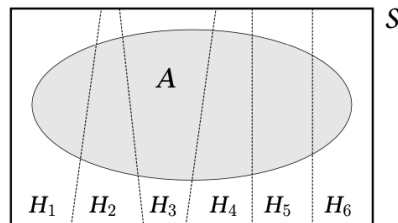
V tem primeru jim pravimo **privzetki ali hipoteze** in jih označimo s H_i :

Vse H_i sestavljajo poln sistem, ta nabor izčrpa vse možnosti

zato velja formula za popolno verjetnost (angl. total probability formula)

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

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



Slika 1.3 — Ilustracija formule za popolno verjetnost. Izključujoči se pogoji oziroma hipoteze H_i so disjunktne množice, ki razkosajo vzorčni prostor S in s tem tudi poljuben dogodek A .

Ko v zvezi z Bayesovo formulo govorimo o hipotezah in danem dogodku A imejmo v mislih, **da so hipoteze H_i obstajale pred dogodkom A** . Hipoteze dogodek A pogojujejo. Običajni razmisleki nam tipično podajo verjetnost za dogodek A ob predpostavki hipoteze H_i , torej $P(A|H_i)$. Mi pa bi radi določili kaksna je bila verjetnost za izhodiščno hipotezo ob kasnejšem izidu A , torej $P(H_i|A)$.

Znova se oprimo na zvezo (1.10), to pot na njen desni enačaj, od koder preberemo $P(H_i|A)P(A) = P(A|H_i)P(H_i)$ oziroma

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

Imenovalec tega izraza podaja enačba (1.15) in končni rezultat je slavna Bayesova formula [5]

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{P(A|H_1)P(H_1) + \dots + P(A|H_n)P(H_n)}, \quad i = 1, 2, \dots, n. \quad (1.16)$$

V naključnem poskusu smo ob določenem trenutku lahko vedno znova pričali dogodkom A , toda dogodki H_i , ki izid A pogojujejo — vsak z ustrezno verjetnostjo $P(H_i)$ — so se zgodili *pred* A . Verjetnostim $P(H_i)$ zato pravimo *apriorne*, saj so v poskusu načeloma znane vnaprej. Leva stran Bayesove formule pa meri verjetnost, da velja izhodiščna hipoteza H_i ob *kasnejšem* izidu A . Pogojno verjetnost $P(H_i|A)$ zato imenujemo *aposteriorno*, saj na podlagi trenutnega izida A podaja verjetnost za dogodek H_i , ki se je zgodil pred njim. Bayesovo formulo zato imenujemo tudi *izrek o verjetnosti hipotez*.

$P(H_i|A)$: verjetnost za izhodiščno hipotezo ob kasnejšem izidu A :

aposteriorna verjetnost za hipotezo po izidu A

$P(H_i)$: **apriorne verjetnosti za hipotezo**

Pogosto dogodek A predstavlja meritev podatkov D (data). Ustrezno **Bayesovo formulo** za verjetnost hipoteze H_i ob dani meritvi podatkov zapišemo kot

$$P(H_i|D) = \frac{P(D|H_i) P(H_i)}{\sum_j P(D|H_j) P(H_j)}$$

Dva pristopa k verjetnosti:

frekventistični pristop (frequentist view) : pri dani hipotezi H_i nas zanima pogojna verjetnost da bomo pri opravljeni meritvi dobili podatke D

Bayes pristop (Bayesian approach): po dobljeni meritvi podatkov D nas zanima verjetnost hipoteze H_i .

Zgled (str 20)

Ob krvni preiskavi nas zanima ali imamo bolezen Zahodnega Nila.

Verjetnost da naključno izbrani član celotne populacije zboli je 10^{-6}

Krvni test pokaže pozitiven izvid (visoki titer V) v 99 % primerov, v katerih je bolezen prisotna:

Krvni test pokaže negativen izvid (nizki titer N) v 95 % primerov, v katerih je bolezen ni prisotna

Hipoteza da je bolezen Nila prisotna (bolni): H_B

Hipoteza da bolezen Nila ni prisotna (zdravi): H_Z

Test nam pokaže pozitiven izvid. Kolikšna je verjetnost, da smo okuzeni s tem virusom?

Ta verjetnost je pravzaprav pogojna verjetnost da smo bolni ob pogoju da je test pokazal visoki titer, torej $P(H_B|V)$.

	zdravi: H_Z	bolni: H_B
nizki titer N	$P(N H_Z)=0.95$	$P(V H_B)=1-0.99$
visoki titer V	$P(V H_Z)=1-0.95$	$P(V H_B)=0.99$

$$P(H_B)=10^{-6}$$

$$P(H_B|V) = \frac{P(V|H_B)P(H_B)}{P(V|H_B)P(H_B) + P(V|H_Z)P(H_Z)} = \frac{0.99 \cdot 10^{-6}}{0.99 \cdot 10^{-6} + 0.05 (1 - 10^{-6})} \approx 1.98 \cdot 10^{-5}$$

motivacija za porazdelitve: $P(\text{dogodek})$ bi radi izrazili z $P(\text{številsko vrednost})$; slednje nam bodo podajale verjetnostne porazdelitve. A najprej obravnavajmo porazdelitve na splošno.

2. PORAZDELITVE IN VERJETNOSTNE PORAZDELITVE

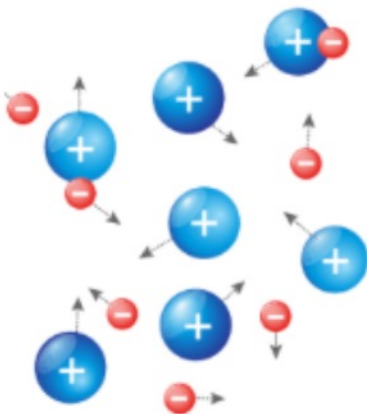
Literature: MAFI 11.1

Preden se posvetimo verjetnostnim porazdelitvam, si ogledimo porazdelitev kake dobro znane fizikalne količine, na primer mase

porazdelitev mase
v zvezdni lupini
od "delec" obravnavamo
kot "diskretno"

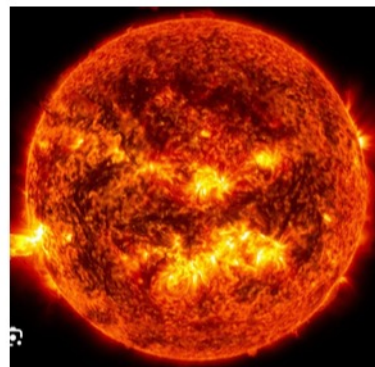


na skali μm je
relevantna diskretna
porazdelitev pline
ioniziranih atomov v plazmi

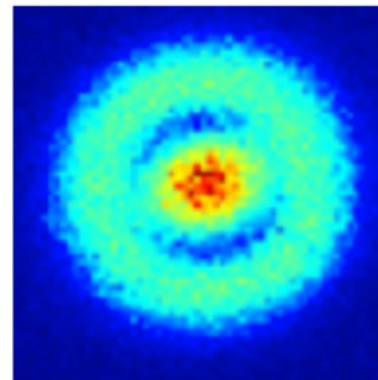


Vendar: sleherni
opis (zvezni/diskretni)
je poenostavljen
matematični model.

bližje zvezde je
relevantna zvezna
porazdelitev v zvezdi



na skali mm
je relevantna
zvezna porazdelitev
znotraj delujoč
ioniziranih atomov



Delta je prikladna za idealizacijo točkastega telesa.

2.1 DIRACOVA DELTA "funkcija" / porazdelitev

Pri diskretni porazdelitvi mase imamo v mislih končno maso točkastega (neskončno majhnega) delca, drugje pa je gostota enake nič, če opis teke porazdelitve je prikladne Diracove delte. V 1D primeru ko je točkasta masa pri $x=0$

$$\delta(x) = \lim_{\sigma \rightarrow 0} \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$$

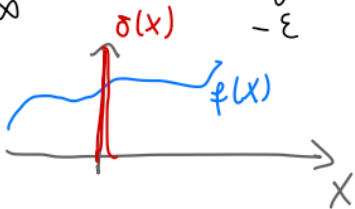
$$\text{ali } \delta(x) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \text{ (rectangular pulse of width } \Delta \text{ and height } 1/\Delta \text{)} = \delta(x)$$

Ker v limiti $\sigma \rightarrow 0$ oz. $\Delta \rightarrow 0$ vrednost pri $x=0$ divergira, $\delta(x)$ ni funkcija v pravem pomenu besede.

Zares smiselne rezultate do jela ko maso poob integriramo:

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

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



Diracova delta $\delta(x)$ je neke vrste funkcional, saj slike med prostori funkcij f in zalogo vrednosti $f(0)$

$$\delta : f \rightarrow f(0)$$

$$\text{prostor funkcij} \rightarrow \mathbb{R}$$

Funkcija delta izlušči tisto vrednost funkcije f kjer je argument delte enak nič.

1D:

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

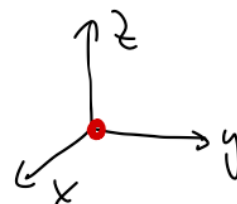
se pogosto privzame kot definicij delte.



$$[\delta(x)] = 1/[x]$$

enote: $\delta(x) = \frac{1}{m}$, $\delta(t) = \frac{1}{t}$ ker $\int \delta(x) dx = 1$

3D: $\delta(\vec{r}) \equiv \delta(x) \delta(y) \delta(z)$ $[\delta(\vec{r})] = \frac{1}{m^3}$



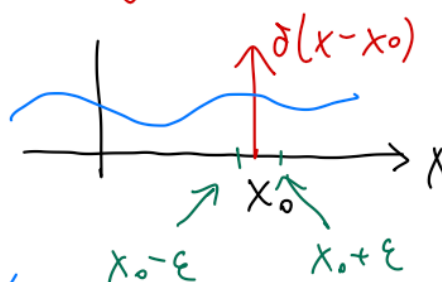
$$\int \delta^3(\vec{r}) d\vec{r} = \int \int \int \delta(x) dx \delta(y) dy \delta(z) dz = 1$$

Izpeljene lastnosti delte funkcije/poredelitve

• $\int f(x) \delta(x-x_0) dx = f(x_0)$
če integriramo preko x_0

Dz: $x_0 \pm \epsilon$ "

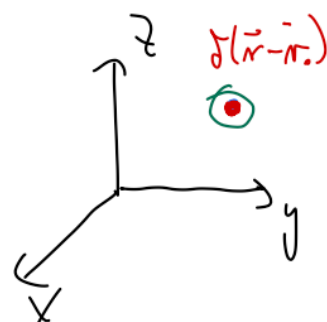
$$\int_{x_0-\epsilon}^{x_0+\epsilon} f(x) \delta(x-x_0) dx = f(x_0) \underbrace{\int \delta(x-x_0) dx}_1$$



• $\int f(\vec{r}) \delta(\vec{r}-\vec{r}_0) d\vec{r} = f(\vec{r}_0)$
če integriramo preko $\vec{r}_0$

"

Dz $\int f(\vec{r}) \delta(\vec{r}-\vec{r}_0) d\vec{r} = f(\vec{r}_0) \underbrace{\int \delta(\vec{r}-\vec{r}_0) d\vec{r}}_1$
O majhne okolice okoli $\vec{r}_0$



Funkcija delta izlušči tisto vrednost funkcije f kjer je argument delte enak nič.

$$\delta(ax) = \frac{1}{|a|} \delta(x)$$

$$\int_{-\infty}^{\infty} \delta(ax) dx \geq 0 \text{ ker } \delta(ax) \geq 0 \text{ povsod}$$

$$Dz: \int_{-\infty}^{\infty} \delta(ax) dx = \begin{cases} a > 0: \int_{-\infty}^{\infty} \delta(y) \frac{1}{a} dy = \frac{1}{a} \\ a < 0: \int_{\infty}^{-\infty} \delta(y) \frac{1}{a} dy = -\frac{1}{a} \end{cases} = \frac{1}{|a|} = \int_{-\infty}^{\infty} \frac{1}{|a|} \delta(x) dx$$

$$y = ax \quad dy = a dx$$

$x = -\infty \rightarrow y = \infty$ če $a < 0$: mejnava mej

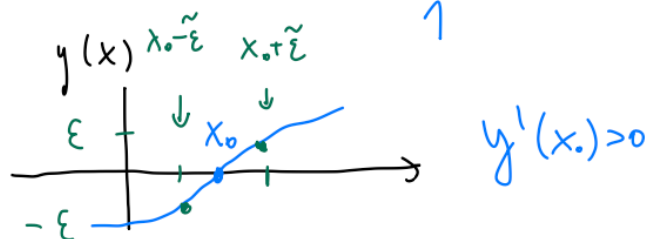
$$\delta(A\vec{x}) = \frac{1}{|\det(A)|} \delta(\vec{x})$$

$A = \text{matrice}$

$$Dz': \int \delta(\underbrace{A\vec{x}}_{\vec{y}}) dV(\vec{x}) = \int \delta(\vec{y}) dV(A^{-1}\vec{y}) = |\det A^{-1}| \int \delta(\vec{y}) dV(\vec{y})$$

JACOBIAN

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



$$y'(x_0) > 0$$

Če ima $y(x)$ zgolj eno ničlo: x_0

$$\int_{x_0 - \tilde{\epsilon}}^{x_0 + \tilde{\epsilon}} \delta(x - x_0) dx = \int_{x_0 - \tilde{\epsilon}}^{x_0 + \tilde{\epsilon}} \delta(y) dy = \int_{x_0 - \tilde{\epsilon}}^{x_0 + \tilde{\epsilon}} \delta(y(x)) y'(x) dx = y'(x_0) \int_{x_0 - \tilde{\epsilon}}^{x_0 + \tilde{\epsilon}} \delta(y(x)) dx$$

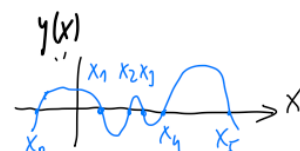
$y = y(x), \quad dy = y'(x) dx$ posledni odvod

PRIMERJAMO LEVO IN
DESNO STRAN, KER JE
ISTA INT. SPR. (X) IN
KER SO ISTE MEJE ($x_0 - \tilde{\epsilon}, x_0 + \tilde{\epsilon}$)

To je bilo če $y'(x_0) > 0$, če $y'(x_0) < 0$ pa bi
obrnili mejo in bi prišlo $|y'(x_0)|$.

$$\delta(y(x)) = \sum_{i=0}^n \frac{\delta(x - x_i)}{|y'(x_i)|}$$

$$\int f(x) \delta(y(x)) dx = \sum_{i=0}^n \frac{f(x_i)}{|y'(x_i)|}$$



Funkcija delta izlušči tisto vrednost funkcije f kjer je argument delte enak nič.

$$\delta(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dx e^{-ikx} \quad (**)$$

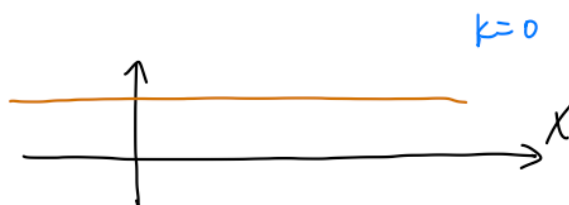
$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} dx [\cos kx - i \sin kx]$$



Če $k \neq 0$ se + in - prispevki pokrepijo v mejah $[-\infty, \infty]$

NEOBVEZNI DODATEK

Tole argument, reno menigostno:



Če $k = 0$:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} dx \rightarrow \infty$$

normalizacije nismo utemeljili
Nejlažje se jo utemelji prek
Fourierove analize (Net IV):

že tiste, ki so tu spoznali Fourierov integral: NEOBVEZNI DODATEK!

$$f(x) = \frac{1}{2\pi} \int dk e^{ikx} \left[\int_{-\infty}^{\infty} d\tilde{x} e^{-ik\tilde{x}} f(\tilde{x}) \right]$$

$F(k)$: Fourierova transformiranka $f(x)$
v prostor valovnih vektorjev k

če $f(x) = 1$ preverimo (*):

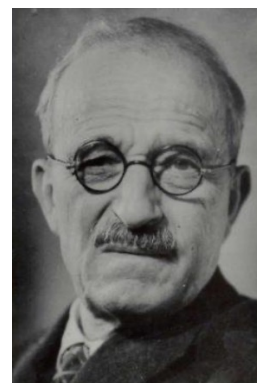
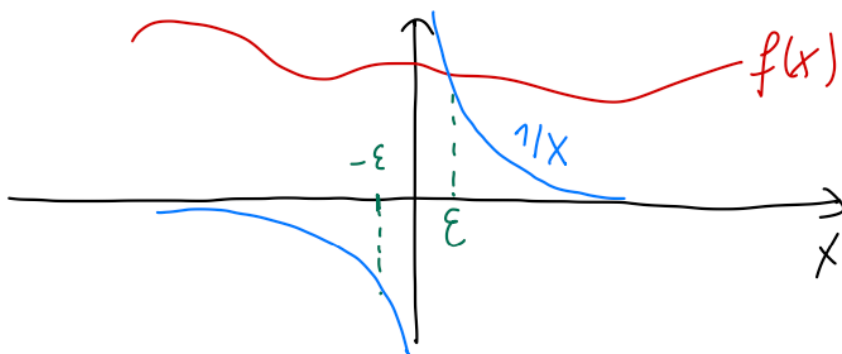
$$\frac{1}{2\pi} \int dk e^{ikx} \underbrace{\int_{-\infty}^{\infty} d\tilde{x} e^{-ik\tilde{x}} \cdot 1}_{2\pi \delta(k)} = \int dk e^{ikx} \delta(k) = 1 \quad \checkmark$$

neobvezni dodatek

• $\frac{1}{x \pm i\varepsilon} = \mp i\pi \delta(x) + \mathcal{P} \frac{1}{x}$ Izrek SOKHOTSKI PLEHLEJ
v smislu $\varepsilon \rightarrow 0$

ima smisel ko obe strani integriramo

$$\lim_{\varepsilon \rightarrow 0} \int \frac{f(x)}{x \pm i\varepsilon} dx = \mp i\pi f(0) + \lim_{\varepsilon \rightarrow 0} \int_{|x| > \varepsilon} \frac{f(x)}{x} dx$$



Josip Plemelj

(1873-1967)

Za Plemlja je izrek ključen korak pri reševanju Riemann-Hilbertovega problema (1908)

P: glavne (posplošne, principielne) vrednosti int.
Cauchy principle value:

Nekateri integrale, ki niso definirani v rigoroznem smislu, A principielno vrednostjo mepravimo regularne tako da se izognemo integraciji okoli singularnosti na simetričen način.

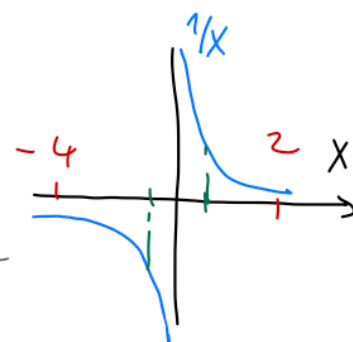
$$\int_{-4}^2 \frac{1}{x} dx = \int_{-4}^{-\varepsilon} \frac{1}{x} dx + \underbrace{\int_{-\varepsilon}^{\varepsilon} \frac{1}{x} dx}_{\text{neregularno; uzmimo pa me prispeve}} + \int_{\varepsilon}^2 \frac{1}{x} dx$$

$$\mathcal{P} \int_{-4}^2 \frac{1}{x} dx = \int_{-4}^{-\varepsilon} \frac{1}{x} dx +$$

$$= \ln(\varepsilon/4) +$$

$$+ \int_{\varepsilon}^2 \frac{1}{x} dx$$

$$\ln(2/\varepsilon) = \ln(2/4) = -\ln(2)$$

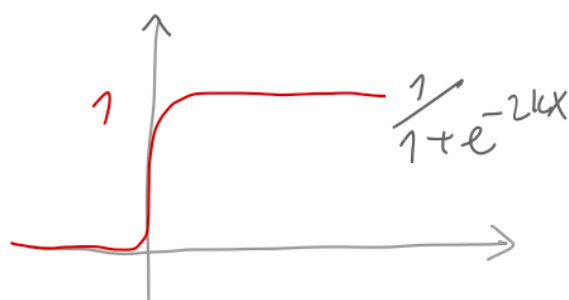
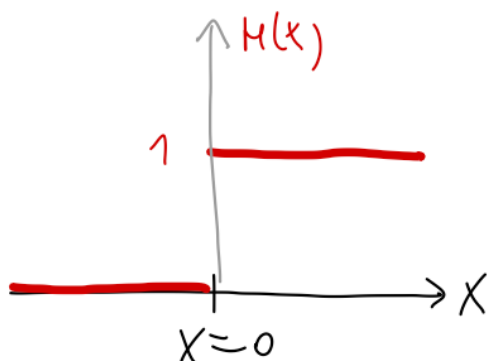


Integral dx/x na intervalu $[-\varepsilon, \varepsilon]$ ali pa tudi na intervalu $[-1, 1]$ NI Darbujevo integrabilen, in tudi NI Riemanovo integrabilen

2.2 Heavisidove stopničaste funkcija

$$H(x) \equiv \int_{-\infty}^x \delta(x) dx = \begin{cases} 0 & x < 0 \\ 1 & x > 0 \end{cases}$$

Primočni analitični približek: $H(x) = \lim_{k \rightarrow \infty} \frac{1}{1 + e^{-2kx}}$



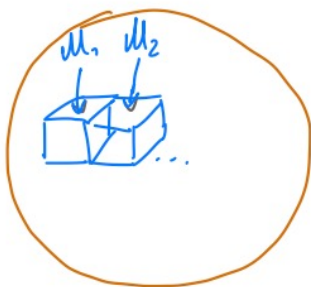
2.3 DISKRETNE IN ZVEŽNE PORAZDELITVE (na splošno) distribution

Vpeljimo porazdelitev na primeru poraza. More predstavljamo si telo, ki smo ga razdelili na več delov. Vsak del je določen z množico točk $\mathcal{U}_i$, z objema in volumen $V(\mathcal{U}_i)$ in ima maso $m(\mathcal{U}_i)$. Mase in funkcija kroja, njene bistvene lastnosti so, da je aditivne

$$m(\mathcal{U}_1 \cup \mathcal{U}_2) = m(\mathcal{U}_1) + m(\mathcal{U}_2) \text{ če } \mathcal{U}_1 \cap \mathcal{U}_2 = \emptyset$$

3D PORAZDELITVE

ZVEŽNA PORAZDELITEV MASE
v zvezdi



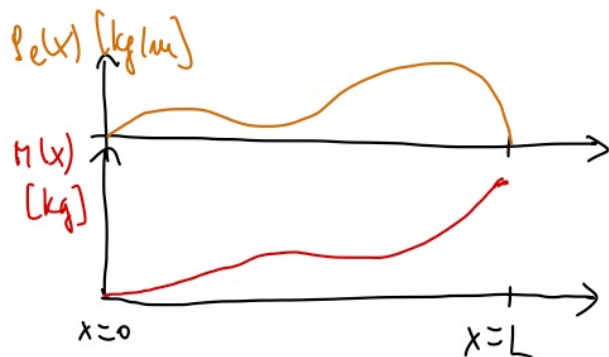
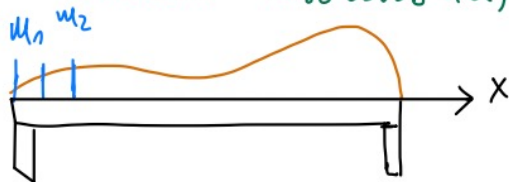
$$\rho(\mathbf{r}_i) = \frac{dm(\mathbf{r}_i)}{dV(\mathbf{r}_i)} = \rho(\mathbf{r})$$

za infinitesimalno majhne $\mathbf{r}_i \rightarrow 0$
V zgornji enačbi, $\frac{dm}{dV}$ NI ODOVOD ker
m ni funkcija V, $m \neq m(V)$!

1D porazdelitev: tu najlažje
ZVEŽNA PORAZDELITEV MASE

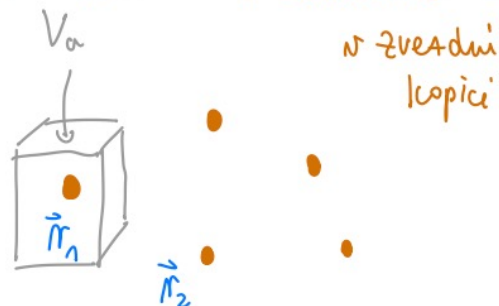
$$\rho(x) = \frac{dm(\mathbf{r}_i)}{dx(\mathbf{r}_i)} = \frac{dn(x)}{dx}$$

TO NI ODOVOD TO JE ODOVOD $n(x)$



$H(x)$: masa peske med int $[0, x]$:
KUMULATIVNA FUN., pišemo z velikimi črkami:
 $H(x) = m([0, x]) = \int_0^x \rho(\tilde{x}) d\tilde{x}$

DISKRETNA PORAZDELITEV MASE



$$\mathbf{r}_i = \{\vec{r}_i\}, \quad m(\mathbf{r}_i) = m_i, \quad dV(\mathbf{r}_i) \rightarrow 0$$

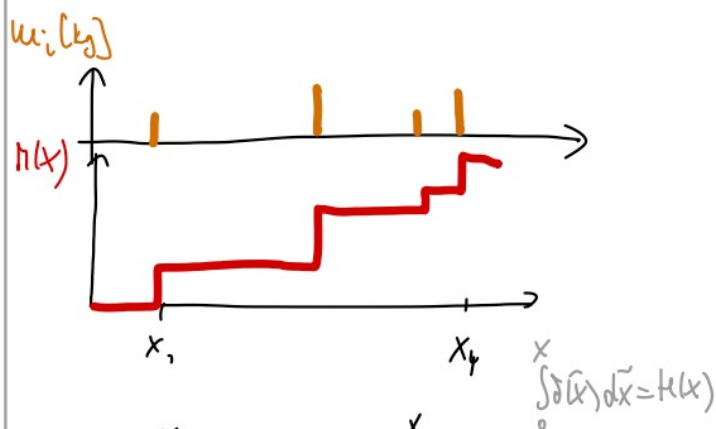
$$\rho(\mathbf{r}_i) = \frac{m(\mathbf{r}_i)}{dV(\mathbf{r}_i)} \rightarrow \rho(\vec{r}) = \sum m_i \delta(\mathbf{r} - \vec{r}_i)$$

$$\int_{V_0} \rho(\vec{r}) d\vec{r} = m_1 \underbrace{\int \delta(\vec{r} - \vec{r}_1) d\vec{r}}_1 = m_1$$

uvvedemo tudi KUMULATIVNE por.

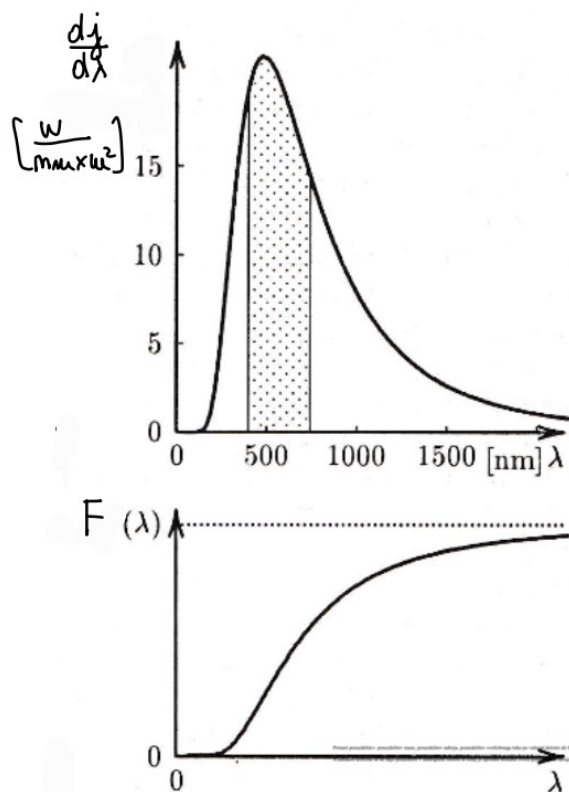
me maske
DISKRETNA PORAZDELITEV MASE

$$\rho_e(x) = \sum_i m_i \delta(x - x_i)$$

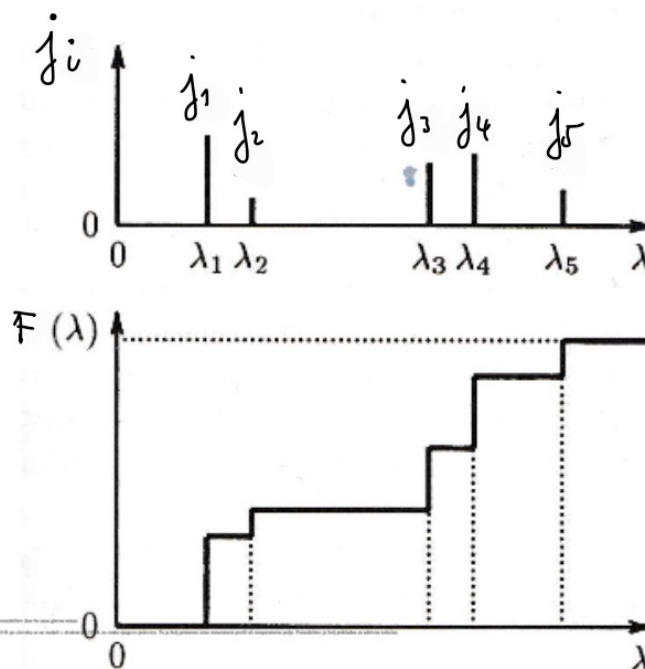


$$\begin{aligned} H(x) &= \int_0^x \rho_e(\tilde{x}) d\tilde{x} = \int_0^x \sum_i m_i \delta(\tilde{x} - x_i) d\tilde{x} \\ &= \sum_i m_i H(x - x_i) = \sum_{x_i < x} m_i \end{aligned}$$

ZVEZNA PORAZDELITEV TOKA svetlobe po λ



DISKRETNA PORAZDELITEV TOKA svetlobe po λ



porazdelitev

$$\frac{dJ}{d\lambda}$$

$$\frac{dJ}{d\lambda} = \sum J_i \delta(\lambda - \lambda_i)$$

kumulativna
porazdelitev

$$F(\lambda) = \int_0^{\lambda} \frac{dJ}{d\lambda} d\lambda$$

$$F(\lambda) = \sum_{\lambda_i \leq \lambda} J_i$$

$$[J] = \frac{W}{m^2 \cdot s}, \quad [\delta(\lambda)] = \frac{1}{m}$$

Primeri porazdelitev: porazdelitev mase, porazdelitev naboja, porazdelitev svetlobnega toka po valovni dolžini ali frekvenci, porazdelitev sile po površini telesa, verjetnostna porazdelitev (kar bo naša glavna tema). Porazdelitev je prikladna za aditivne količine.

A nekatere količine se ne dajo porazdeliti v analognem smislu in tedaj je uporaba besedne zveze porazdelitev zavajajoča. Temperaturna porazdelitev je nesmiselna: temperature 310 K po človeku se ne razdeli v dvakrat po 155 K za vsako njegovo polovico. Tu je bolj primeren izraz temperaturni profil ali temperaturno polje. Porazdelitev je bolj prikladna za aditivne količine.

VERJETNOSTNE PORAZDELITVE

$m(\mathcal{M}_1 + \mathcal{M}_2) = m(\mathcal{M}_1) + m(\mathcal{M}_2)$ ce $\mathcal{M}_1 \cap \mathcal{M}_2 = \{\}$ mase je porazdeljena po delih telesa, M_i : i-ti del telesa
 $P(\mathcal{A}_1 + \mathcal{A}_2) = P(\mathcal{A}_1) + P(\mathcal{A}_2)$ ce $\mathcal{A}_1 \cap \mathcal{A}_2 = \{\}$ verjetnost je porazdeljena po dogodkih, A_i : i-ti dogodek

2.5 Enorazsežne (1D) diskretne verjetnostne porazdelitve

X: naključna spremenljivka, ki lahko zavzame končno število različnih diskretnih vrednosti
Naključne spremenljivke bomo označevali z velikimi tiskanimi črkami.

$x_i, i=1, \dots, n$: **številski vrednosti naključne spremenljivke X**,
te vrednosti označujemo z malo črko; verjetnosti namreč želimo izraziti kot funkcije števil x , ki predstavljajo elemente vzorčnega prostora

Bolj formalno: Spremenljivka X vsakemu izidu iz vzorčnega prostora S priredi število x .

Zgledi

*) Pri metu ene igralne kocke z vzorčnim prostorom

$$S = \left\{ \begin{array}{|c|} \hline \cdot \\ \hline \end{array}, \begin{array}{|c|} \hline \cdot \cdot \\ \hline \end{array}, \begin{array}{|c|} \hline \cdot \cdot \cdot \\ \hline \end{array}, \begin{array}{|c|} \hline \cdot \cdot \cdot \cdot \\ \hline \end{array}, \begin{array}{|c|} \hline \cdot \cdot \cdot \cdot \cdot \\ \hline \end{array}, \begin{array}{|c|} \hline \cdot \cdot \cdot \cdot \cdot \cdot \\ \hline \end{array} \right\}$$

nam naključna spremenljivka X ="število pik na kocki" sporoči številski izid poskusa $X(\begin{array}{|c|} \hline \cdot \cdot \\ \hline \end{array}) = x_3 = 3$

*) Pri opazovanju plinskih molekul v posodi je vzorčni prostor hitrosti precejšen (to ni primer s končnim vzorčnim prost.)
Spremenljivka V ="velikost hitrost molekule" nam sporoči številski izid meritve $V(\text{green circle}) = v = 300 \text{ m/s}$

$P_i = P(X = x_i) = f_X(x_i)$: **Verjetnost, da ima v določeni ponovitvi poskusa X vrednost x_i**

f_X je **diskretna verjetnostna porazdelitev** (discrete probability distribution)
oziroma **verjetnostna funkcija** (probability function)

Lastnosti f_X : $P_i > 0 \rightarrow f_X(x) > 0$, za poln sistem dogodkov x_i ($i=1, \dots, n$): $\sum_{i=1}^n f_X(x_i) = 1$

Smiselno je definirati tudi verjetnost, da naključna spremenljivka X zavzame vrednost, ki je manjša ali enaka od neke vrednosti x . To verjetnost podaja

$F_X(x) = P(X \leq x)$, $-\infty < x < \infty$: F_X je **kumulativna porazdelitvena funkcija**

$$F_X(x) = \sum_{x_i \leq x} P_i = \sum_{x_i \leq x} f_X(x_i).$$

enota $[P] = [F] = 1$ (vedno), $[f] = 1$ (za diskretno)

Zgled: met dveh igralnih kock: X =vsota pik iz obeh kock

$$f_X = \sum_i P_i \delta(x - x_i)$$

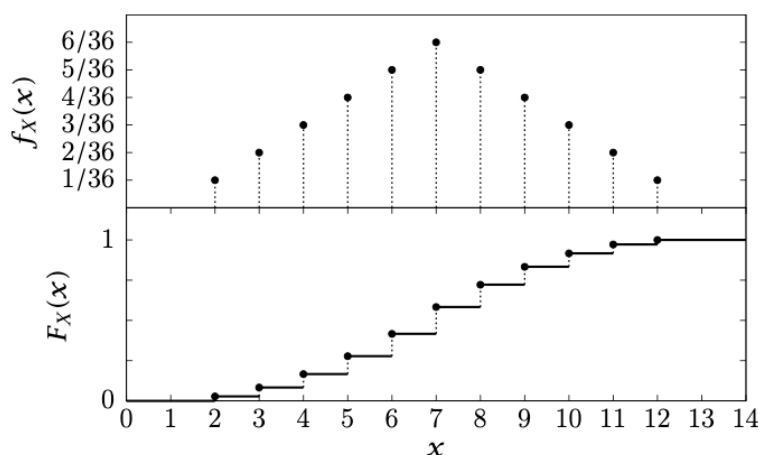
$n=36$

$x=2$: $1/36$ 1+1

$x=3$: $2/36$ 2+1, 1+2

..

$x=12$: $1/36$ 6+6



$F(1) = P(X \leq 1) = 0$

$F(2) = P(X \leq 2) = P_2 = 1/36$

$F(3) = P(X \leq 3) = P_2 + P_3 = 1/36 + 2/36$

..

$F(12) = 1$

Lastnosti kumulativne prazdelitvene funkcije:

nepadajoča

$$\lim_{x \rightarrow -\infty} F_X(x) = 0, \quad \lim_{x \rightarrow \infty} F_X(x) = 1.$$

2.6 1D zvezne verjetnostne porazdelitve

X : naključna spremenljivka, ki lahko zavzame zvezne vrednosti

x : vrednosti naključne spremenljivke X (lahko v omejenem območju $[x_{\min}, x_{\max}]$ ali na neomejenem $[-\infty, \infty]$)

Pri zveznih verjetnostnih porazdelitvah nikoli ne moremo govoriti o "verjetnosti, da zvezna spremenljivka X zavzame vrednost x ". Ta verjetnost bi bila nična. Smiselno je govoriti o "verjetnosti da zavzame spremenljivka X vrednost x na intervalu $[x_1, x_1 + \Delta x]$ ali na intervalu $[x_1, x_2]$.

$$dP(x \leq X \leq x + dx) = f_X(x) dx : f_X : \text{verjetnostna gostota (probability density function)}$$

$f_X(x) dx$ podaja verjetnost da je vrednost spremenljivke X na intervalu $[x, x+dx]$

Ker je infinitizemalen interval dx , je tudi verjetnost infinitizemalna

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f_X(x) dx$$

$$\text{Lastnosti } f_X: \quad f_X(x) \geq 0, \quad \int_{-\infty}^{\infty} f_X(x) dx = 1. \quad \text{enote: } [P]=[F]=1 \rightarrow [f]=1/[x]$$

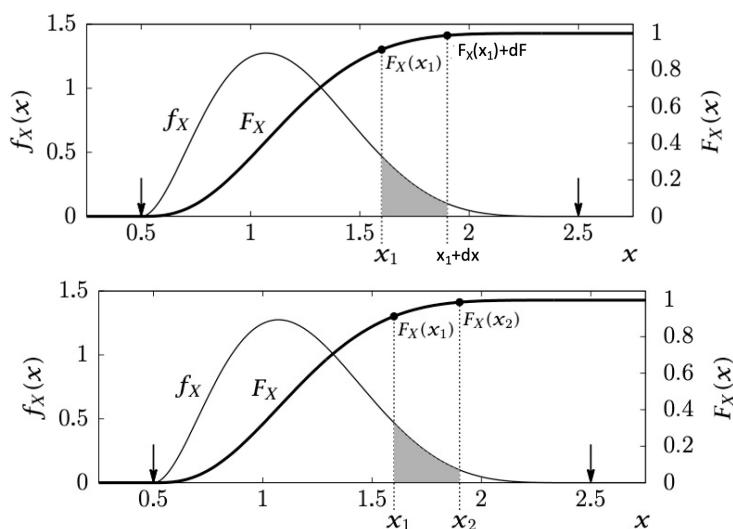
Iz "delčkov verjetnosti" $f_X(\tilde{x}) d\tilde{x}$ dobimo verjetnost, da je $X \leq x$

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(\tilde{x}) d\tilde{x}$$

F_X : kumulativna porazdelitvena funkcija

Odvod integrala po zgornji meji je enak integrandu, zato sta verjetnostna gostota in kumulativna funkcija povezani preko spodnje zveze, kar se razbere tudi za $x_2 = x_1 + \Delta x$ iz slike

$$f_X(x) = \frac{dF_X}{dx} = F'_X(x)$$



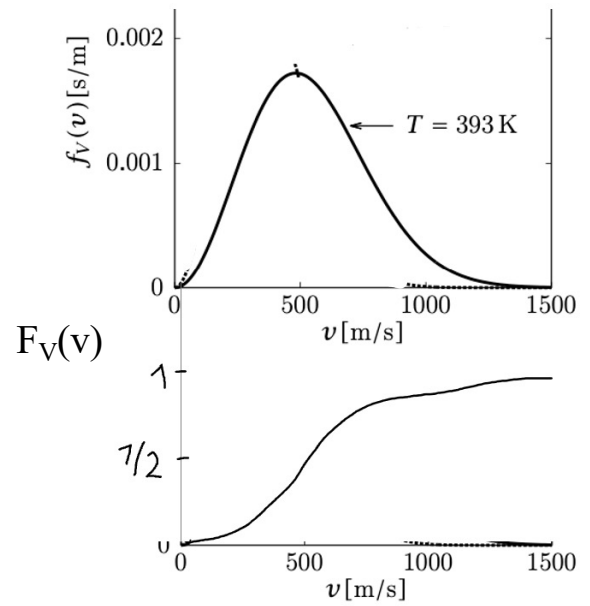
Slika 2.5 — Primer normirane verjetnostne gostote f_X (tanka krivulja, leva ordinata) za zvezno porazdelitev verjetnosti, ki je od nič različna samo na intervalu $[0.5, 2.5]$ (puščici) in ustrezna porazdelitvena funkcija F_X (odebeljena krivulja, desna ordinata). Verjetnost $P(x_1 \leq X \leq x_2)$ je enaka površini senčenega območja pod krivuljo funkcije f_X in obenem razliki vrednosti $F_X(x_2) - F_X(x_1)$.

Zgled: V =velikost hitrosti N_2 molekule v posodi s plinov pri $T=393\text{ K}$ (ima vlogo X)

v =vrednost velikosti hitrosti (ima vlogo x)

$$f_V(v) = \frac{dF_V}{dv} = \left(\frac{m}{2\pi k_B T} \right)^{3/2} 4\pi v^2 \exp\left(-\frac{mv^2}{2k_B T}\right)$$

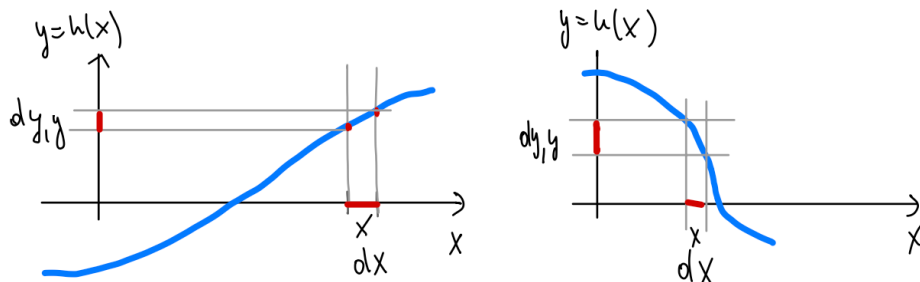
$$[F]=1, [f]=1/[v]=1/(\text{m/s})$$



2.7 Transformacije naključnih spremenljivk (1D)

Naključni spremenljivki X naj ustreza porazdelitev z gostoto f_X . Določimo porazdelitev f_Y naključne spremenljivke Y , ki je znana funkcija spremenljivke X : $Y=h(X)$, $y=h(x)$

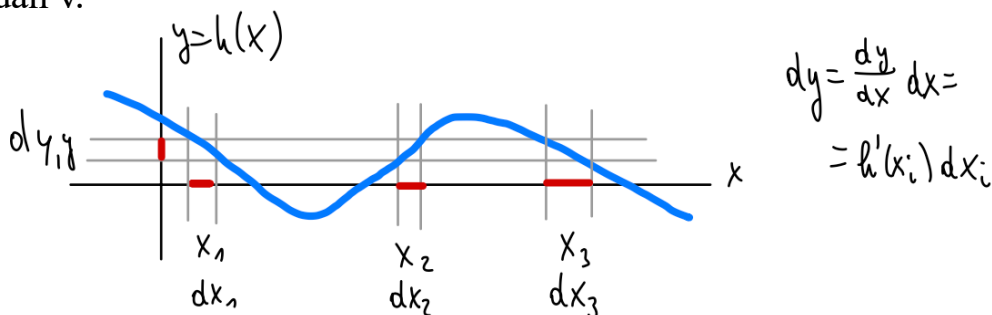
Če je funkcija h na celotnem območju x injektivna (monotona, ves čas naraščajoča ali padajoča): to je trivialen primer: verjetnost dP , da je x na intervalu $[x, x+dx]$ je enaka verjetnosti na ustreznem preslikanem intervalu $[y, y+dy]$ spremenljivke y



$$dP = f_X(x) |dx| = f_Y(y) |dy|, \text{ absolutni predznak ker so } P, f_X, f_Y \text{ po definiciji } \geq 0$$

$$\begin{aligned} f_Y(y) &= \text{vse izraženo z } y & \text{vse izraženo z } x & f_X(x) \left| \frac{dx}{dy} \right| = f_X(x) \frac{1}{\left| \frac{dy}{dx} \right|} & h(x)=y \\ &= f_X(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right| = f_X(x) \frac{1}{|h'(x)|} & x=h^{-1}(y) \end{aligned}$$

Če je funkcija h NI injektivna, več vrednosti x se preslika v določeno vrednost y , nas pa zanima verjetnostna gostota f_Y pri taki vrednosti y . Tedaj moramo sešteti po okolicah vseh x_i , ki se preslikajo v dan y .

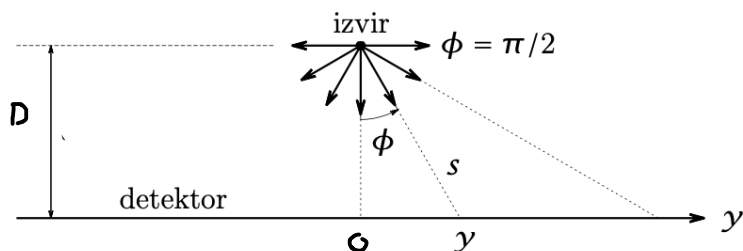


$$dP = \sum_{x_i} f_X(x_i) |dx_i| = f_Y(y) |dy|$$

Ker inverz funkcije h v tem primeru ne obstaja ("funkcija" $h^{-1}(y)$ ki pri določenem y vodi do več vrednosti x pravzaprav ni funkcija), se opremo na izrek o inverzni preslikavi, ki pravi, da za odprte podmnožice območja okoli x_i lahko tvorimo inverzno funkcijo $g_i = h^{-1}$

$$\begin{aligned} f_Y(y) &= \text{vse izraženo z } y & \text{vse izraženo z } x & \sum_{x_i} f_X(x) \left| \frac{dx}{dy} \right| = \sum_{x_i} f_X(x) \frac{1}{\left| \frac{dy}{dx} \right|} & h(x)=y \\ &= \sum_{x_i} f_X(g_i(y)) \left| \frac{dg_i(y)}{dy} \right| = \sum_{x_i} f_X(x) \frac{1}{|h'(x)|} & \text{lokalno okoli } x_i: x=h^{-1}(y)=g_i(y) \end{aligned}$$

Zgled Obravnavajmo ravninski problem točkastega radioaktivnega izvira in linearnega detektorja (na primer dolge tanke elektrode v žični komori za sledenje delcev), ki je od izvira na najbližjem mestu oddaljen za razdaljo d (slika 2.6).

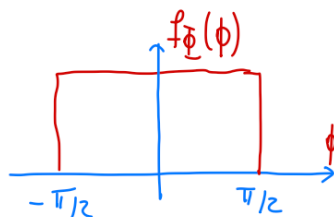


Slika 2.6 — Ravninski problem s točkastim radioaktivnim izvirom in neskončno dolgim tankim detektorjem. Izotropno sevanje izvira (enakomerna porazdelitev po kotih ϕ) je vzdolž detektorja (koordinata y) porazdeljeno po Cauchyjevi porazdelitvi.

$f_{\Phi}(\phi) = ?$

$$f_{\Phi}(\phi) = \frac{1}{\pi}, \quad \Phi = [-\pi/2, \pi/2]$$

$$\int_{-\pi/2}^{\pi/2} f_{\Phi}(\phi) d\phi = 1$$



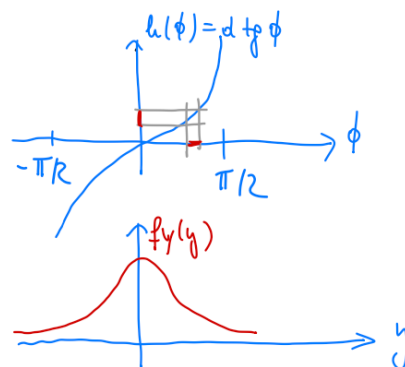
$f_Y(y) = ?$

$y = D \tan \phi = h(\phi) : h \text{ bijektivna na } \phi \in [-\pi/2, \pi/2]$

$y_0 = 0$

$$h^{-1} = \arctan \frac{y}{D}$$

$$f_Y(y) = \frac{1}{\pi} \cdot \left| \frac{d}{dy} \arctan \left(\frac{y}{D} \right) \right| = \frac{1}{\pi} \frac{1}{D^2 + y^2}$$

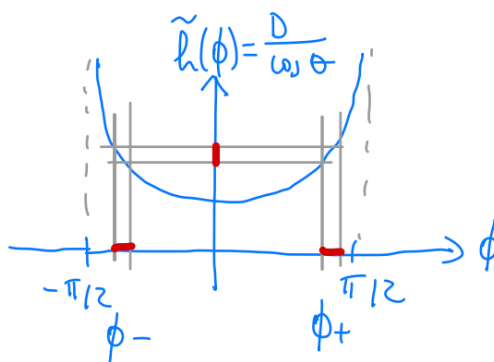


medenič uporabimo drugo oznako

$f_S(s) = ?$

$$s = \frac{D}{\cos \phi} = \tilde{h}(\phi)$$

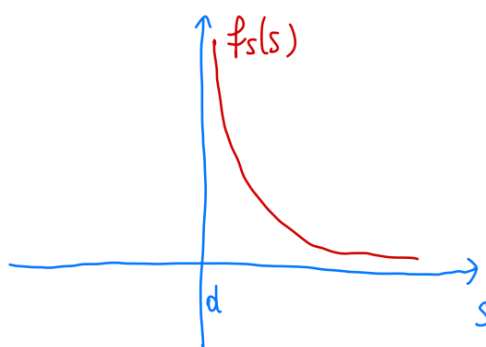
-) $\tilde{h}$ ni bijektivna na $\phi \in [-\pi/2, \pi/2]$
-) razdelimo na dve bijektivni območji: $\pm$
-) $\phi > 0$: $\tilde{h}^{-1}(s) = \arccos \frac{D}{s} = g(y)$ je bijektivna za $\phi \in [0, \pi/2]$



$$f_S(s) = \sum_{\phi_+, \phi_-} f_{\Phi}(g(s)) \left| \frac{dg}{ds}(s) \right| =$$

$$= 2 \times \frac{1}{\pi} \times \left| \frac{d}{ds} \arccos \frac{D}{s} \right| =$$

$$= 2 \cdot \frac{1}{\pi} \cdot \frac{D}{s \sqrt{s^2 - D^2}}, \quad \int_0^{\infty} f_S(s) ds = 1$$



2.8 in 2.9: DVOBRZSEŽNE PORAZDELITVE

(komentarji, besedilo preberite iz knjige)

Pri 2D porazdelitvah imamo opravke z dvema neenjainimi spremenljivkama, npr. X in Y , in zanju skupno (joint) verjetnost

DISKRETNA 2D PORAZDEL.

skupne verjetnostne funkcije je $f_{X,Y}$
(joint probability function)

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

verjetnost da ima X vrednost x
in Y vrednost y

Denimo da X zavzema vr. $\{x_1, \dots, x_m\}$
 Y $\{y_1, \dots, y_n\}$

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

lastnosti: $f_{X,Y} > 0$, $\sum_{i=1}^m \sum_{j=1}^n f_{X,Y}(x_i, y_j) = 1$

verjetnost za $X=x_i$ brez kakršne koli zahteve za Y

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

žrgonsko previmo, da smo marginalizirali po y in dobili marginalno oz. robno porazdelitev po x .

ŽVEŽNA 2D PORAZDEL.

skupne verjetnostne gostote $f_{X,Y}$
(joint probability density function):

$$\begin{aligned} dP(X=[x, x+dx], Y=[y, y+dy]) &= \\ &= f_{X,Y}(x,y) \cdot dx dy \end{aligned}$$

pove verjetnost dP da je X med $[x, x+dx]$ in Y med $[y, y+dy]$

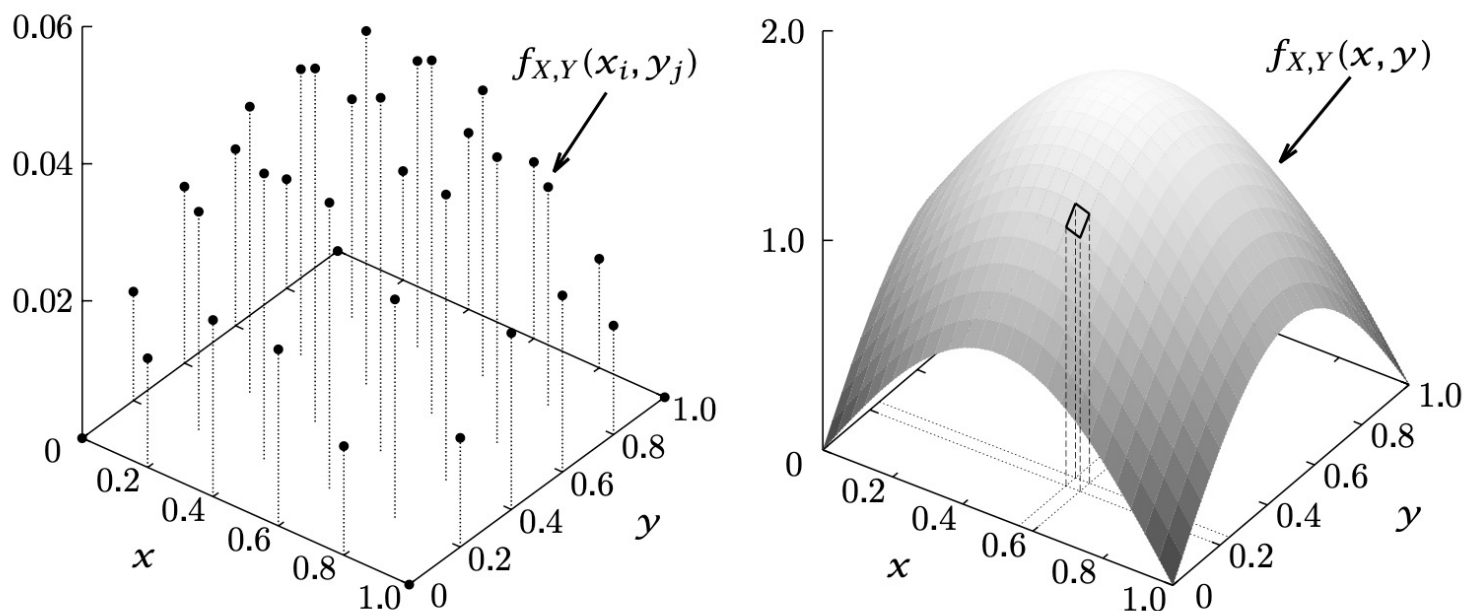
verjetnost da je točno $X=x, Y=y$ je pa še vedno nič.

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

verjetnost da je X med $[x, x+dx]$ brez zahteve za Y je

$$dP = \underbrace{f_X(x)}_P dx = \left[\int_{-\infty}^{\infty} \underbrace{f_{X,Y}(x,y)}_P dy \right] dx$$

verjetnostno gostoto - po X smo dobili z marginalizacijo po y



Slika 2.9 — [LEVO] Diskretna dvorazsežna porazdelitev verjetnosti s skupno verjetnostno funkcijo $f_{X,Y}(x_i, y_j) = \mathcal{N}(\sin \pi x_i + \sin \pi y_j)$. [DESNO] Zvezna dvorazsežna porazdelitev s skupno verjetnostno gostoto $f_{X,Y}(x, y) = \mathcal{N}(\sin \pi x + \sin \pi y)$.

verjetnost za $Y=y_j$ brez kakršne koli zahteve za X

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

žerjousko previmo, da smo marginalizirali po x in dobili marginalno oz. robno porazdelitev po y

Ležnostti f_X, f_Y :

$$\sum_{i=1}^m f_X(x_i) = \sum_{j=1}^m f_Y(y_j) = 1$$

Neključni spr. X in Y ste neodvisni če

$$P(X=x, Y=y) = P(X=x)P(Y=y) \text{ od koder}$$

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

2D skupne kumulativne fun. pove verjetnost da $X \leq x$ in $Y \leq y$

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

Pogojna verjetnost da se $X=x_i$ ob pogoju, da se je zgodil $Y=y_j$ je

$$P(X=x_i | Y=y_j) = \frac{f_{X,Y}(x_i, y_j)}{f_Y(y_j)} = f_{X|Y}(x_i, y_j)$$

$$dP = f_Y(y)dy = \left[\int_{-\infty}^{\infty} f_{X,Y}(x, y) dx \right] dy$$

Neključni spr. X in Y ste neodvisni če

$$\begin{aligned} d^2P(X=[x, x+dx], Y=[y, y+dy]) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy \\ &= \underbrace{dP(X=[x, x+dx])}_{f_X(x) dx} \underbrace{dP(Y=[y, y+dy])}_{f_Y(y) dy} \end{aligned}$$

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

$$F_{X,Y}(x, y) =$$

$$P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(\tilde{x}, \tilde{y}) d\tilde{x} d\tilde{y}$$

Pogojna verjetnost da bo $X=[x, x+dx]$ ob pogoju da je $Y=[y, y+dy]$

$$d^2P(X=[x, x+dx] | Y=[y, y+dy]) = \frac{f_{X,Y}(x, y) dx dy}{f_Y(y) dy} = f_{X|Y}(x, y) dx$$

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

2.11.4 Neodvisnost naključnih spremenljivk v dveh razsežnostih

Denimo, da sta koordinata q in hitrost p nekega telesa naključni, da vselej ležita med dvema skrajnima vrednostma q_1 in q_2 oziroma p_1 in p_2 , ter da smo pri njuni meritvi uzrli porazdelitev točk v faznem prostoru (q, p) , kot jo prikazuje prva slika. Zanima nas delež točk v faznem prostoru, ki ustreza desetodstotnemu odstopanju od premice

$$p = p_1 + 2 \frac{\Delta p}{\Delta q} (q - q_1),$$

kjer je $\Delta q = q_2 - q_1$ in $\Delta p = p_2 - p_1$. Premica je narisana s črtkano črto, pogoj

$$\left| \frac{p - p_1}{\Delta p} - 2 \frac{q - q_1}{\Delta q} \right| \leq 0.1$$

pa je označen senčeno. V brezdimenzijskih spremenljivkah $X = (Q - q_1)/\Delta q$ in $Y = (P - p_1)/\Delta p$ ima pogoj obliko $|Y - 2X| \leq 0.1$ in je prikazan na spodnji sliki.

Seveda lahko dogodke v označenem območju kar preštujemo, vendar se poskušajmo domisliti preprostega modela. Postopno naraščanje gostote točk od q_1 proti q_2 in od p_1 proti p_2 nas napeljuje na misel, da utegne biti v ozadju porazdelitev zveznih spremenljivk X in Y s skupno verjetnostno gostoto

$$f_{X,Y}(x, y) = \begin{cases} Cxy & ; \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1, \\ 0 & ; \quad \text{sicer,} \end{cases}$$

kjer je C konstanta. Porazdelitev normiraj, nato pa izračunaj ① $P(|Y - 2X| \leq 0.1)$, torej verjetnost, da sta vrednosti spremenljivk X in Y omejeni na senčeni pas na sliki. ② Izračunaj še enorazsežni verjetnostni gostoti $f_X(x)$ in $f_Y(y)$. Ali sta spremenljivki X in Y odvisni? (Ne pozabi, da je to samo model!)

 Skupno gostoto najprej normiramo:

$$1 = \int_0^1 dx \int_0^1 dy f_{X,Y}(x, y) = C \int_0^1 x dx \int_0^1 y dy = C \frac{x^2}{2} \Big|_0^1 \frac{y^2}{2} \Big|_0^1 = \frac{C}{4} \Rightarrow C = 4.$$

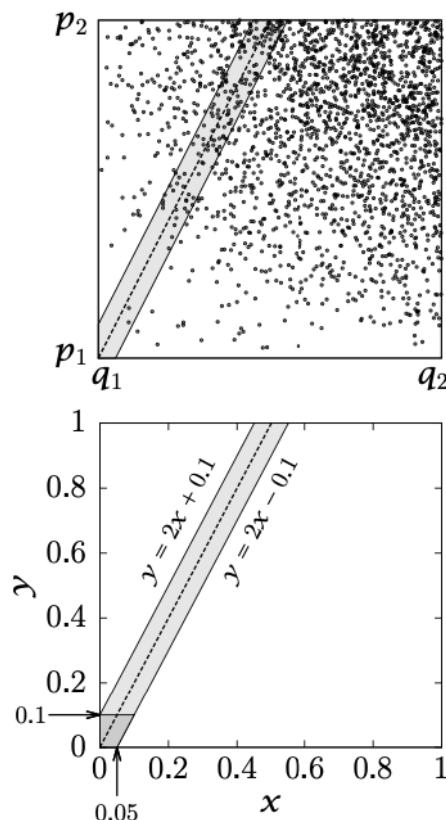
① Iskano verjetnost dobimo z integralom verjetnostne gostote po dveh območjih: po temnejši senčenem območju, določenem s pogojem $0 \leq x \leq (y + 0.1)/2$ in $0 \leq y \leq 0.1$, in po svetlejšem območju, določenem z $(y - 0.1)/2 \leq x \leq (y + 0.1)/2$ in $0.1 \leq y \leq 1$:

$$\begin{aligned} P(|Y - 2X| \leq 0.1) &= \int_0^{0.1} dy \int_0^{(y+0.1)/2} f_{X,Y}(x, y) dx + \int_{0.1}^1 dy \int_{(y-0.1)/2}^{(y+0.1)/2} f_{X,Y}(x, y) dx \\ &= \frac{1}{2} \left[\int_0^{0.1} y(y + 0.1)^2 dy + \int_{0.1}^1 0.4y^2 dy \right] \approx 0.0667. \end{aligned}$$

② Spremenljivki X in Y sta neodvisni, kajti

$$\begin{aligned} f_X(x) &= \int_0^1 f_{X,Y}(x, y) dy = 4x \int_0^1 y dy = 2x, \quad 0 \leq x \leq 1, \\ f_Y(y) &= \int_0^1 f_{X,Y}(x, y) dx = 4y \int_0^1 x dx = 2y, \quad 0 \leq y \leq 1, \end{aligned}$$

torej res velja $f_{X,Y}(x, y) = 4xy = f_X(x)f_Y(y) = 2x \cdot 2y$.



2.11.6 Porazdelitev maksimalnih in minimalnih vrednosti

Naj bodo $X_1, X_2, \dots, X_n$ neodvisne in identično porazdeljene zvezne naključne spremenljivke. ① Kako je porazdeljena njihova maksimalna vrednost

$$U = \max\{X_1, X_2, \dots, X_n\}?$$

(Po porazdelitvi spremenljivke U se seveda lahko sprašujemo, ker je tudi sama naključna.) Izpelji splošen izraz in ga uporabi v primeru, da vsem spremenljivkam X_i ustrezajo verjetnostne gostote $f_X(x) = \lambda e^{-\lambda x}$, kjer je $x \geq 0$, $\lambda > 0$.

② Kako je porazdeljena *minimalna* vrednost takšnih spremenljivk,

$$V = \min\{X_1, X_2, \dots, X_n\}?$$

 Nalogo ① rešimo z uporabo porazdelitvenih funkcij. Da bi bila maksimalna vrednost vseh X_i manjša od nekega x , morajo biti vsi X_i hkrati manjši od x . Torej

$$F_U(x) = P(U \leq x) = P(X_1 < x, X_2 < x, \dots, X_n < x) = [P(X_1 \leq x)]^n = [F_X(x)]^n.$$

Za posamezen eksponentno porazdeljen X je

$$P(X \leq x) = F_X(x) = \int_0^x f_X(t) dt = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x},$$

torej je

$$F_U(x) = [1 - e^{-\lambda x}]^n.$$

Do iskane verjetnostne gostote, ki ustreza spremenljivki U , nas loči le še odvod:

$$f_U(x) = F'_U(x) = n\lambda e^{-\lambda x} [1 - e^{-\lambda x}]^{n-1}.$$

2.10 Transformacije naključnih spremenljivk v dveh ali več razsežnostih

Spoznali smo, kako se transformirajo verjetnostne gostote pri transformaciji spremenljivk $y=h(x)$ v 1D, zdaj bi to radi posplošili na transformacijo v D razsežnostih

$\vec{X} \in \mathbb{R}^D$, $\vec{Y} \in \mathbb{R}^D$, med vektorjema naključnih spremenljivk X in Y slika vektorska funkcija $\vec{h}$
 Tokrat imejmo zaradi preprostosti v mislih le bijektivne preslikave.

D različnih naključnih spremenljivk $\vec{X} = (X^{(1)}, \dots, X^{(D)})$ z vrednostmi $\vec{x} = (x^{(1)}, \dots, x^{(D)})$

D različnih naključnih spremenljivk $\vec{Y} = (Y^{(1)}, \dots, Y^{(D)})$ z vrednostmi $\vec{y} = (y^{(1)}, \dots, y^{(D)})$

$$\vec{y} = \begin{pmatrix} y^{(1)} \\ \vdots \\ y^{(D)} \end{pmatrix} = \begin{pmatrix} h_1(x^{(1)}, \dots, x^{(D)}) \\ \vdots \\ h_D(x^{(1)}, \dots, x^{(D)}) \end{pmatrix} = \vec{h}(\vec{x})$$

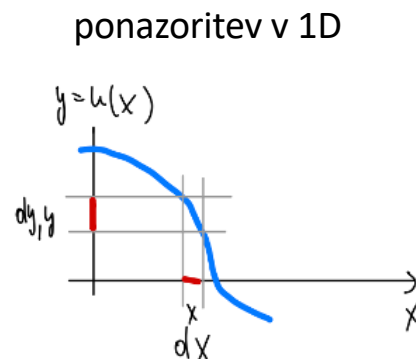
Iz znane verjetnostne gostote $f_{\vec{X}}$ bi radi določili verjetnostno gostoto $f_{\vec{Y}}$

Verjetnost $d^n P$ da je

$\vec{X}$ v n dimenzionalnem volumskem elementu $d\vec{x}$ je enaka verjetnosti,
 da je $\vec{Y}$ v preslikanem volumskem elementu $d\vec{y}$

$$d^n P = f_{\vec{X}}(\vec{x}) |d\vec{x}| = f_{\vec{Y}}(\vec{y}) |d\vec{y}|$$

$$f_{\vec{Y}} = f_{\vec{X}} \frac{1}{\left| \frac{d\vec{y}}{d\vec{x}} \right|} = f_{\vec{X}} \frac{1}{|J_{\vec{h}}|} = f_{\vec{X}} |J_{\vec{h}^{-1}}|$$



Tu je J_h Jacobijeva matrika totalnega odvoda

$$J_h(\mathbf{x}) = \det \left(\frac{\partial \mathbf{h}}{\partial \mathbf{x}}(\mathbf{x}) \right) \quad \frac{\partial \vec{h}}{\partial \vec{x}}(\vec{x}) = \begin{pmatrix} \frac{\partial h_1}{\partial x^{(1)}} & \cdots & \frac{\partial h_1}{\partial x^{(D)}} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_D}{\partial x^{(1)}} & \cdots & \frac{\partial h_D}{\partial x^{(D)}} \end{pmatrix}$$

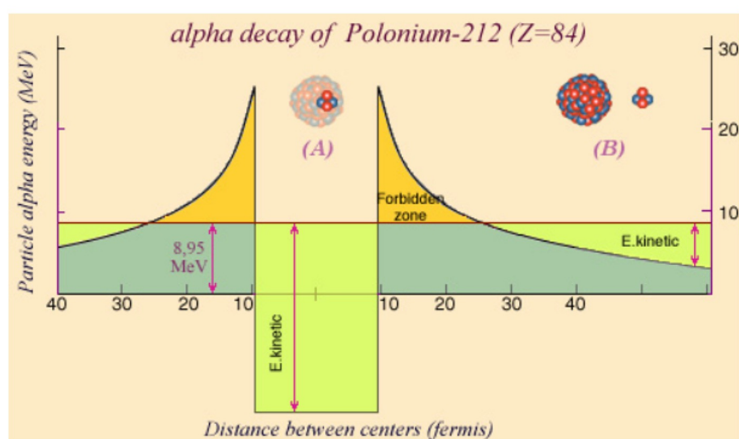
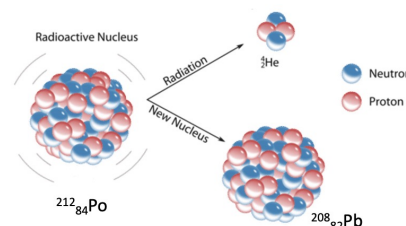
3. Posebne zvezne verjetnostne porazdelitve

Pogosto uporabne zvezne verjetnostne porazdelitve so Gaussova (normalna), enakomerna, Maxwelllova, eksponentna in Lorentzova (Cauchyeva ali Breit—Wignerjeva) porazdelitev. Primere nekaterih bomo obravnavali potem, ko bomo spoznali pojem pričakovanih vrednosti. Na tem mestu se posebej posvetimo le eksponentni porazdelitvi, ker bi jo radi fizikalno utemeljili. V knjigi so v tretjem poglavju opisane tudi porazdelitev χ^2 (chi-kvadrat, f_{χ^2}), Studentova porazdelitev (f_T) in porazdelitev F, ki so zelo uporabne, a na tem mestu še ni jasno čemu sluzijo. Mi jih bomo predstavili šele ko jih bomo potrebovali.

3.1 Eksponentna verjetnostna porazdelitev

Z eksponentno verjetnostno porazdelitvijo opišemo procese, pri katerih je verjetnost za določen dogodek na časovno enoto konstantna. Zelo znan primer je časovni potek razpada radioaktivnih jeder, druge primere pa bomo navedli kasneje v tem podpoglavju. Razpad nestabilnega atomskega jedra je naključni proces *par excellence*. Za eno samo jedro ne moremo napovedati trenutka njegovega razpada; verjetnost za razpad v nekem časovnem intervalu je odvisna samo od dolžine tega intervala, Δt , ne pa od starosti jedra. Pravimo, da se jedra ne “starajo” in da je radioaktivni razpad proces “brez spomina”.

Neobvezno: Preden se lotimo matematičnega opisa, vsaj grobo utemeljimo zakaj je tako. Preprost model jedrskega razpada alfa je gibanje alfa delca v potencialu preostalega jedra. Ta ima privlačni del zaradi močne jedrske sile in odbojni del zaradi Coulomskega odboja med pozitivnimi protoni. Slednja povzroči bariero v potencialu, ki v klasični fiziki preprečuje delcu alfa, da zapusti jedro. V kvantni fiziki pa lahko alfa delec zapusti jedro le zaradi določene verjetnosti za tuneliranje skozi bariero. Ta verjetnost je odvisna le od višine bariere in kinetične energije alfa delca, in neodvisna od starosti jedra saj se razmere v jedru s časom ne spreminjajo. Zato je verjetnost za tuneliranje in tako tudi za razpad jedra neodvisna od starosti jedra.



Oglejmo si analogijo eksperimenta, kjer je verjetnost za razpad na časovno enoto Δt konstantna. Ta verjetnost naj bo kar $\frac{1}{2}$ za kar bo poskrbel met kovanca. Razpad 50 jeder modelirajmo naključno tako, da vržemo 50 kovancev in “razpadla” so tista za katere je padla cifra. S preostalimi jedri zopet ponovimo poskus, ki ga ponavljamo v enakih časovnih razmikih Δt . Opazujemo število jeder oz kovancev v odvisnosti od časa.

<https://www.youtube.com/watch?v=sc2caB0gc7Q>

Zdaj vse skupaj opredelimo se matematično:

T= naključna spremenljivka, ki predstavlja čas oziroma trenutek ob katerem jedro razpade

Pravimo, da se jedra ne “starajo” in da je radioaktivni razpad proces “brez spomina”: denimo, da na razpad jedra zaman čakamo že čas t ; verjetnost, da se razpad končno zgodi po $t + \Delta t$, je neodvisna od t ,

$$P(T > t + \Delta t | T > t) = P(T > \Delta t) . \quad (3.2)$$

Če je interval Δt dovolj kratek, lahko privzamemo, da je verjetnost za razpad sorazmerna z Δt in tedaj jedru ostane le izbira

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

kjer je $\lambda = 1/\tau$ *verjetnost za razpad na časovno enoto* [s^{-1}] ali tako imenovana *razpadna konstanta*, τ pa je *karakteristični ali razpadni čas*. Verjetnost, da jedro po času $n\Delta t$ še *ni* razpadlo, je $(1 - \lambda\Delta t)^n$. Verjetnost, da ni razpadlo po dolgem času $t = n\Delta t$, se pravi, da bo razpadlo ob nekem času $T > t = n\Delta t$, je torej

$$\text{veretnost, da jedro do časa } t \text{ še ni razpadlo} \\ P(T > t) = \lim_{n \rightarrow \infty} (1 - \lambda\Delta t)^n = \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda t}{n}\right)^n = e^{-\lambda t} . \quad (3.3)$$

Ker velja $P(T > t) = 1 - P(T \leq t) = 1 - F_T(t)$, lahko takoj izračunamo tudi ustrezno verjetnostno gostoto,

$$f_T(t) = \frac{dF_T(t)}{dt} = \frac{d}{dt} (1 - e^{-\lambda t}) = \lambda e^{-\lambda t} , \quad t \geq 0 , \quad (3.4)$$

verjetnost dP, da jedro razpade ob trenutku T na intervalu [t,t+dt] je enaka dP=f_T(t) t= dt

Število jeder ob času $t=0$ je N_0 , in opazujemo njihovo število v odvisnosti od časa. Jedra so neodvisna, zato ustrezna števila preživetja in razpadov dobimo iz zgornjih verjetnosti, tako da jih zgolj pomnožimo z N_0 .

Za vsako je verjetost, da do časa t še ni razpadlo podano z enačbo (3.3). Torej je *število jeder, ki do časa t še ni razpadlo*

$$N(t) = N_0 P(T > t) = N_0 e^{-\lambda t} . \quad (3.5)$$

Število jeder dN , ki v časovnem oknu razpade je podano z verjetnostjo $dP = f_T(t) dt$ (3.4)

$$dN(t) = - dP N_0 = - f_T(t) dt N_0 = - \lambda e^{-\lambda t} dt = - \lambda N(t) dt . \text{ oz } dN(t)/N(t) = - \lambda dt$$

kar je seveda enako kot če (3.5) odvajamo. Minus nastopa saj se število jeder dN zmanjša. Zvezo $dN(t) = - \lambda N(t) dt$ bi lahko seveda zapisali že takoj ob predpostavki, da je število razpadov sorazmerno le s trenutnim številom delcev in dolžine časovnega okna dt .

Drugi primeri eksponentne verjetnostne porazdelitve:

T je tu trenutek ko se zgodi dogodek (razpad, trk, okvara,...)

Verjetnost za trk molekule plina v drugo molekulo na časovno enoto se s časom ne spreminja. Verjetnost, da je torej molekula potovala brez trka do časa t je torej porazdeljena $P(T > t) = e^{-\lambda t}$

Verjetnost za kemijsko reakcijo $A+B \rightarrow C+D$ na časovno enoto je pri nespremenljivih okoliščinah je neodvisna od tega koliko časa reaktanti A in B že niso reagirali. Torej je za dane reaktante verjetnost da niso reagirali do časa t eksponentna $P(T > t) = e^{-\lambda t}$

Če pride do okvare neke naprave zaradi naključnih (na primer kvantno-mehanskih) procesov in ne zaradi staranja, bo verjetnost za dobro delovanje naprave do časa t eksponentno $P(T > t) = e^{-\lambda t}$ Primer je življanski čas računalniškega diska ali žarnice LED.

Če je pečenje avtomobila po cesti mimo določene točke naključno, in če se verjetnost prečenoa na časovno enoto ne spreminja s časom, bo verjetnost za prehod do časa t zopet enaka $P(T > t) = e^{-\lambda t}$

Minizgled Opazujemo dva enaka izotopa s karakterističnim razpadnim časom $\tau = \frac{1}{\lambda} = 84$ s. Eno jedro je pravkar nastalo, drugo pa je nastalo pred 3 minutami. Za katerega bo večja verjetnost, da bo razpadlo v naslednjih $\Delta t = 30$ sekundah?

Ta verjetnost je za oba enaka. To vemo ze iz osnovne predpostavke, da verjetnost ni odvisna od starosti jedra. DN Preverite se računsko, da spodnja pogojna verjetnost res ni odvisna od starosti jedra t . Resitev je na str 58 knjige: V obeh primerih je verjetnost 0.035.

$$P(T > t + \Delta t | T > t) = P(T > \Delta t) .$$

Zgled, str 58 (strežbe na bančnih okencih -> pri predavanjih: okvare računalniških diskov)

Zgled Osebi A in B istočasno prideta k blagajniškima okencema. Čas njune postrežbe je naključna spremenljivka z eksponentno porazdelitvijo s časovnim parametrom λ_A za prvo in λ_B za drugo okence. Izračunajmo verjetnost, da B od okenca odide prej kot A!

Naključni spremenljivki, ki merita časa dejanske postrežbe, označimo s T_A oziroma T_B . Verjetnost, da stranka A do časa t_A še ni postrežena, je $e^{-\lambda_A t_A}$. Ustrezna verjetnost za stranko B je $e^{-\lambda_B t_B}$. Ker sta procesa čakanja med seboj neodvisna, je njuna skupna verjetnostna gostota produkt posameznih verjetnostnih gostot:

$$f_{T_A, T_B}(t_A, t_B) = \lambda_A e^{-\lambda_A t_A} \cdot \lambda_B e^{-\lambda_B t_B} .$$

Iskana verjetnost je tedaj

$$P(T_B < T_A) = \int_0^\infty dt_A \int_0^{t_A} f_{T_A, T_B}(t_A, t_B) dt_B = \int_0^\infty dt_A \lambda_A e^{-\lambda_A t_A} (1 - e^{-\lambda_B t_A}) = \frac{\lambda_B}{\lambda_A + \lambda_B} .$$

Tudi limiti sta smiselni: če je uslužbenec, ki streže B, nadvse počasen ($\lambda_B \rightarrow 0$), gre $P(T_B < T_A) \rightarrow 0$, v nasprotnem primeru pa $P(T_B < T_A) \rightarrow 1$. $\triangleleft$

Zgled: str 59 : primer eksponentne verjetnostne porazdelitve, ki ni funkcija t temveč x

Zmotno je prepričanje, da srečamo eksponentne porazdelitve le pri naključnih spremenljivkah, katerih vrednosti imajo kaj opraviti s časom. Zamislimo si posodo z velikim številom majhnih kroglic s premerom D , med katerimi je delež p črnih in $1 - p$ belih [2]. Kroglice vlečemo iz posode in jih zlagamo v vrsto eno zraven druge. Denimo, da smo ravnokar izvlekli črno kroglico. Kolikšna je verjetnost, da je razdalja x med njenim središčem in središčem naslednje črne kroglice natanko iD ($i = 1, 2, \dots$)? Opazujemo zaporedja izvlečenih kroglic oziroma "dogodke"

$$\bullet | \bullet, \quad \bullet | \circ \bullet, \quad \bullet | \circ \circ \bullet, \quad \bullet | \underbrace{\circ \circ \dots \circ \circ}_{(i-1)D} \bullet,$$

torej je iskana verjetnost očitno

$$P(x = iD) = (1 - p)^{i-1} p.$$

Ker so dogodki nezdružljivi, je ustrezna porazdelitvena funkcija kar vsota vseh verjetnosti za posamezna zaporedja:

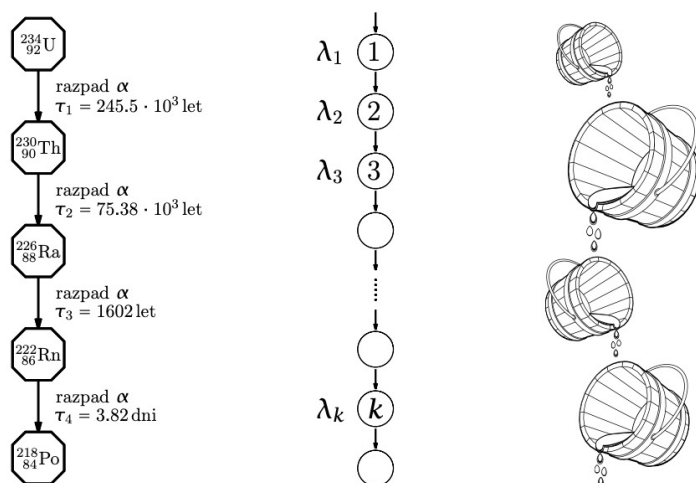
$$F_X(x) = P(x \leq iD) = p + (1 - p)p + \dots + (1 - p)^{i-1} p = 1 - (1 - p)^i.$$

Z oznakama $D = 1/n$ in $np = \lambda$ to prepišemo kot

$$F_X(x) = 1 - \left(1 - \frac{\lambda}{n}\right)^{nx},$$

saj je $i = x/D = nx$. Denimo, da naredimo limiti $n \rightarrow \infty$ in $p \rightarrow 0$ (to pomeni, da je v posodi zelo majhen delež črnih kroglic in da imajo zelo majhne premere), ob katerih pa ostaneta λ in x nespremenjena: tedaj gre $F_X(x) \rightarrow 1 - e^{-\lambda x}$, ustrezna gostota pa je $f_X(x) = dF_X/dx = \lambda e^{-\lambda x}$, kar je res enako kot v (3.4).

Zgled: str 77, Primer jedrskega razpadnega niza, kjer vsako jedro razpade le na en način



Slika 3.13 — [LEVO] Uranov razpadni niz v segmentu, v katerem je na vsaki stopnji možen samo en razpad. [SREDINA] Shema k zaporednih procesov z enim samim izhodom, ki jih opisuje hipoeksponentna porazdelitev. [DESNO] Nazorni prikaz jedrskega razpadnega niza; primerjaj ga s sliko 3.12 (desno).

✎ Predpostavimo, da se razpadni niz začne z jedri vrste 1, ki nimajo še nobenih potomcev, in da v to vrsto jeder ne razpada nobeno drugo jedro. Razpadni niz tedaj opiše sistem diferencialnih enačb

$$\begin{aligned} \dot{N}_1 &= -\lambda_1 N_1, \\ \dot{N}_2 &= -\lambda_2 N_2 + \lambda_1 N_1, \\ \dot{N}_3 &= -\lambda_3 N_3 + \lambda_2 N_2, \\ \dots &= \dots, \end{aligned}$$

4 Pričakovane vrednosti (za zvezne 1D verjetnostne porazdelitve)

V tem poglavju obravnavamo izračun količin, ki jih smemo pri posameznih naključnih spremenljivkah ali funkcijah teh spremenljivk — glede na verjetnostne porazdelitve, ki te spremenljivke opisujejo — pričakovati po velikem številu poskusov: imenujemo jih pričakovane vrednosti naključnih spremenljivk. Najpomembnejša takšna količina je **povprečna vrednost**, ki je pričakovana vrednost spremenljivke v osnovnem, najožjem pomenu besede, v nadaljevanju pa spoznamo še **varianco**=(standardni-odklon)², **pričakovane vrednosti višjih momentov** itd...

4.1 Pričakovana (povprečna) vrednost same spremenljivke

Pričakovano vrednost zvezne spremenljivke X izračunamo tako, da njeno vrednost x pomnožimo z verjetnostjo $f_X(x)dx$, da ima ta spremenljivka po velikem številu poskusov to vrednost x na intervalu $[x, x+dx]$. Potem tako utežene prispevke seštejemo oziroma integriramo po vseh vrednostih x

pričakovana (povprečna) vrednost X
$$\bar{X} = E[X] = \int_{-\infty}^{\infty} x f_X(x) dx .$$

Tako $E[X]$ kot $\bar{X}$ in μ_X pomenijo “operacijo povprečenja” nad spremenljivko X .

Komentar glede notacije : Za povprečje uporabljamo obe oznaki, poleg tega pa še μ_X . v poglavjih 4–6 vse tri pomenijo eno in isto reč, medtem ko bomo v poglavjih 7–10 z $\bar{X}$ in $\underline{x}$ označevali povprečje vzorca, za pričakovano vrednost populacije (pravo povprečno vrednost kot jo obravnavamo v tem poglavju) pa bomo uporabljali $E[\bullet]$.

4.4 Pričakovane vrednosti funkcij naključnih spremenljivk

Najpreprostejši funkciji naključnih spremenljivk sta vsota $X+Y$ dveh spremenljivk in linearna kombinacija $aX+b$, kjer sta a in b poljubni konstanti. Ker je pričakovana vrednost zvezne spremenljivke, $E[X]$, definirana z integralom, pričakovani vrednosti $E[X+Y]$ in $E[aX+b]$ podedujeta vse lastnosti integrala, med njimi linearnost: $E[X+Y] = E[X] + E[Y]$, $E[aX+b] = aE[X] + b$, $E[aX+bY] = aE[X] + bE[Y]$

Določimo še povprečno vrednost spremenljivke $Y=g(X)$ ob dani porazdelitvi $f_X(x)$. Vrednost naključne spremenljivke $g(x)$ pomnožimo z njeno verjetnostjo pojavljanja $f_X(x)dx$ na intervalu $[x, x+dx]$, in vse tako utežene vrednosti $g(x)$ seštejemo/integriramo

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

Sem sodi še komentar o zelo pogosto uporabljenem približku, ki je lahko grda napaka ali pa dobra bližnjica do rešitve: to je aproksimacija

$$\overline{g(X)} \neq g(\bar{X}), \quad \overline{g(X)} \simeq g(\bar{X})$$

Bolj ko je porazdelitev f_X spremenljivke X ozka, boljši je ta približek.

4.5 Razpršenost (varianca) ter efektivni odmik (standardni odklon, standardna deviacija)

Z izračunom pričakovane vrednosti naključne spremenljivke X nekaj izvemo o tem, kje znotraj njenega definicijskega območja bodo pri mnogih ponovitvah naključnega poskusa večinoma pristale njene vrednosti. Zdaj nas zanima še, kakšna je raztresenost (razpršenost) vrednosti okrog njihovega povprečja $E[X]=\bar{X}$. Mera za razpršenost je varianca (angl. variance), definirana kot

	splošna definicija	za zvezno porazdelitev
varianca (mera za razpršenost)	$\text{var}[X] = E[(X - E[X])^2] = \overline{(X - \bar{X})^2}$	$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \bar{X})^2 f_X(x) dx$

standardni odklon= standardna deviacija	$\sigma_X = \sqrt{\text{var}[X]}$
--	-----------------------------------

Velika varianca pomeni veliko razpršenost okrog povprečja in obratno. Pozitivni kvadratni koren variance, imenujemo standardna deviacija (angl. standard deviation) — zlasti kadar imamo v mislih normalno porazdelitev — včasih pa tudi efektivni odmik ali standardni odklon.

$$\sigma_X^2 = E[(X - \bar{X})^2] = E[X^2 - 2\bar{X}X + \bar{X}^2] = E[X^2] - 2\bar{X}E[X] + \bar{X}^2 = \overline{X^2} - 2\bar{X}^2 + \bar{X}^2 = \overline{X^2} - \bar{X}^2$$

4.7 Momenti naključne spremenljivke

Povprečje in varianca sta posebna primera pričakovanih vrednosti, imenovanih momenti. Uporabljata se dve definiciji p -tega momenta naključne spremenljivke

	splosna definicija	za zvezno porazdelitev	
centralni p -ti moment	$M_p = E[(X - \bar{X})^p]$	$M_p = \int_{-\infty}^{\infty} (x - \bar{X})^p f_X(x) dx$	M_2 =standardni odk.
algebrajski p -ti moment	$M'_p = E[X^p]$	$M'_p = \int_{-\infty}^{\infty} x^p f_X(x) dx$	M'_1 =povprečna vr.

Pričakovane vrednosti za primer naključne spremenljivke X , kjer ima ta diskretne vrednosti $x_1, \dots, x_n$ določimo analogno, le da integral zamenjamo z vsoto $\sum_{i=1}^n$, npr

standarni odklon za diskretno porazdelitev	$\sigma_X^2 = \sum_{i=1}^n (x_i - \bar{X})^2 f_X(x_i)$
--	--

4.10.1 Pričakovani čas okvare računalniškega diska

Računalniški disk krmili pet elementov ($i = 1, 2, 3, 4, 5$). Čas do nepopravljive napake v vsakem elementu je porazdeljen eksponentno, s posamezno časovno konstanto λ_i . Disk deluje, dokler sočasno delujejo elementi 1, 2 in 3, elementi 3, 4 in 5, ali pa, očitno, vseh pet elementov naenkrat. Kako dolgo po proizvodnji diska smemo pričakovati, da disk še deluje brez napake?

 Verjetnost, da i -ti element do časa t še ni pokvarjen (verjetnost, da je čas, ko se pokvari, večji od t), eksponentno pada in je enaka $e^{-\lambda_i t}$. Za okvaro diska so odgovorni trije ključni dogodki:

dogodek A : okvara elementov 1 in 2 po času t : $P(A) = e^{-(\lambda_1 + \lambda_2)t}$,
 dogodek B : okvara elementa 3 po času t : $P(B) = e^{-\lambda_3 t}$,
 dogodek C : okvara elementov 4 in 5 po času t : $P(C) = e^{-(\lambda_4 + \lambda_5)t}$.

Disk deluje, dokler $(A \cap B \cap \bar{C}) \cup (\bar{A} \cap B \cap C) \cup (A \cap B \cap C) \neq \{\}$. Verjetnost, da po času t disk še deluje, je torej

$$\begin{aligned} P(T > t) &= P(t) = P(A \cap B \cap \bar{C}) + P(\bar{A} \cap B \cap C) + P(A \cap B \cap C) \\ &= P(A)P(B)[1 - P(C)] + [1 - P(A)]P(B)P(C) + P(A)P(B)P(C) \\ &= P(B)[P(A) + P(C) - P(A)P(C)] \\ &= e^{-(\lambda_1 + \lambda_2 + \lambda_3)t} + e^{-(\lambda_3 + \lambda_4 + \lambda_5)t} - e^{-(\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5)t}. \end{aligned}$$

T=čas (trenutek) ob katerem pride do okvare diska

Dolociti moramo verjetnostno porazdelitev naključne spremenljivke T in $f_T(t)dt$ naj označuje verjetnost da se disk pokvari od trenutku T na intervalu $[t, t+dt]$.

$f_T(t)$ dobimo iz verjetnosti $P(T > t)$ zgoraj kot $f_T(t)dt = P(t) - P(t+dt) = -P'(t)dt$ torej je $f_T(t) = -P'(t)$ kar smo izpeljali tudi že ob začetku obravnave eksponentne porazdelitve.

$$\begin{aligned} f_T(t) = -P'(t) &= \left[(\lambda_1 + \lambda_2 + \lambda_3)e^{-(\lambda_1 + \lambda_2 + \lambda_3)t} + (\lambda_3 + \lambda_4 + \lambda_5)e^{-(\lambda_3 + \lambda_4 + \lambda_5)t} \right. \\ &\quad \left. - (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5)e^{-(\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5)t} \right] \end{aligned}$$

Nal: Določi povprečen T in standardni odklon T , ter primerjaj povprečje(T^3) ter $(\text{povprečje}(T))^3$ za primer

$$\lambda_n = n \cdot \lambda$$

$$\begin{aligned} \int_0^\infty f_T(t) dt &= 1 + 1 - 1 = 1 & \bar{T} &= \int_0^\infty t \left[(\lambda_1 + \lambda_2 + \lambda_3)e^{-(\lambda_1 + \lambda_2 + \lambda_3)t} + (\lambda_3 + \lambda_4 + \lambda_5)e^{-(\lambda_3 + \lambda_4 + \lambda_5)t} \right. \\ &&& \left. - (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5)e^{-(\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5)t} \right] dt \\ &&&= \frac{1}{\lambda_1 + \lambda_2 + \lambda_3} + \frac{1}{\lambda_3 + \lambda_4 + \lambda_5} - \frac{1}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5}. \end{aligned}$$

```
f[t_] := 6 * λ * Exp[-6 * λ * t] + 12 * λ * Exp[-12 * λ * t] - 15 * λ * Exp[-15 * λ * t];
```

```
Assuming[λ > 0,
```

```
tav = Integrate[f[t] * t, {t, 0, Infinity}];
```

```
var = Integrate[f[t] * (t - tav) ^ 2, {t, 0, Infinity}];
```

```
M3p = Integrate[f[t] * (t) ^ 3, {t, 0, Infinity}];
```

```
];
```

Slednji sta zelo različni; približnega ujemanja ne pričakujemo ker je porazdelitev široka

$$\text{av}(T) = \frac{0.18}{\lambda}, \quad \text{sig}(T) = 0.16 \sqrt{\frac{1}{\lambda^2}}, \quad \text{av}(T^3) = \frac{0.029}{\lambda^3}, \quad ((\text{av}(T))^3 = \frac{0.0062}{\lambda^3})$$

zvezne verjetnostne porazdelitve v Wolframovi Mathematici
datoteke: [posebne-zvezne-porazdelitve.nb](#)

PDF: ProbabilityDensityFunction : $f_X(x)$

CDF: CumulativeDensityFunction, $F_X(x)$

InverseCDF : F_X^{-1}

NormalDistribution[μ , σ]

UniformDistribution[{ a , b }]

ExponentialDistribution[λ]

MaxwellDistribution[σ]

(* kasneje bomo vpeljali tudi spodnje *)

CauchyDistribution[a , b]

ChiSquareDistribution[ν]

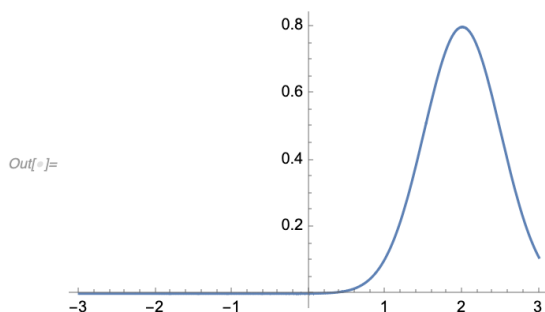
StudentTDistribution[ν]

za vse porazdelitve se uporablja enake
funkcije PDF, CDF, InverseCDF, ..

In[]:= PDF[NormalDistribution[μ , σ], x]

$$\text{Out[]} = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}$$

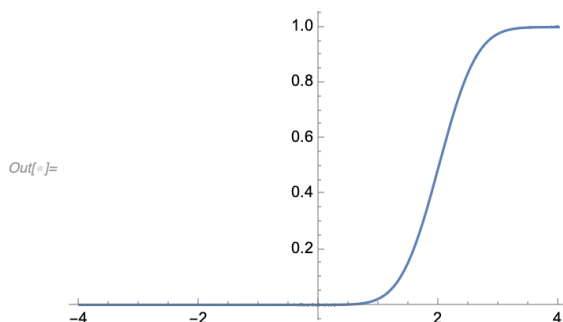
Plot[PDF[NormalDistribution[2, 0.5], x], { x , -4, 4}]



In[]:= CDF[NormalDistribution[μ , σ], x]

$$\text{Out[]} = \frac{1}{2} \operatorname{Erfc}\left[\frac{-x + \mu}{\sqrt{2}\sigma}\right]$$

In[]:= Plot[CDF[NormalDistribution[2, 0.5], x], { x , -4, 4}]



In[]:= InverseCDF[NormalDistribution[2, 0.5], 0.9]

Out[]:= 2.64078

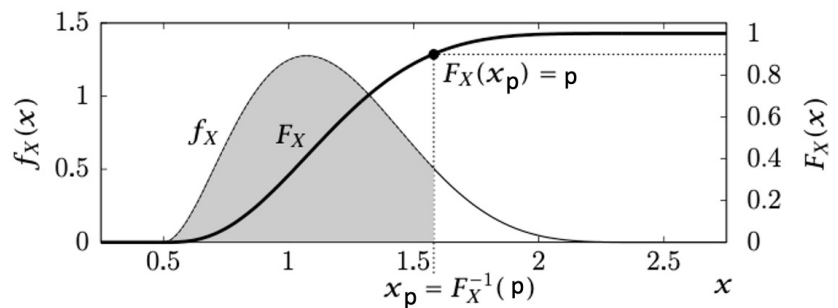
In[]:= Assuming[{ $\sigma > 0$, $\mu > 0$ }, Integrate[PDF[NormalDistribution[μ , σ], x] * x , { x , -Infinity, Infinity}]]

Out[]:= μ

In[]:= Assuming[{ $\sigma > 0$, $\mu > 0$ }, Integrate[PDF[NormalDistribution[μ , σ], x] * ($x - \mu$) ^ 2, { x , -Infinity, Infinity}]]

Out[]:= σ^2

4.3 Kvantili, percentili



Slika 4.2 — Definicija kvantila zvezne porazdelitve. Integral verjetnostne gostote $f_X(x)$ od $-\infty$ (oziroma od skrajnega levega konca definicijskega območja) do $x = x_p$ mora biti enak p . Slika prikazuje verjetnostno gostoto $f_X(x) = \frac{21}{32}(x - \frac{1}{2})^2(\frac{5}{2} - x)^5$, $0.5 \leq x \leq 2.5$, z ustrezno porazdelitveno funkcijo in 90. percentilom ($p = 0.90$), ki je enak $x_p = 1.58$.

Vrednost naključne spremenljivke (x_p), pod katero po mnogih poskusih najdemo določen delež vseh dogodkov (p), imenujemo kvantil njene porazdelitve (lat. quantum, “koliko”). Za zvezno porazdelitev to pomeni, da je integral verjetnostne gostote od $-\infty$ do x_p enak p (slika 4.2).

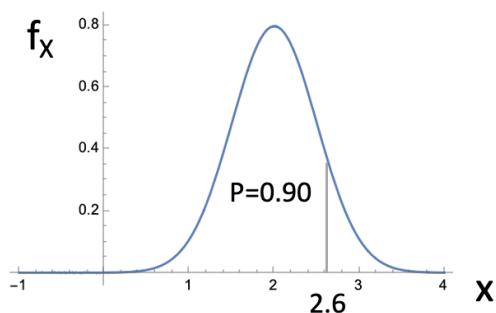
Primer: 0.90-ti kvantil normalne porazdelitve $N(2, 0.5)$ je enak $x_{0.90} = 2.6$.

Za izražanje p -tega kvantila so smiselne vse vrednosti $0 \leq p \leq 1$, vendar so v rabi še njegove logične izpeljanke za značilne vrednosti p : celoštevilskim vrednostim (izraženim v odstotkih) ustrezajo **percentili**.

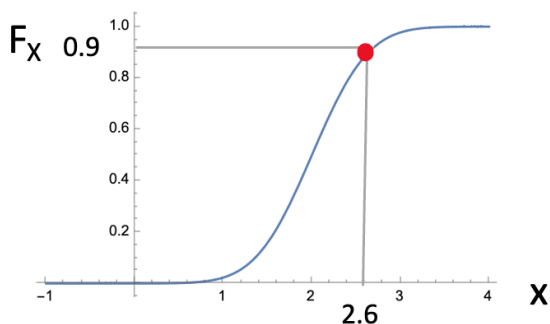
90-ti percentil normalne porazdelitve $N(2, 0.5)$ je enak $x_{0.90} = 2.6$.

0.70-ti kvantil eksponentne porazdelitve z $\lambda=1$ je enak $x_{0.70} = 1.2$.

`Plot[PDF[NormalDistribution[2, 0.5], x], {x, -1, 4}]`



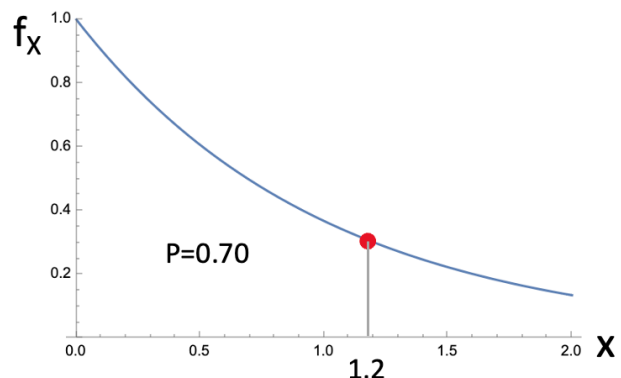
`Plot[CDF[NormalDistribution[2, 0.5], x], {x, -1, 4}]`



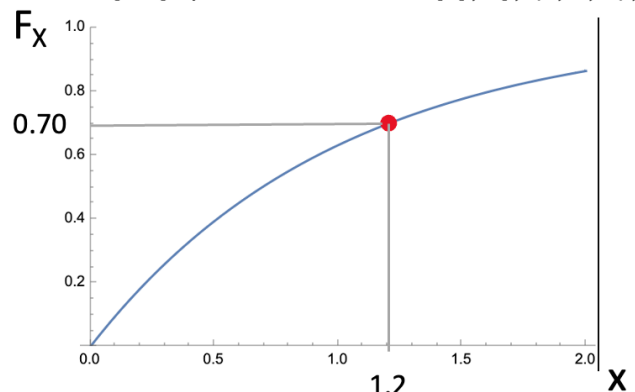
`InverseCDF[NormalDistribution[2, 0.5], 0.9]`

2.64078

`Plot[PDF[ExponentialDistribution[1], x], {x, 0, 2}]`



`Plot[CDF[ExponentialDistribution[1], x], {x, 0, 2}]`



`InverseCDF[ExponentialDistribution[1], 0.7]`

1.20397

4.2 Mediana in modus

Mediana naključne spremenljivke X (diskretne ali zvezne) je tista njena vrednost $x = \text{med}[X]$, za katero velja

$$P(X < x) \leq \frac{1}{2} \quad \text{in} \quad P(X > x) \leq \frac{1}{2}. \quad (4.4)$$

V primeru zvezne spremenljivke X neenakosti v tej enačbi postaneta enakosti,

$$P(X < x) = P(X > x) = \frac{1}{2} \quad \Leftrightarrow \quad \text{med}[X] = F_X^{-1}(1/2),$$

saj je vselej mogoče najti x , ki ploščino pod krivuljo verjetnostne gostote razdeli natanko na dve polovici: verjetnosti, da zavzame X vrednost pod oziroma nad mediano, sta natanko 50 %.

Modus spremenljivke X je tista vrednost z največjo verjetnostjo (v intervalu z danim dx): $\text{mod}[X] = X_{\max}$, $df/dx|_{\{x=X_{\max}\}} = 0$.

Motivacija za tovrstne pričakovane vrednosti izhaja predvsem pri obravnavi diskretnih naključnih spremenljivk pri statistiki, kjer se nekatere vrednosti x_i v vzorcu močno razlikujejo od glavnine – pravimo jim **ubežniki (eng outliers)**. Povprečje je lahko zelo občutljivo na take ubežnike, mediana in modus pa sta robustna na ubežnike.

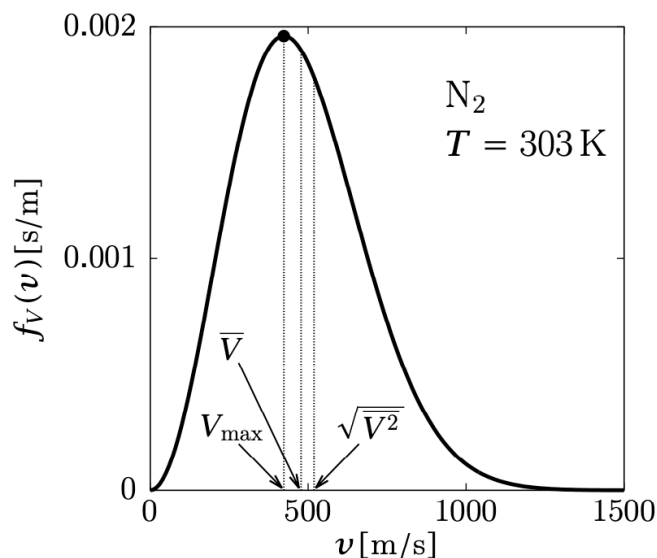
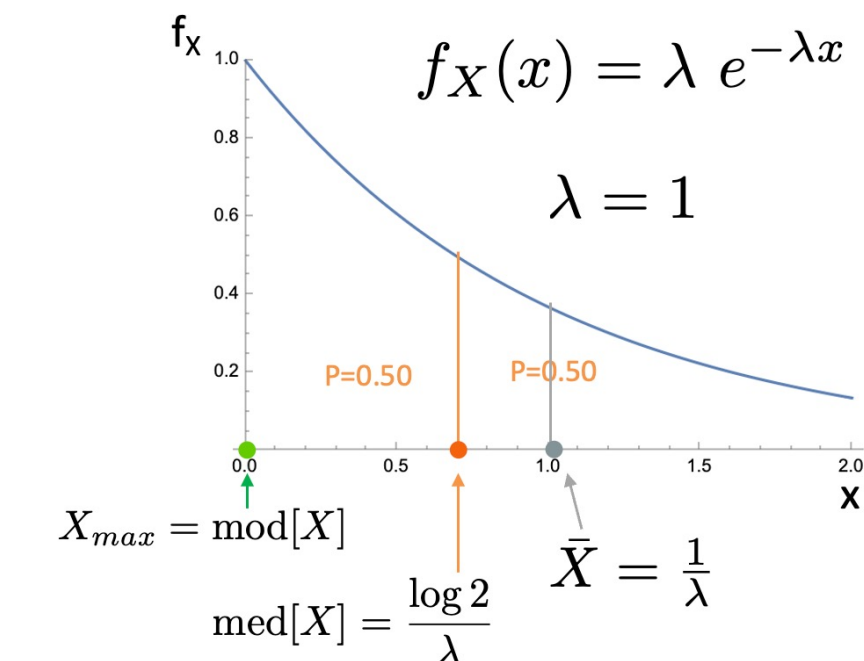
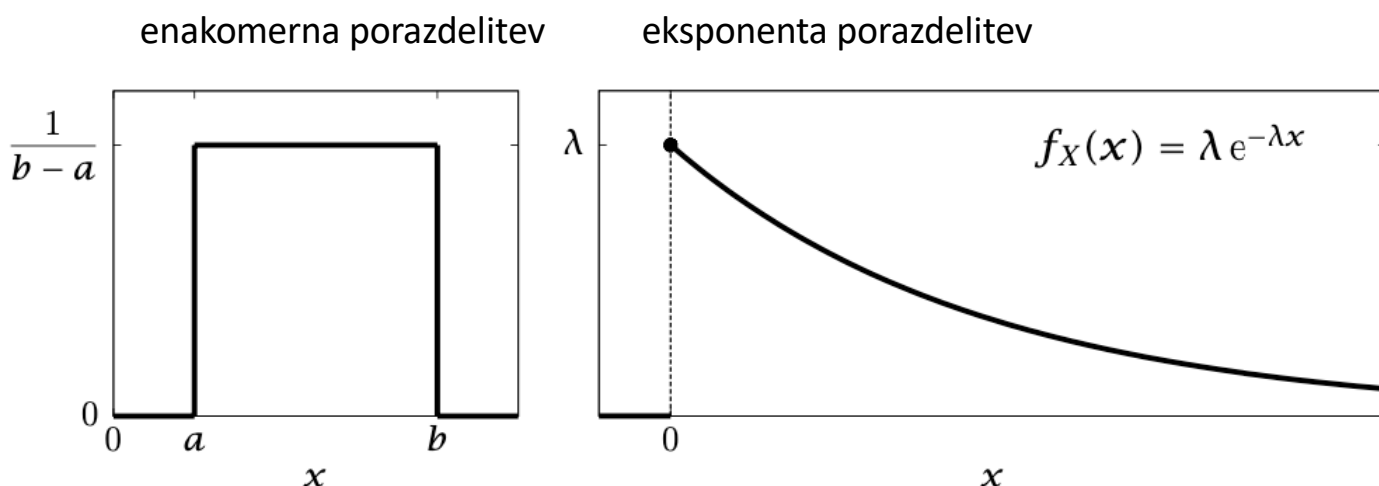


Tabela pricakovanih vrednosti za izbrane zvezne porazdelitve

Porazdelitev	povprečje	mediana	modus	varianca	
$U(a, b)$ (3.1)	$\frac{a+b}{2}$	$\frac{a+b}{2}$	/	$\frac{(b-a)^2}{12}$	
$\text{Exp}(\lambda)$ (3.4)	$\frac{1}{\lambda}$	$\frac{\log 2}{\lambda}$	0	$\frac{1}{\lambda^2}$	
$N(\mu, \sigma^2)$ (3.7)	μ	μ	μ	σ^2	
Cauchy (3.20)	/	x_0	x_0	/	vaje , kasneje
$\chi^2(\nu)$ (3.21)	ν	$\nu - \frac{2}{3}^\dagger$	$\nu - 2^\ddagger$	2ν	kasneje
$t(\nu)$ (3.22)	0^*	0	0	$\frac{\nu}{\nu-2}^*$	kasneje



Slika 3.1 — [LEVO] Verjetnostna gostota enakomerne porazdelitve $U(a, b)$. [DESNO] Verjetnostna gostota eksponentne porazdelitve s parametrom λ .

3.3 Normalna (Gaussova) porazdelitev

Po normalni (Gaussovi) porazdelitvi ali vsaj približno v skladu z njo so porazdeljene neštete količine iz sleherne sfere človeškega obstoja in narave. Normalno ali približno normalno so na primer porazdeljeni premeri navojev vijakov, ki v tisočih letijo iz stružnice, telesne mase ljudi, izpitne ocene in hitrosti plinskih molekul. Delne razlage in utemeljitve za to vseprisotnost Gaussove porazdelitve bomo deležni v razdelku 6.3 in zlasti poglavju 11, kjer bomo obravnavali **centralni limitni teorem**: ta v grobem pravi, da je povprečje velikega števila identičnih in neodvisnih naključnih spremenljivk ($X^{(i)}$) porazdeljeno po Gaussovi porazdelitvi, četudi posamezne spremenljivke $X^{(i)}$ niso porazdeljene Gaussovo.

Gaussova porazdelitev je dvoparametrične z verjetnostno gostoto

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty,$$

$N(\mu, \sigma^2)$

Že zgolj iz definicijskega območja je razvidno, zakaj je normalna porazdelitev mnogokrat le približek: telesne mase ne morejo biti negativne in izpitne ocene ne morejo biti neskončne. Porazdelitev je simetrična okrog vrednosti μ , širino njenega vrha pa določa standardna deviacija σ ; v točkah $x = \mu \pm \sigma$ ima funkcija f_X prevoj. Splošno sprejeta "kratica" za normalno porazdelitev je $N(\mu, \sigma^2)$.

μ je njeno povprečje, σ pa njeno standardna deviacija

Standardizirana normalna porazdelitev

Pri delu z gaussovske porazdeljenimi podatki se je smiselno znebiti odvisnosti od položaja izhodišča in širine krivulje, tako da od spremenljivke X odštejemo μ in jo delimo s σ ter tako tvorimo novo, *standardizirano* naključno spremenljivko

$$Z = \frac{X - \mu}{\sigma}.$$

Porazdelitev spremenljivke Z tedaj imenujemo *standardizirana normalna* in jo označimo z $N(0, 1)$ (povprečje nič in enotska varianca). Ustreza ji namreč verjetnostna gostota

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad (3.10)$$

$N(0, 1)$

porazdelitvena funkcija pa je

$$F(z) = \Phi(z) = P(Z \leq z) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^z e^{-t^2/2} dt = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right) \right]. \quad (3.11)$$

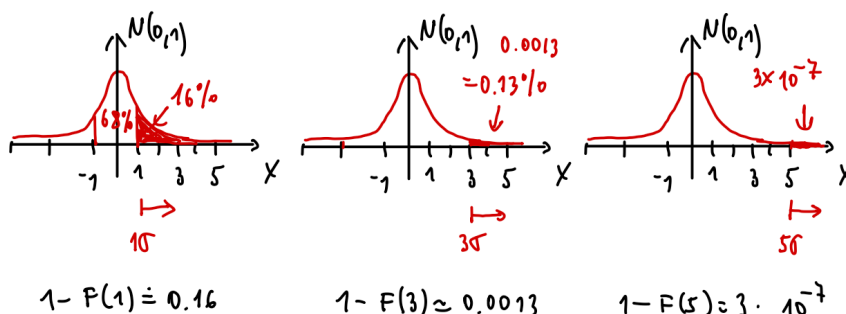
Vrednosti določenih integralov standardizirane normalne porazdelitve

$$\frac{1}{\sqrt{2\pi}} \int_0^z e^{-t^2/2} dt = \frac{1}{2} \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right) \quad (3.12)$$

za z med 0 in 5 v korakih po 0.01, ki zadoščajo za vsakdanjo rabo, so zbrane v tabeli [D.1](#). Abscise $x = \mu \pm n\sigma$ oziroma $z = \pm n$ ($n = 1, 2, \dots$) so še posebej pomembne. Ploščine pod krivuljo $f_Z(z)$ na teh intervalih,

$$P_{n\sigma} = P(\mu - n\sigma \leq X \leq \mu + n\sigma) = P(-n \leq Z \leq n) = \operatorname{erf}\left(\frac{n}{\sqrt{2}}\right),$$

funkcija napake (error function) je definirana kot: $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$



V fiziki osnovnih delcev trdimo, da smo **odkrili nov delec**, če ga odkrijemo s **statistično pomembnostjo 5σ** ali več. **Statistična pomembnost (eng. *significance*) 5σ pri odkritju novega delca pomeni**, da je verjetnost, da je signal posledica statistične fluktuacije (in ne obstoja novega delca) enaka $P=3 \times 10^{-7}$. Pri tem imamo v mislih verjetnost za signal, ki je enak eksperimentalno opaženemu ali pa še večji. To seveda nakazuje izredno majhno verjetnost, da je opažen signal zgolj statistična fluktuacija; imamo torej utemeljene razloge, da signal pripisemo obstoju novega delca. Komentar: Pri verjetnosti $P=3 \times 10^{-7}$ upoštevamo le možnosti, ki izhajajo vrednosti naključne spremenljivke nad povprečjem (torej povečan signal nad ozadjem).

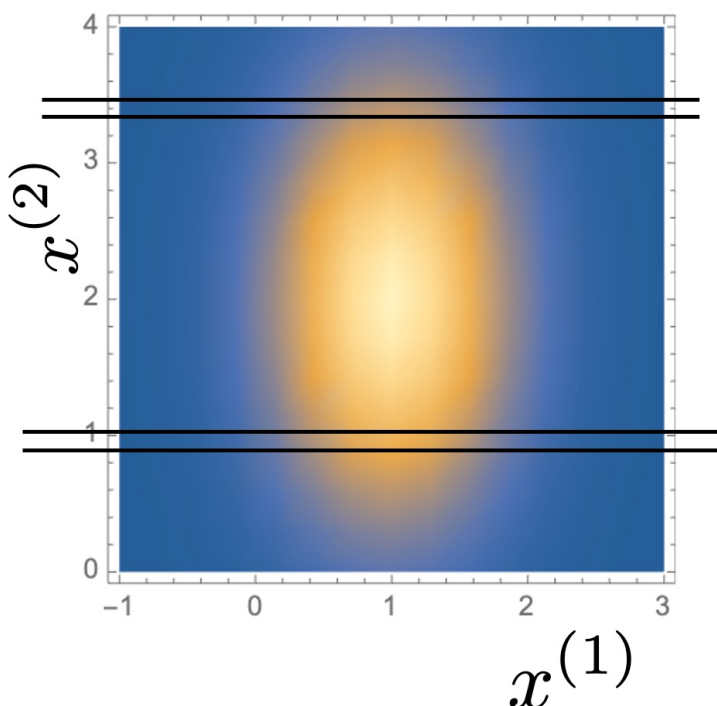
$$P(X > \mu + 5\sigma) = 3 \cdot 10^{-7}$$

Opomba: pri Gassovski porazdelitvi je to ravno verjetnost da opazljivka X preseže povprečje za 5σ ali več.

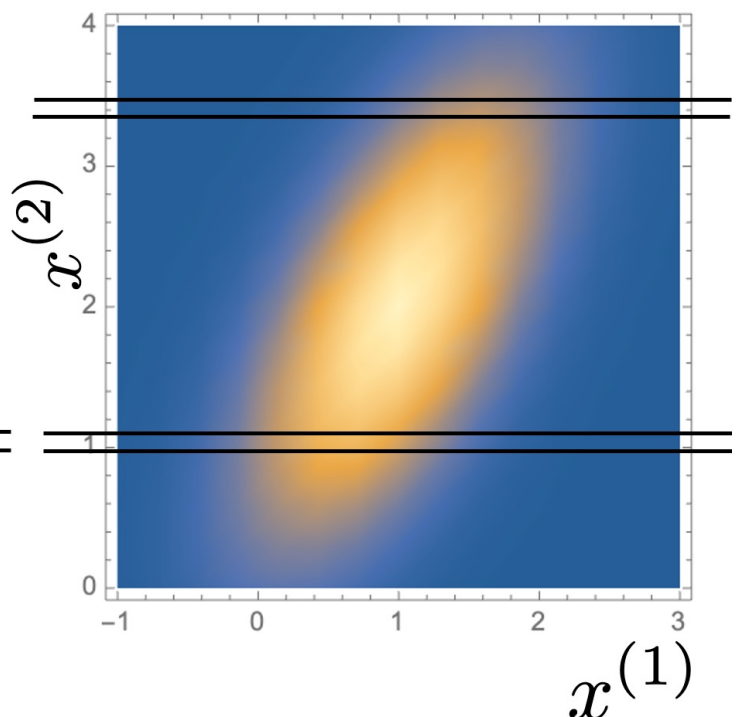
Tudi pri količinah, ki niso porazdeljene po Gaussovi porazdelitvi, se o odkritju z pomembnostjo 5σ , ko je verjetnost, da je opažen signal (ali še večji) posledica statistične fluktuacije enaka $P=3 \times 10^{-7}$. Statistična pomembnost 5σ je v tem primeru mera za pomembnost oz zanesljivost odkritja, ki se jo lahko uporablja in v grobem primerja pri vseh porazdelitvah.

Več in bolj podrobno o tem bomo govorili v poglavju o statističnih testih in hipotezah.

neodvisni x_1 in x_2



x_1 in x_2 nista neodvisni



Pričakovane vrednosti pri D-razsežnih verj. porazd.

D meključnih spremenljivk: $X^{(1)}, \dots, X^{(k)}, \dots, X^{(D)}$ $k=1, \dots, D$
 z vrednostmi $x^{(1)}, \dots, x^{(k)}, \dots, x^{(D)} \in \mathbb{R}$

Če so spremenljivke diskretne $x_i^{(1)}, x_i^{(k)}, \dots, x_i^{(D)}$ $k=1, \dots, D$
 $i=1, \dots, M$
 $x_1^{(k)}, \dots, x_m^{(k)}$ so meključne diskretne vrednosti spr. $x^{(k)}$
 različne spremenlj.

Verjetnostne gostote

$$d^D P = f_{\vec{X}}(x^{(1)}, \dots, x^{(D)}) dx^{(1)} \dots dx^{(D)} = f_{\vec{X}}(\vec{x}) d\vec{x}$$

Če so spremenljivke neodvisne:

$$f_{\vec{X}}(x^{(1)}, \dots, x^{(D)}) = f_{X^{(1)}}(x^{(1)}) \cdot \dots \cdot f_{X^{(D)}}(x^{(D)})$$

a to v splošnem ni res.

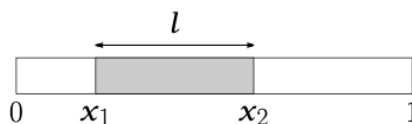
$$\bar{X}^{(k)} = E[X^{(k)}] = \int x^{(k)} f_{\vec{X}}(\vec{x}) d\vec{x} \xrightarrow{\text{za neodv.}} \int x^{(k)} f_{X^{(k)}}(x^{(k)}) dx^{(k)} \underbrace{\int f_{X^{(1)}}(x^{(1)}) dx^{(1)} \dots}_{1}$$

$$E[g(\vec{X})] = \int g(\vec{x}) f_{\vec{X}}(\vec{x}) d\vec{x}$$

Zgled (str 95) za pričakovano vrednost funkcije g pri dvorazsežni porazdelitvi; glede na notacijo iz prejsnje strani bi bili indeksi 1 in 2 za obe dolžini zgoraj

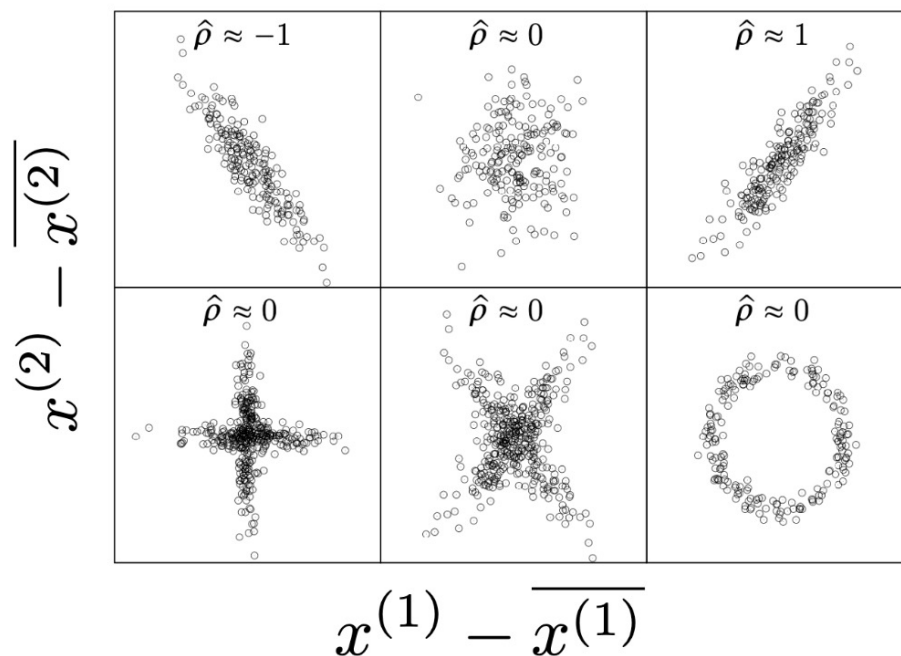
Zgled Veliko število palic dolžine 1 z udarcem naključno zlomimo na dveh mestih. Zanima nas povprečna dolžina srednjih koščkov. Ob vsakem

udarcu se palica zlomi na mestih $0 < x_1 < 1$ in $0 < x_2 < 1$, pri čemer so vrednosti x_1 in x_2 enakomerno porazdeljene po intervalu $[0, 1]$, poleg tega pa je lahko $x_1 < x_2$ ali $x_1 > x_2$. Iščemo torej pričakovano vrednost naključne spremenljivke $L = |x_2 - x_1|$ (z vrednostmi l) glede na verjetnostno gostoto $f_{X,Y}(x_1, x_2) = 1$:



$$\bar{L} = \int_0^1 \int_0^1 |x_2 - x_1| dx_1 dx_2 = \int_0^1 dx_2 \int_0^{x_2} (x_2 - x_1) dx_1 + \int_0^1 dx_2 \int_{x_2}^1 (x_1 - x_2) dx_1 = \frac{1}{3}.$$

Premisli: kako se rezultat spremeni, če verjetnost, da se bo ob udarcu palica zlomila, linearno narašča od vrednosti nič v izhodišču do konca palice? $\triangleleft$



Slika 7.8: Ponazoritev korelacije na primeru dveh spremenljivk $X^{(1)}$ in $X^{(2)}$ ter končnih vzorcev. Slike ponazarjajo množice izmerkov $(x^{(1)}, x^{(2)})$, ki jih je mogoče opisati z dvorazsežno porazdelitvijo spremenljivk $X^{(1)}$ in $X^{(2)}$, in ustrezne ocene za vzorčni korelacijski koeficient $\rho = \rho^{12}$. [Zgoraj, od desne proti levi] Skoraj popolnoma korelirani podatki ($\rho \approx 1$), nekorelirani podatki ($\rho \approx 0$) in skoraj povsem antikorelirani podatki ($\rho \approx -1$). [Spodaj] Trije primeri realizacije nekoreliranih spremenljivk, ki med seboj nista statistično neodvisni. Pogoju $\rho = 0$ je zgolj potreben pogoj za statistično neodvisnost.

$$\sigma_{X^{(k)}}^2 = E[(X^{(k)} - \bar{X}^{(k)})^2] = \int (x^{(k)} - \bar{x}^{(k)})^2 f_{\bar{X}}(\bar{x}) d\bar{x} \equiv \sigma_k$$

Pri dveh ali več ressejnih porazdelitvah je relevantna tudi pričakovana vrednost kovariacije dveh spremenljivk $X^{(k)}$ in $X^{(e)}$

kovariacija spremenljivk $X^{(k)}$ in $X^{(e)}$

korelacijski koeficient

$$\sum^{ke} = \text{cov}[X^{(k)}, X^{(e)}] = \int (x^{(k)} - \bar{x}^{(k)}) (x^{(e)} - \bar{x}^{(e)}) f_{\bar{X}}(\bar{x}) d\bar{x} \equiv \rho^{(ke)} \sigma_{X^{(k)}} \sigma_{X^{(e)}}$$

to je definicija

$$\sum^{ke} = \text{kovariacija metrike} \equiv E[(X^{(k)} - \bar{X}^{(k)})(X^{(e)} - \bar{X}^{(e)})]$$

$\sum^{ke}$ je matrični element

$$\sum^{ke} \equiv \text{cov}[X^{(k)}, X^{(e)}] = \sum_{i_1, \dots, i_D}^m (x_{i_1}^{(k)} - \bar{x}^{(k)}) (x_{i_1}^{(e)} - \bar{x}^{(e)}) f_{\bar{X}}(x_{i_1}^{(1)}, \dots, x_{i_D}^{(D)}) = \rho^{(ke)} \sigma_{X^{(k)}} \sigma_{X^{(e)}}$$

zo diskretno

indeksi i , pri $x_{i_1}^{(1)}$ tčajo $i_1 = 1, \dots, m$ in podobno ostali

Zgled: posoda s plinom.

Naključne opazljivke: tlak, število trkov na s, povprečna hitrost molekule,...

Te naključne opazljivke so korelirane, če jim merimo na steni iste posode s plinom.

Niso korelirane, če jih merimo na različnih posodah s plinom.

$$\text{cov}[X^{(k)}, X^{(e)}] = E[(X^{(k)} - \bar{X}^{(k)})(X^{(e)} - \bar{X}^{(e)})] = E[X^{(k)}X^{(e)}] - \bar{X}^{(k)}\bar{X}^{(e)} \quad (\#)_1$$

to velja splošno

$$\begin{aligned} \text{var}[X^{(k)} \pm X^{(e)}] &= E[(X^{(k)} \pm X^{(e)} - \bar{X}^{(k)} \pm \bar{X}^{(e)})^2] = \\ &= E[(X^{(k)} - \bar{X}^{(k)})^2] + E[(X^{(e)} - \bar{X}^{(e)})^2] \pm 2E[(X^{(k)} - \bar{X}^{(k)})(X^{(e)} - \bar{X}^{(e)})] \\ &= \sigma_{X^{(k)}}^2 + \sigma_{X^{(e)}}^2 \pm 2\text{cov}[X^{(k)}, X^{(e)}] \end{aligned}$$

$$\text{var}[X^{(k)} \pm X^{(e)}] = \text{var}[X^{(k)}] + \text{var}[X^{(e)}] \pm 2\text{cov}[X^{(k)}, X^{(e)}] \quad (\#)_2$$

to velja splošno

Ogledimo si kako se zgoraj zveze poenostavijo za neodvisni $X^{(1)}$ in $X^{(2)}$:

$$E[X^{(1)}X^{(2)}] = \int \int \underbrace{X^{(1)}}_{\downarrow \bar{X}^{(1)}} \underbrace{X^{(2)}}_{\downarrow \bar{X}^{(2)}} \underbrace{f_{X^{(1)}}(x^{(1)})}_{\downarrow \bar{X}^{(1)}} dx^{(1)} \underbrace{f_{X^{(2)}}(x^{(2)})}_{\downarrow \bar{X}^{(2)}} dx^{(2)} = \bar{X}^{(1)}\bar{X}^{(2)}$$

toréj je glede na $\oplus_1$ zgoraj $\text{cov}[X^{(1)}, X^{(2)}] = 0$

in glede na $\oplus_2$ zgoraj $\text{var}[X^{(1)} + X^{(2)}] = \text{var}[X^{(1)}] + \text{var}[X^{(2)}]$

Variance neodvisnih nekejskih spremenljivk se seštevajo.

4.8.2 Korelacija (correlation) ne implicira kavzalnosti (causality)

Ničelni korelacijski koeficient spremenljivk X_1 in X_2 ne pomeni, da sta ti med seboj stohastično neodvisni: za vsako gostoto f_{X_1, X_2} , ki je soda funkcija odmikov $x_1 - \mu_1$ in $x_2 - \mu_2$, velja $\rho_{12} = 0$. Z drugimi besedami, $\rho_{12} = 0$ je zgolj potreben pogoj za neodvisnost, ni pa zadosten (glej spodnji del slike 7.8, ki ponazarja korelacijo ob primeru končnih vzorcev).

Čeprav med dvema spremenljivkama (naboroma vrednosti, meritvama, pojavoma) opazimo korelacijo, to ne pomeni nujno, da med njima obstaja neposredna vzročna zveza (causal relation): korelacija ne implicira kavzalnosti. Kadar imamo korelirani količini in med njima navidezno odvisnost, je najpogosteje vpleten neki skupni dejavnik.

Zgled: sposobnost računanja in velikost majice pri otrocih sta pozitivno korelirani. A če otroku kupimo večjo majico ne bomo vplivali oziroma izboljšali njegovega matematičnega znanja. V ozadju je skupni dejavnik : starost otroka

Korelacija in kavzalnost sta pojma iz različnih svetov: prvi pomeni ugotovitev na podlagi verjetnostne teorije, drugi pa opisuje strogo fizikalno reč, v katere ozadju sta čas in vzročna povezanost sedanjih dogodkov s preteklimi.

4.8.1 Multi variatne normalne porazdelitev

Zapišeli bi radi $f_{\vec{X}}(\vec{x})$ za mekorelirane ter korelirane $X^{(k)}$.

•) Za neodvisne $X^{(k)}$, torej mekorelirane $X^{(k)}$, je $f_{\vec{X}}$ ker produkt

$$f_{\vec{X}}(\vec{x}) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(X^{(1)} - \mu_1)^2}{2\sigma_1^2}\right) \cdot \dots \cdot \frac{1}{\sqrt{2\pi}\sigma_D} \exp\left(-\frac{(X^{(D)} - \mu_D)^2}{2\sigma_D^2}\right)$$

$$= \frac{1}{(2\pi)^D (\det \Sigma_D)^{1/2}} \exp\left(-\frac{1}{2} \vec{\delta}^T \Sigma_D^{-1} \vec{\delta}\right) \quad (*)$$

$$\vec{\delta} = \vec{X} - \vec{\mu}, \quad \Sigma_D = \begin{vmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_D^2 \end{vmatrix}, \quad \sigma_1 \dots \sigma_D = (\det \Sigma_D)^{1/2}$$

↓
residual, odstopanje od povprečne vrednosti

• Za korelirane $X^{(k)}$ z mediagonalno korenjeno mat

$$\vec{\Sigma} = \vec{X} - \vec{\mu}; \quad \overline{\varepsilon^{(k)} \varepsilon^{(k')}} = \Sigma^{kk'} : \quad \Sigma = \text{simetrične realne m.}$$

mojprej lahko poiščemo $\vec{\delta} = \vec{X} - \vec{\mu}$ ki ima diag kov. Σ_D

$$\vec{\Sigma} = U \vec{\Sigma}_D U^T, \quad \vec{\delta} = U \vec{\delta}$$

↓
kov. metrike
za $\vec{\Sigma}$
↓
kov. metrike
za $\vec{\delta}$
 $\Sigma_D^{ee'}$

$$\Sigma^{(kk')} = \overline{\varepsilon^{(k)} \varepsilon^{(k')}} = U_{ke} \overline{\delta^{(e)} \delta^{(e')}} U_{e'k'}^T = (U \Sigma_D U^T)_{kk}$$

in za mekorelirane $\vec{\delta}$ je porazdelitev ker produkt ki je zapisen zgoraj (*). Izrežimo ga z $\vec{\delta}$

$$\det \Sigma_D = \det \Sigma$$

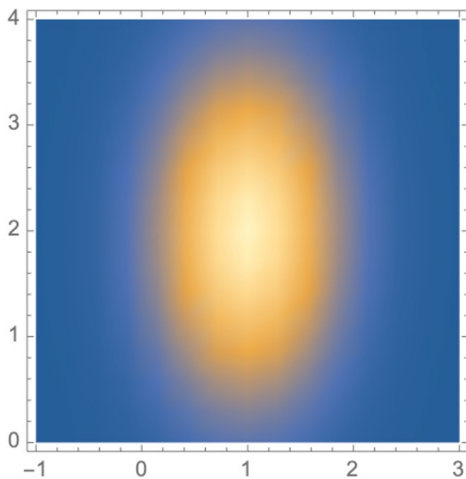
$$f = \frac{1}{\sqrt{2\pi}^D} (\det \Sigma)^{-\frac{1}{2}} \exp \left(-\frac{1}{2} \vec{\tilde{E}}^T \underbrace{U^{T-1} \Sigma_0^{-1} U^{-1}}_{(U \Sigma_0 U^T)^{-1} = \Sigma^{-1}} \vec{\tilde{E}} \right)$$

$$f_{\vec{x}}(\vec{x}; \vec{\mu}, \Sigma) = \frac{1}{\sqrt{2\pi}^D} (\det \Sigma)^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu}) \right)$$

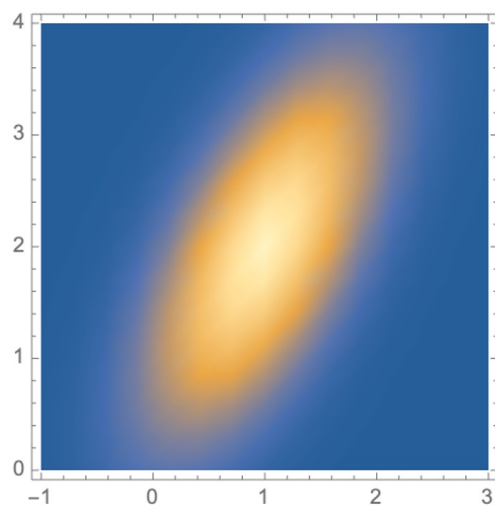
To je verjetnostna gostota za multivariatsko normalno porazdelitev z deno $\vec{\mu}$ in kovarnost. Σ .

`MultinormalDistribution[mu, cov]`

`mu = {1, 2};`
`cov = {{0.3, 0}, {0, 1}};`



`mu = {1, 2};`
`cov = {{0.4, 0.4}, {0.4, 1}};`



`Covariance[MultinormalDistribution[mu, cov]]`
`{{0.5, 0.4}, {0.4, 1}}`

4.9 Širjenje napak

$$\text{var}[X^{(k)} \pm X^{(l)}] = \text{var}[X^{(k)}] + \text{var}[X^{(l)}] \pm 2\text{cov}[X^{(k)}, X^{(l)}] \quad (4.20)$$

V mislih imejmo naključne spremenljivke $X^{(k)}$, ki so relativno ozko porazdeljene okoli povprečne vrednosti zaradi napak pri meritvi. Če bi znali formulo (4.20) posplošiti na poljubno funkcijo poljubnega števila spremenljivk, bi odgovorili na pomembno vprašanje širjenja napak. Kaj sploh pomeni "napaka naključne spremenljivke"? V uvodnih razdelkih smo spoznali, da si vsako meritev neke količine lahko predstavljamo kot realizacijo naključne spremenljivke, katere vrednost statistično fluktuiira. Tak naključen odklon od njene pričakovane vrednosti imenujemo statistična negotovost ali "napaka". Z analizo širjenja napak želimo ugotoviti, kako se negotovosti vrednosti določenega nabora spremenljivk odražajo v negotovosti vrednosti funkcije teh spremenljivk. Značilen primer je določitev moči na uporniku iz padca napetosti na njem: če je negotovost meritve napetosti ΔU , upornost R pa poznamo le na ΔR natančno, kolikšna je negotovost izračunane moči $P = U^2/R$?

Določimo napako ne skalarne količini $Y = Y(\vec{X}) = Y(x^{(1)}, \dots, x^{(n)})$ če poznamo korenično metriko Σ te spremenljivke $\vec{X}$. Obremerejmo primer ko so porazdelitve $\vec{X}$ dovolj ozke, torej so σ_{kk} majhne in zadošče prvi red v Taylorjevem raz.

$$Y(\vec{X}) \doteq Y(\vec{\mu}) + \sum_{k=1}^D (X^{(k)} - \mu_k) \left. \frac{\partial Y}{\partial X^{(k)}} \right|_{\vec{X}=\vec{\mu}} + \dots \quad (*)$$

Napako ne količini Y bo določala variance $\text{var}[Y(\vec{X})]$

$$\text{var}[Y(\vec{X})] = E[(Y(\vec{X}) - \underbrace{\overline{Y(\vec{X})}}_{Y(\vec{\mu})})^2] = E[(Y(\vec{X}) - Y(\vec{\mu}))^2] \quad (*)$$

ze ozke porazdelitve $\overline{Y(\vec{X})} \doteq Y(\vec{\mu}) = Y(\vec{\mu})$

$$= \sum_{ke} E[(X^{(k)} - \mu_k)(X^{(e)} - \mu_e)] \left(\frac{\partial Y}{\partial X^{(k)}} \frac{\partial Y}{\partial X^{(e)}} \right)_{\vec{X}=\vec{\mu}} = \sum_{ke} \sum_{ke} \left(\frac{\partial Y}{\partial X^{(k)}} \frac{\partial Y}{\partial X^{(e)}} \right)_{\vec{X}=\vec{\mu}}$$

$\text{Cov}[X^{(k)}, X^{(e)}] = \sum_{ke}$

za neodvisne spremenljivke je $\sum_{ke} = \sigma_k^2 \delta_{ke}$ in

$$\text{var}[Y(\vec{X})] = \sum_k \text{var}[X^{(k)}] \left(\frac{\partial Y}{\partial X^{(k)}} \right)_{\vec{X}=\vec{\mu}}^2$$

Zgled: A in B naj bosta neodvisni naključni spremenljivki.
 s pričakovanimi vrednostmi $\mu_{A,B}$ in variancama $\sigma_{A,B}^2$.
 Izračunaj varianco naključne spr. $C = A^m B^m$!

$$\bar{C} = \overline{A^m B^m} = \overline{A^m} \cdot \overline{B^m} \approx (\bar{A})^m \cdot (\bar{B})^m = \mu_A^m \mu_B^m$$

$\downarrow$ točno ker A in B neodvisni $\downarrow$ približno

Če neodvisne A in B vzamemo zvezo $\otimes$, zgorej:

$$\sigma_C^2 = \sigma_A^2 \left(\frac{\partial C}{\partial A} \right)^2 + \sigma_B^2 \left(\frac{\partial C}{\partial B} \right)^2 =$$

$$= \sigma_A^2 (m A^{m-1} B^m)^2 + \sigma_B^2 (m A^m B^{m-1})^2 \quad \Big|$$

$$A = \mu_A, B = \mu_B$$

$$= (\mu_A^m \mu_B^m)^2 \left[\sigma_A^2 \frac{m^2}{\mu_A^2} + \sigma_B^2 \frac{m^2}{\mu_B^2} \right]$$

$$\rightarrow \left(\frac{\sigma_C}{\mu_C} \right)^2 = m^2 \left(\frac{\sigma_A}{\mu_A} \right)^2 + m^2 \left(\frac{\sigma_B}{\mu_B} \right)^2$$

Pozor: to je za neodvisne A in B,
 če tudi nekaj približkov je v izrazu $\otimes$.

4.9.1 Posplošitev širjenja meroz na več funkcij in transformacija korenicne metrike

Določimo korenicno metriko meroz $\Sigma(\vec{Y})$ te spremenljivke $\vec{Y} = (Y^{(1)}, \dots, Y^{(m)})$, če so te funkcije medijčnih spr. $\vec{X}$: $Y^{(k)} = Y^{(k)}(X^{(1)}, \dots, X^{(n)})$

in pothemo korenicno metriko $\Sigma(\vec{X})$ in $E(\vec{X}) = \vec{\mu}$

$$Y^{(k)}(\vec{X}) = Y^{(k)}(\vec{\mu}) + \sum_{i=1}^n \left(\frac{\partial Y^{(k)}}{\partial X^{(i)}} \right)_{\vec{X}=\vec{\mu}} (X^{(i)} - \mu^{(i)}) \quad \begin{matrix} i=1, \dots, n \\ k=1, \dots, m \end{matrix}$$

$$\vec{Y}(\vec{X}) = \vec{Y}(\vec{\mu}) + D(\vec{X} - \vec{\mu}) \quad D_{ki} = \frac{\partial Y^{(k)}}{\partial X^{(i)}} \Big|_{\vec{X}=\vec{\mu}} : m \times n \text{ metrika}$$

$$\Sigma_{kl}(\vec{Y}) = E[(Y^{(k)}(\vec{X}) - Y^{(k)}(\vec{\mu})) (Y^{(l)}(\vec{X}) - Y^{(l)}(\vec{\mu}))]$$

$$= \sum_{i,j} \underbrace{\left(\frac{\partial Y^{(k)}}{\partial X^{(i)}} \right)_{\vec{X}=\vec{\mu}}}_{D_{ki}} \underbrace{\left(\frac{\partial Y^{(l)}}{\partial X^{(j)}} \right)_{\vec{X}=\vec{\mu}}}_{D_{lj}} E[\underbrace{(X^{(i)} - \mu^{(i)})(X^{(j)} - \mu^{(j)})}_{\Sigma_{ij}(\vec{X})}]$$

$$= \sum_{i,j} D_{ki} \Sigma_{ij}(\vec{X}) D_{lj}^T$$

$$\Sigma(\vec{Y}) = D \Sigma(\vec{X}) D^T$$

To je približna zveza, glej približna pri izpeljavi.

↓
kover. met $\vec{Y}$

↓
kover. metrike $\vec{X}$

D je metrike prvih odvodov

5. Posebne diskretne verjetnostne porazdelitve

V tem poglavju obravnavamo porazdelitve diskretnih naključnih spremenljivk, med katerimi sta najpomembnejši **binomska in Poissonova**.

5.1 Binomska porazdelitev

Z binomsko porazdelitvijo imamo opravka, kadar sta pri danem številu (Z) naključnih, med seboj neodvisnih, poskusov možni samo dve vrsti izidov: nekaj se zgodi (verjetnost p , recimo temu "ugoden" izid oz dogodek A) ali pa se ne zgodi (verjetnost $q = 1 - p$, recimo vsemu ostalemu "neugoden" izid oz dogodek $\bar{A}$). Pri metu kovanca pade grb ali cifra; rodi se deklica ali deček; pri metu kocke padeta 2 piki ali ne; jedro razpade ali ne; kvantnomehaniski proces se je zgodil ali ne. Takemu zaporedju neodvisnih poskusov pravimo tudi Bernoullijevo zaporedje.

- Z : celotno število poskusov
- N : naključna diskretna spremenljivka, ki šteje kolikokrat se je zgodil "ugoden" izid "tipa p ", številska vrednost te naključne spremenljivke N je n .

Verjetnost, da pri Z neodvisnih poskusih najprej n -krat zaporedoma dobimo ugoden izid in potem $(Z-n)$ krat zaporedoma neugoden izid je: $p^n (1-p)^{Z-n}$. Verjetnosti smo kar zmnožili ker gre za neodvisne poskuse.

Presteti moramo le se na koliko načinov lako izmed Z poskusov izberemo n ugodnih in $Z-n$ neugodnih: to je podano z Binomskim simbolom

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

Utemeljimo: prvi ugoden izid izbiramo med Z poskusi, drugi ugoden izid med preostalimi $(Z-1)$, n -ti ugoden izid med preostalimi $(Z-n+1)$, torej smo našli

$Z \cdot (Z-1) \cdot \dots \cdot (Z-n+1) = \frac{Z!}{(Z-n)!}$ načinov izbora. A pri tem smo šteli različna zaporedja izbranih izidov kot različna, pa so enaka enaka, zato moramo število načinov deliti z $n!$

Zaključimo: Verjetnost za n ugodnih izidov pri Z neodvisnih poskusih je

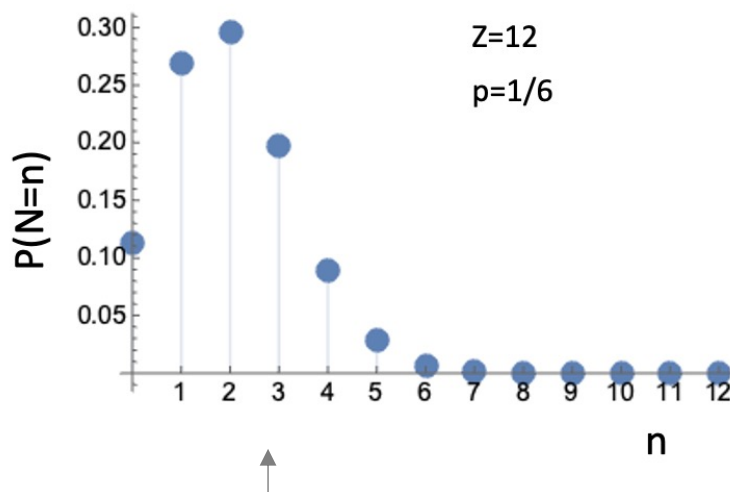
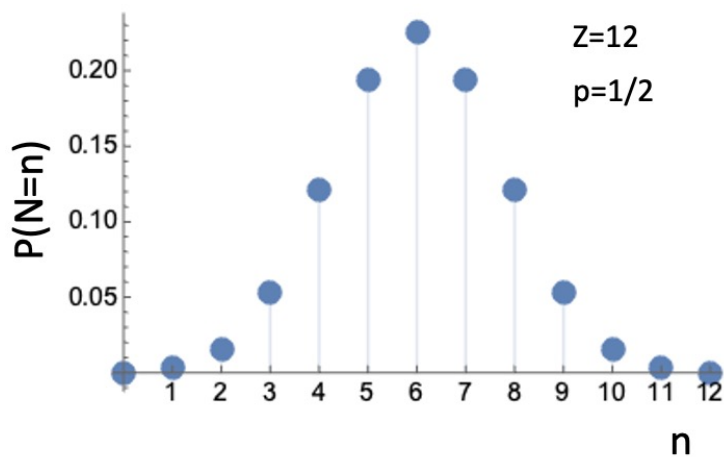
$f_N(n; Z, p) = P(N = n; Z, p) = P(N = n) = \binom{Z}{n} p^n (1-p)^{Z-n}$	Binomska porazdelitev
---	-----------------------

kjer je p verjetnost ugodnega izida pri posameznem poskusu. Z in p sta parametra porazdelitve, ponavadi ju od argumenta ločimo z dvopicjem.

Porazdelitev je seveda pravilno normirana

$$\sum_{n=0}^Z P(N = n) = \sum_{n=0}^Z \binom{Z}{n} p^n q^{Z-n} = (p + q)^Z = 1^Z = 1$$

Zgornjih oznak se bova pri predavanjih in vajah z Miho Mihovilovicem držala (vsaj skozi celotno poglavje 5). Kot ponavadi isto crko uporabljamo za določeno spremenljivko v veliko zacetnico (N) in malo zacetnico (n). Opozorilo: zgornji N je v knjigi Verjetnost označen z X ; zgornji Z je v knjigi Verjetnost označen z N .



DiscretePlot[PDF[**BinomialDistribution[12, 1/6]** , n], {n, 0, 12}]

Zgled Vržemo 12 poštenih igralnih kock. Kolikšna je verjetnost, da bo število pik 3 nastopilo največ enkrat ($N \leq 1$)?

Verjetnost za ugodni izid "3 pike" pri vsakem metu je $p=1/6$, verjetnost za kateri koli drugi izid pa $q=1-p=5/6$. Iskana verjetnost je

$$P(n \leq 1) = P(n=0) + P(n=1) = \binom{12}{0} \cdot \left(\frac{1}{6}\right)^0 \cdot \left(\frac{5}{6}\right)^{12} + \binom{12}{1} \cdot \left(\frac{1}{6}\right)^1 \cdot \left(\frac{5}{6}\right)^{11} = 0.11 + 0.27 = 0.38$$

$Z=12$



Zgled Čas čakanja v vrsti v menzi je naključna spremenljivka, porazdeljena eksponentno s povprečjem 4 minute. Kolikšna je verjetnost, da bo študent v vsaj štirih dneh od naslednjih šestih postrežen v manj kot treh minutah?

Verjetnost, da na izbrani dan študent v 4 minutah še *ni* postrežen, eksponentno pada: $P(t) = e^{-t/\tau}$, kjer je $\tau = 4$ min. Verjetnost, da je postrežen prej kot v treh minutah, je $p = 1 - e^{-3/4}$. Upoštevati moramo le še vse možnosti v zaporednih dneh, kar opisuje binomska porazdelitev. Vsak dan sta namreč samo dve možnosti: postrežen je prej kot v 3 minutah (verjetnost p) ali kasneje (verjetnost $1 - p$). Iskana verjetnost je torej

$$\sum_{n=4}^6 \binom{6}{n} p^n (1-p)^{6-n} = \sum_{n=4}^6 \binom{6}{n} \left(1 - e^{-3/4}\right)^n \left(e^{-3/4}\right)^{6-n} \approx 0.3969.$$

opomba: tu verjetnost p ni majhna, in tudi število poskusov čakanja v menzi ($Z=6$) ni majhno zato uporaba Poissonove porazdelitve namesto Binomske ni upravičena.

Povprečje in varianca pri binomski porazdelitvi

Želimo določiti povprečje in varianco meključme spr N ki pove število "ugodnih" izidov izmed z poskusov:

$$\bar{N} = \sum_{m=0}^z m P(N=m) \quad \sigma_N^2 = \sum_{m=0}^z (m - \bar{N})^2 P(N=m) = \overline{N^2} - \bar{N}^2$$

Te vsote najlažje izračunamo s "trikom" ko nadomestimo $p \rightarrow dp$ in $q \rightarrow dq$ pri binomski porazdelitvi in odvodom po d

$$\sum_m \binom{z}{m} (dp)^m q^{z-m} = (dp + q)^z$$

1. odvod po d

$$\sum_m \binom{z}{m} m d^{m-1} \overbrace{p^n}^{P(m)} q^{z-m} = z (dp + q)^{z-1} \cdot p$$

$\downarrow d=1$

Pri $z=12$ metu kocke zgoraj
je povprečje števila ugodnih
izidov s tremi pikami res
 $\bar{N} = 12 \cdot \frac{1}{6} = 2$

$$\boxed{\bar{N} = \sum m P(m) = z \cdot p} \quad \text{pričakovani rezultat}$$

2. odvod po d :

$$\sum \binom{z}{m} m(m-1) d^{m-2} p^m q^{z-m} = z(z-1) (dp + q)^{z-2} p^2$$

$\downarrow d=1$

$$\overline{N(N-1)} = z(z-1) p^2 \rightarrow \overline{N^2} = z(z-1) p^2 + \overbrace{z \cdot p}^{\bar{N}}$$

$$\boxed{\sigma_N^2 = \overline{N^2} - \bar{N}^2 = z(z-1)p^2 + z p - z^2 p^2 = -z p^3 + z p = z p (1-p) = z p q}$$

če σ_N razumemo kot negotovost izmerjenega števila dogodkov zapišemo kar

$$N_{izm} = N \pm \sigma_N = Z p \pm \sqrt{Z p q}$$

5.2 Multinomnske porazdelitve

"bi" nomnske $\rightarrow$ 2 razrede (verjetnost p za A , $1-p$ za $\bar{A}$)

"multi" nomnske $\rightarrow$ k razredov z verjetnostmi $p_i, i=1, \dots, k, \sum_{i=1}^k p_i = 1$

Ostajamo pri neodvisnih poskusih.

Neključne spremenljivke N_i naj predstavljajo število izidov vrste i , n_i pa so njihove številske vrednosti.

Verjetnost za n_1 izidov vrste 1, ..., n_k izidov vrste k

$$P(n_1, \dots, n_k) = \binom{z}{n_1, \dots, n_k} p_1^{n_1} \dots p_k^{n_k} \quad \sum_{i=1}^k n_i = z$$

$f_{\vec{N}}(\vec{n}) =$

↓
multinomski simbol $\binom{z}{n_1, \dots, n_k} = \frac{z!}{n_1! \dots n_k!}$ predstavlja

število permutacij če so n_1 med seboj medločljivi, ..., n_k med seboj medločljivi.

V kakšnem primeru multinomske preide v binomsko?

Predstavljajmo si i -ti izid kot ugoden (A), ostale kot neugodne ($\bar{A}$)

Torej je N_i porazdeljene po binomski porazdelitvi z

$$\bar{N}_i = z \cdot p_i \quad \sigma_{N_i}^2 = z p_i (1 - p_i)$$

in teh pričakovanih vrednosti ni potrebno še enkrat določiti
na podlagi multinomske porazdelitve. Slednje potrebujemo
le za določitev medianskih členov kovariacijske matrike,
tu brez izpeljave prevedemo

$$\boxed{\sum_{i,j} \text{cov}[N_i, N_j] = E[(N_i - \bar{N}_i)(N_j - \bar{N}_j)] = \dots = -z p_i p_j} \quad i \neq j$$

$$\Sigma_{ij} = \text{cov}[N_i, N_j] = \begin{cases} -Z p_i p_j & i \neq j \\ Z p_i (1 - p_i) & i = j \end{cases}$$

$$\Sigma = \begin{pmatrix} Z p_1 (1 - p_1) & \dots & -Z p_1 p_k \\ \vdots & & \vdots \\ -Z p_k p_1 & \dots & Z p_k (1 - p_k) \end{pmatrix}$$

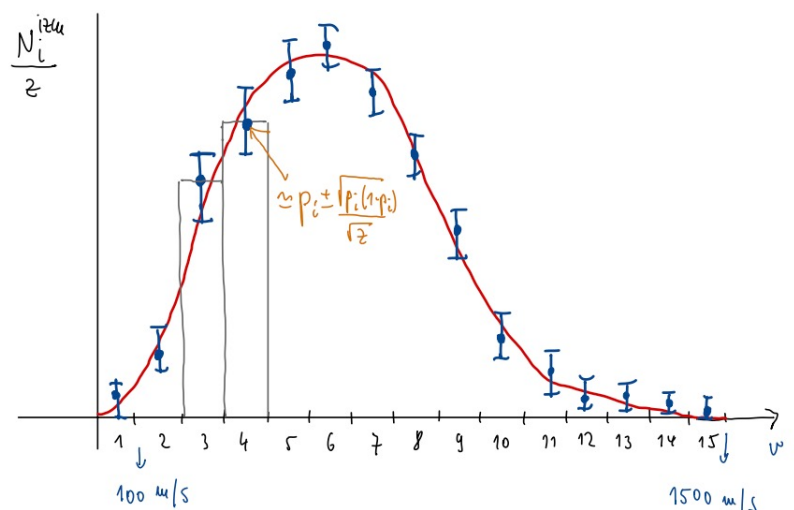
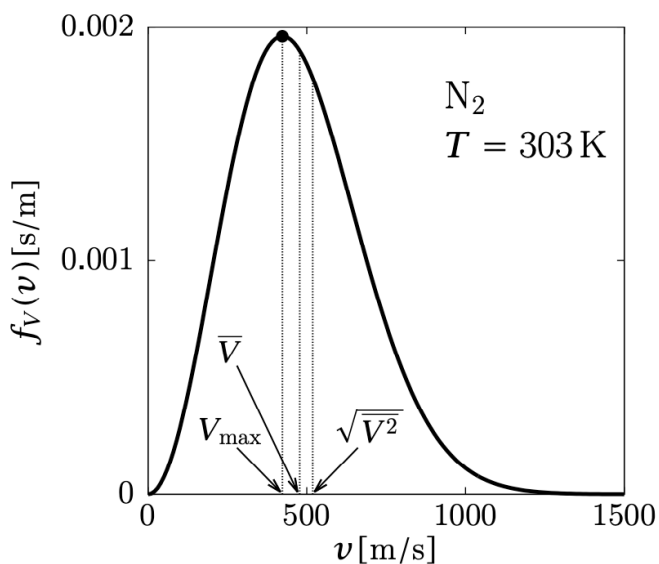
Zgled Merimo porazdelitev molekul po hitrostih; pričakujemo rezultat, podoben sliki 4.1 (desno). Meritve uredimo v histogram s $k = 15$ ekvidistančnimi razredi

1: $[0, 100]$ m/s, 2: $[100, 200]$ m/s ... 15: $[1400, 1500]$ m/s.

V posamezni razred pade n_i molekul; vsi razredi, v katere lahko pade posamezna hitrost, so med seboj neodvisni. Skupno preštejemo $Z = n_1 + n_2 + \dots + n_{15}$ molekul. Tak histogram — ob vsaki novi meritvi bo videti drugače — predstavlja multinomsko porazdelitev. Relativna napaka na razmerju N_i^{izm} pada korensko z večanjem števila meritev Z .

$$N_i^{\text{izm}} = Z p_i \pm \sqrt{Z p_i (1 - p_i)}$$

$$\frac{N_i^{\text{izm}}}{Z} = p_i \pm \frac{\sqrt{p_i (1 - p_i)}}{\sqrt{Z}}$$



5.5 Poissonove porazdelitve

zelo pomembne za fiziko!!

Poissonove porazdelitve je limite binomske v kateri zmanjšujemo verjetnost posameznega izida ($p \rightarrow 0$) in hkrati povečujemo število vseh poskusov ($z \rightarrow \infty$), vendar tako da ostane povprečno število ugodnih izidov $\bar{N} = p \cdot z$ konstantno.

Opazimo jo tako da v binomski izrazimo $p = \frac{\bar{N}}{z}$ in $z \rightarrow \infty$

$$P = \binom{z}{m} p^m (1-p)^{z-m} = \binom{z}{m} \left(\frac{\bar{N}}{z}\right)^m \left(1 - \frac{\bar{N}}{z}\right)^{z-m}$$

$$\lim_{z \rightarrow \infty} \left(1 - \frac{\bar{N}}{z}\right)^{z-m} = \lim_{z \rightarrow \infty} \left(1 - \frac{\bar{N}}{z}\right)^z \left(1 - \frac{\bar{N}}{z}\right)^{-m} = e^{-\bar{N}} \cdot 1$$

$$\lim_{z \rightarrow \infty} \binom{z}{m} \frac{1}{z^m} = \frac{z(z-1) \cdots (z-m+1)}{m! z^m} = \lim_{z \rightarrow \infty} \frac{1}{m!} \left(1 - \frac{1}{z}\right) \cdots \left(1 - \frac{m-1}{z}\right) = \frac{1}{m!}$$

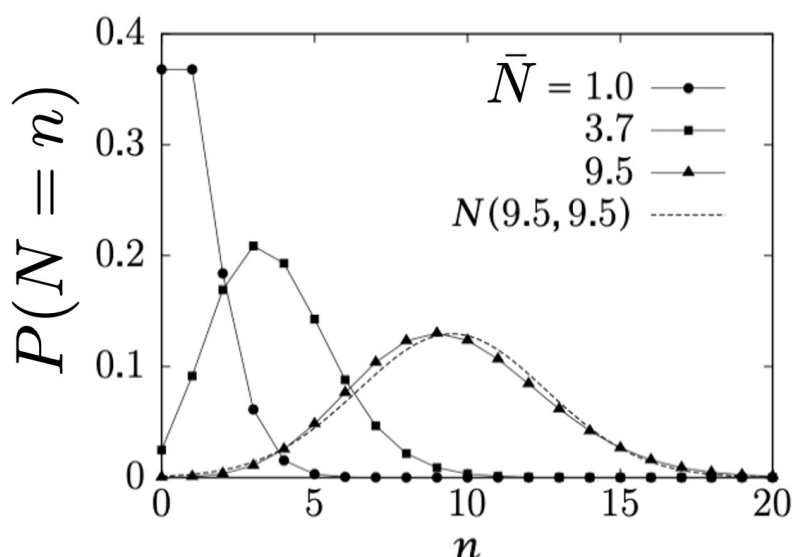
Povzemimo: **Obravnavamo veliko število (Z) neodvisnih poskusov, kjer sta pri vsakem poskusu možni le dve vrsti izidov: nekaj se zgodi (z verjetnostjo p) ali se pa ne zgodi (z verjetnostjo $1-p$). V limiti velikega števila poskusov ($Z \rightarrow \infty$) in majhne verjetnosti $p \rightarrow 0$, je verjetnost za n ugodnih izidov pri danem povprečju ugodnih izidov $\bar{N}$ podana s spodnjo Poissonovo porazdelitvijo:**

$$P(N=n) = \frac{1}{n!} \bar{N}^n e^{-\bar{N}}$$
$$f_N(n; \bar{N}) =$$

POISSONOVA PORAZDELITEV

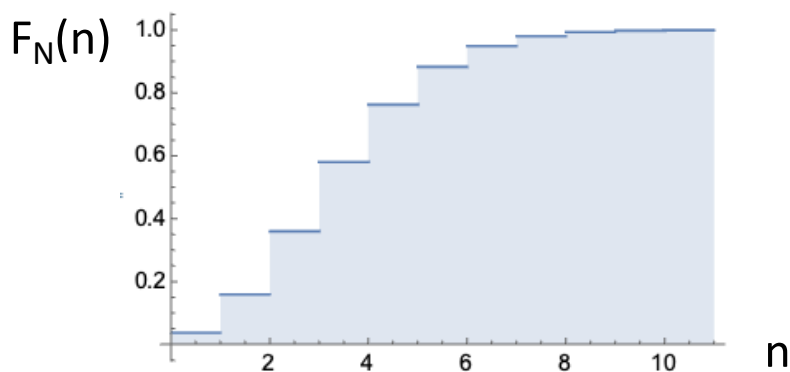
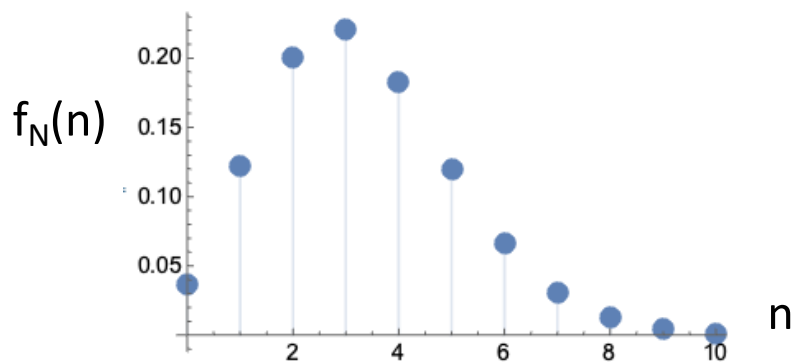
$$n = 0, 1, \dots$$

edini parameter binomske porazdelitve je povprečno število ugodnih izidov $\bar{N}$ v danem številu neodvisnih poskusov



pogosto se povprečno število $\bar{N}$ označi $\bar{N} = \lambda$
(tako je tudi v knjigi S.Sirca)

```
Nav = 3.3;
DiscretePlot[PDF[PoissonDistribution[Nav], n], {n, 0, 10},
DiscretePlot[CDF[PoissonDistribution[Nav], n], {n, 0, 10},
```



Pričakovane vrednosti povprecja in variance za Poissonovo porazdelitev določimo kar iz vrednosti, ki smo jih določili za Binomsko porazdelitev, tako da uporabimo $q=1-p \approx 1$. Vse bomo izrazili z $\bar{N}$, ki je edini parameter te porazdelitve.

binomska

Poissonova

$$E[N] = pZ \rightarrow E[N] = \bar{N},$$

binomska

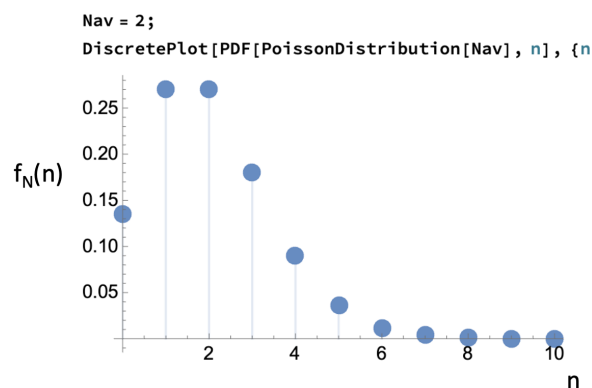
Poissonova

$$\sigma_N = \sqrt{Zpq} \rightarrow \sigma_N = \sqrt{\bar{N}}$$

Zgled (str 116) Cepimo $Z = 2000$ ljudi. Verjetnost p za stranske učinke pri cepljenju je majhna, $p = 0.001$. V povprečju bosta torej stranske učinke opazila samo $\underline{N} = Zp = 2$ človeka. Kolikšna je verjetnost, da bo število ljudi s stranskimi učinki večje od 2? Verjetnost, da natanko n ljudi izkusi kak stranski učinek, je

$$P(N=n) = \frac{1}{n!} \bar{N}^n e^{-\bar{N}}$$

Iskana verjetnost je torej $P(n > 2) = P(3) + P(4) + \dots + P(2000)$. Računanju in seštevanju teh 1998 vrednosti se izognemo tako, da izračunamo verjetnost za nasprotni dogodek: $P(n > 2) = 1 - P(0) - P(1) - P(2) \approx 0.323$. ▸



Zgled (str 116) Zemeljsko površje doseže v povprečju 25 meteoritov na dan. Kolikšna je verjetnost, da bo v desetih letih najmanj enega od $7 \cdot 10^9$ prebivalcev Zemlje zadel meteorit? Geometrijski presek človeškega telesa je približno $S_1 \approx 0.2 \text{ m}^2$, polmer Zemlje pa $R = 6400 \text{ km}$.

Najprej poiscimo odgovor na zastavljeno vprasanje pragmaticno, ob predpostavki da je v igri Poissonova verjetnostna porazdelitev. Da je ta upravicena bomo utemeljili v naslednjem koraku. Poissonova porazdelitev je enoparametricna, z enim samim parametrom $\underline{N}$. V tem primeru je relevanten

$\underline{N}$ = povprečno število zadetkov meteoritov na clovestvo v desetih letih,

$$\underline{N} = 25 * (10 \text{ let} / 1 \text{ dan}) * 7 \cdot 10^9 * S_1 / (4 * \pi * R^2) = 9125 * 7 \cdot 10^9 * S_1 / (4 * \pi * R^2) = 0.248$$

$$\text{ste\u0161ilo meteoritov na 10 let} = 25 * (10 \text{ let} / 1 \text{ dan}) = 91250$$

Nakljucna spremenljivka N z vrednostjo n pa ustreza številu zadetkov meteoritov na clovestvo v desetih letih (število "ugodnih izidov" kar se tice razmisleka pri tej nalogi), verjetnost za $N=n$ pa je

$$P(N=n) = \frac{1}{n!} \bar{N}^n e^{-\bar{N}} = f_N(n)$$

Verjetnost, da zadane vsaj enega je 1 minus verjetnost da ne zadane nobenega

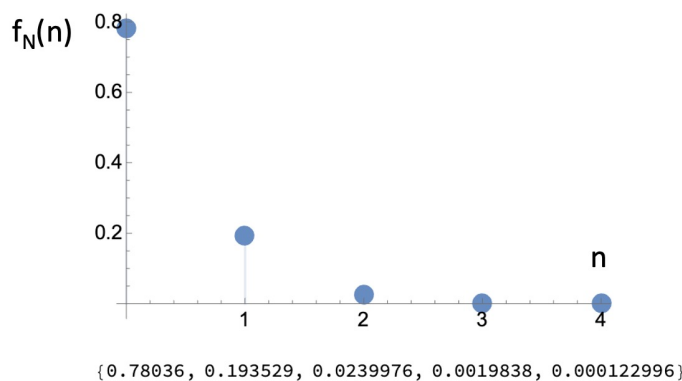
$$P=P(n=1)+P(n=2)+\dots=1-P(n=0)=1 - \bar{N}^0 e^{-\bar{N}}/0! = 1-e^{-0.248} \approx 0.22$$

Zdaj še utemeljimo uporabo Poissonove porazdelitve, ki se jo za ta primer da utemeljiti na več različnih načinov. Vzemimo celotno obdobje $t=10$ let in vsak padec meteorita naj bo neodvisen "poskus".

- Torej je v igri veliko število $Z=25 \cdot (10 \text{ let} / 1 \text{ dan}) = 91250$ neodvisnih poskusov, kar utemeljuje prvo od predpostavk Poissonove porazdelitve.
- Verjetnost p za "ugoden dogodek" (to je zadetek na človeštvo pri padcu enega meteorita) za en poskus (to je padec enega meteorita) pa je $p=7 \cdot 10^9 \cdot S1/(4 \cdot \pi \cdot R^2)=2.7 \cdot 10^{-6}$. Torej je verjetnost za ugoden dogodek pri enem poskusu res $p \ll 1$ in je Poissonova aproximacija binomske porazdelitve utemeljena.

Povprečno število "ugodnih dogodkov" je torej $\bar{N}=Z p=0.248$, kot prej, zgornji razmisek pa utemeljuje uporabo Poissonove porazdelitve "ugodnih dogodkov".

```
Nav = 0.248;  
DiscretePlot[PDF[PoissonDistribution[Nav], n], {n, 0, 4}  
Table[PDF[PoissonDistribution[Nav], n], {n, 0, 4}]
```



Z mislimi na prejšnji zgled **izpeljimo splošnejšo verjetnost za število n ugodnih dogodkov v času t** . Z naj bo število vseh neodvisnih poskusov v tem času: primeri

- (a) Z =število meteoritov (v $t=10$ let)
- (b) Z =število jeder pri poskusu (vsako jedro predstavlja neodvisen poskus ker lahko razpade ali ne) predpostavka v nadaljevanju: število Z ne upoade zaznavno v času t , tako da se aktivnost ne spreminja
- (c) Z = število trkov proton-proton v LHC (na primer v $t=1$ dan)

Za vse te primere je Z velik in je Poissonova porazdelitev s tega stališča upravičena.

Verjetnost p za ugoden dogodek pri enem od Z poskusov naj bo majhna

$$p=a_1 \cdot t, \quad a_1=\text{aktivnost za en meteorit, jedro ali en kvantno-mehaniski proces} = dN_1/dt$$

- (a) aktivnost enega meteorita pri zadetku človeštva
- (b) $a_1=dN_1/dt$ =aktivnost enega jedra= $1/\tau=\lambda$
- (c) a_1 =verjetnost za proces ($p \rightarrow$ končno stanje) pri trku enega para na enoto časa

Tedaj je verjetnost za n ugodnih dogodkov v času t podana s Poissonovo porazdelitvijo z $\bar{N}=p Z = a_1 \cdot t \cdot Z = a \cdot t$, kjer je $a=Z \cdot a_1$ aktivnost celotnega vzorca

$$P(n) = \frac{1}{n!} (at)^n e^{-at}$$

To velja le, če se aktivnost vzorca v času t ne spremeni zaznavno. $P(n)$ tu predstavlja

- (a) n = število zadetkov meteoritov na clovestvo
- (b) n = število razpadov vseh jeder
- (c) n = število opaženih procesov $p \rightarrow$ končno stanje, torej določenih končnih stanj

Vse to v danem času t

Verjetnost, da se ni zgodil noben razpad do časa t je v skladu z eksponentno porazdelitvijo, kjer pomnožimo Zkrat verjetnost za preživetje vsakega izmed jeder

$$P(n = 0) = e^{-at} = e^{-a_1 Zt} = (e^{-a_1 t})^Z, \quad a_1 = 1/\tau = \lambda$$

Klasični zgled V času t zabeležimo n jedrskih razpadov; naj bo $n \gg 1$ (vsaj nekaj deset). Ocena za pravo aktivnost vzorca a (število razpadov celotnega vzorca na časovno enoto) je $a' = n/t$, kjer crtica označuje da gre za oceno. Izmerjeni n opleta okrog $\bar{N} = at$ za $\pm \sigma_N = \pm \sqrt{\bar{N}}$ toda prave vrednosti a in $\bar{N}$ ne poznamo zato naredimo približek $\sigma_N \sim \sqrt{n}$. Pisemo torej

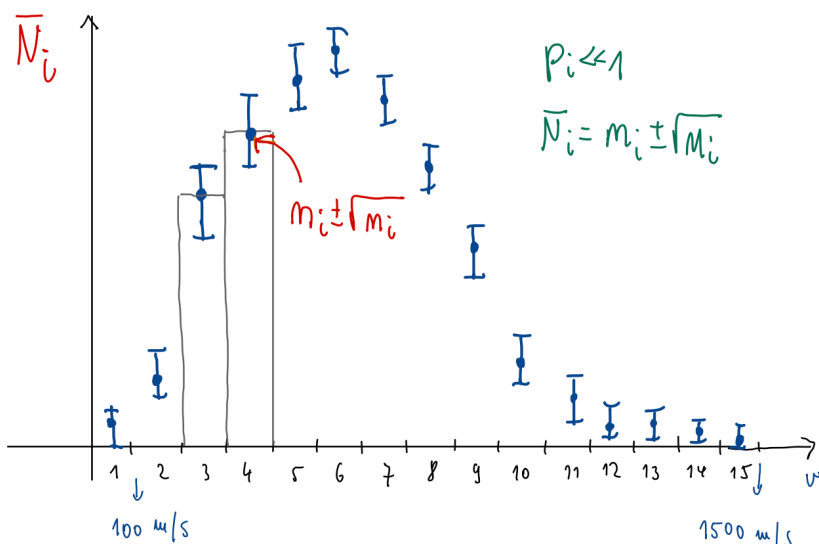
$$n = \bar{N} \pm \sqrt{\bar{N}} \simeq \bar{N} \pm \sqrt{n} \rightarrow \bar{N} = n \pm \sqrt{n}$$

kjer smo desno enačbo dobili s tem, da smo izrazili $\bar{N}$ v levi enacbi.

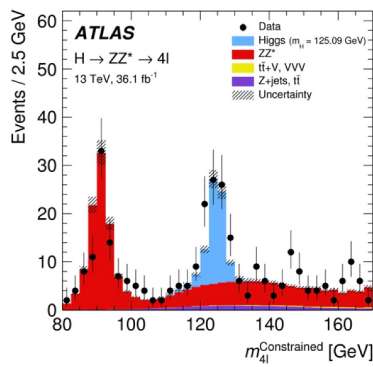
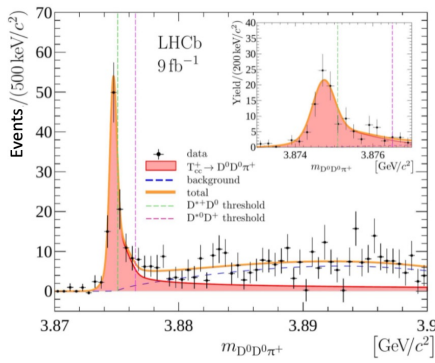
Po deljenju druge enačbe s t ugotovimo, za koliko se v povprečju razlikuje prava aktivnost a od izmerjene vrednosti a' :

$$a = \frac{\bar{N}}{t} = \frac{n \pm \sqrt{n}}{t} = \frac{n}{t} \left(1 \pm \frac{1}{\sqrt{n}}\right) = a' \left(1 \pm \frac{1}{\sqrt{n}}\right)$$

Če hočemo izmeriti aktivnost vzorca na 1 % natančno, moramo torej prešteti 104 razpadov. Ta mlinski kamen nam visi okrog vratu pri vseh poskusih, kjer kar koli "štejemo". Za k -kratno zmanjšanje statistične napake moramo prešteti k^2 -krat več dogodkov oziroma meriti k^2 -krat dlje.



Zgled: vrnilo se k zgledu meritve števila molekul v določenem hitrostnem razredu: če izmerimo n_i molekul v i -tem hitrostnem razredu, povprečje N_i določimo po zgornji zvezi $N_i = n_i \pm \sqrt{n_i}$ kjer smo privzeli da je $q_i = 1 - p_i \sim 1$.



Zgled: tipična določitev napake pri eksperimentih v trkalnikih. Ce izmerimo v razdelku (binu) histograma n procesov ($p \rightarrow$ določeno končno stanje) določimo napako na pravem povprečnem številu dogodkov $\underline{N}$ kar po prejšnji formuli $N = n \pm \sqrt{n}$. Tako so dolovčne najverjetneje tudi napake v levih histogramih (ki predstavljajo signal za eksotični hadron in Higgsov delec)

5.4 Aproximacija binomske in Possonove porazdelitve z normalno

Če je Z velik in niti p niti q nista preblizu nič, je mogoče binomsko porazdelitev zelo dobro aproksimirati z normalno, čeprav je prva diskretna, druga pa zvezna! Ta približek uteleša Laplaceov limitni izrek, ki ga ne bomo dokazali (dokazan je v dodatku B.3.1).

$$\lim_{Z \rightarrow \infty} P\left(a \leq \frac{N - Zp}{\sqrt{Zpq}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

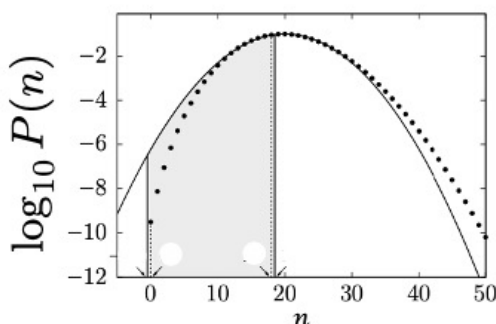
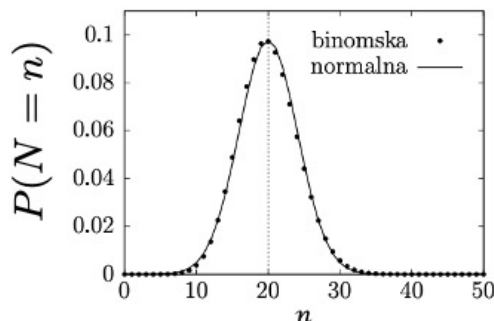
Z drugimi besedami, standardizirana binomska spremenljivka

$$\frac{N - \bar{N}}{\sigma_N}$$

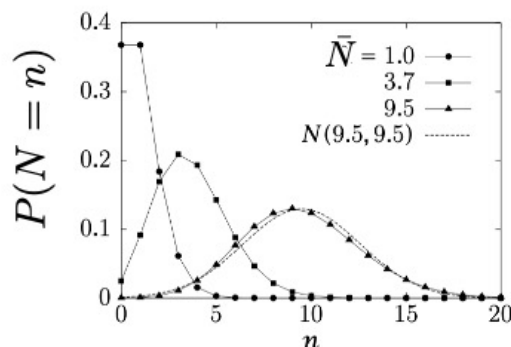
je asimptotsko porazdeljena po normalni porazdelitvi, kar v praksi dobro deluje že pri $Zp, Zq > 5$.

Ker je Poissonova limita binomske porazdelitve, to velja tudi za Poissonovo porazdelitev za dovolj velik $\underline{N} = Zp$, v praksi deluje dovolj dobro že za $\underline{N} > 5$.

binomska: $Z=120, p=1/6$



Poissonova: $\underline{N}=9.5$



6. KONVOLUCIJE, STABILNE (GAUSSOVE) PORAZDELITVE IN CENTRALNI LIMITNI IZREK

To pogleduj obremenjevalne naslednji posamezni problem:

Ymejmo: neodvisne naključne sp. X_i , poznamo $f_{X_i}(x_i)$, $i=1, \dots, D$

Vpr: kakšna je porazd. $f_Z(z)$ njihove vsote $Z = \sum X_i$ ali povprečja?

Viodei bomo: porazdelitev Z padeja KONVOLUCIJA sp. X

V primeru $n \rightarrow \infty$ ob določenih pogojih porazdelitve Z tečijo k stabilnim (Gaussovim) porazdelitvam.

Poskus: met D kock, porazdelitev vsote in povprečja pik

Komentar glede motecij: bolj konsistentno bi bilo namesto

X_i pisati $X^{(i)}$ in vrednosti x_i pisati $x^{(i)}$, saj so $i=1, \dots, D$

različne spremenljivke (ki so lahko porazdeljene po enakih porazdelitvah f_X ali različnih $f_{X_i} = f_i$). A zapiši bodo ustreznosti z spodnjimi indeksi in brez oklepajev.

6.1 KONVOLUCIJA DISKRETNIH PORAZDELITEV

Ymejmo: N_1 porazdeljen po $f_{N_1}(m_1)$ ter $\left. \begin{matrix} N_2 \\ f_{N_2}(m_2) \end{matrix} \right\}$ neodvisni

Določimo porazdelitev $f_N(m)$ za $N = N_1 + N_2$

Verjetnost za izid m_1, m_2 je $f_{N_1}(m_1) \cdot f_{N_2}(m_2)$ ker sta X_1 in X_2 neodv.

Zenimo nas verjetnost za izid m , zato se znebimo $m_2 = m - m_1$ in sestavimo po m_1

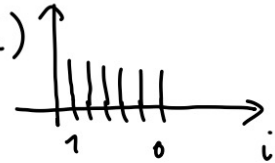
$$N = N_1 + N_2: f_N(m) = \sum_{m_1} f_{N_1}(m_1) f_{N_2}(m - m_1) = (f_{N_1} * f_{N_2})(m)$$

$$N = N_1 + N_2 + N_3: f_N(m) = \sum_{m_1, m_2} f_{N_1}(m_1) f_{N_2}(m_2) f_{N_3}(m - m_1 - m_2) = (f_{N_1} * f_{N_2} * f_{N_3})(m)$$

Operacijo imenujemo konvolucija poreaditev f_{N_1} in f_{N_2} in označimo z $*$.

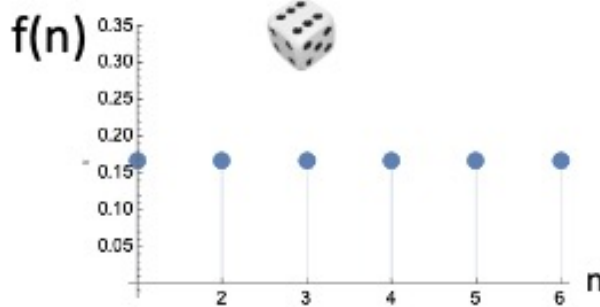
Zgled: Metemo D kock, zapišemo mas $f_N(m)$ za $N = N_1 + \dots + N_D$

Poreaditev za vselo kocko je enakomerna: $f(i)$



$$(f * f)(m) = \sum_{m_1=1}^{m-1} f(m_1) f(m - m_1) = \frac{1}{6} \cdot \frac{1}{6} \times \begin{cases} m=2, m_1=1: 1 \\ m=3, m_1=1,2: 2 \end{cases}$$

tako por.
suo 2 kocket
riseli



D = 2 kocki

$$f_N(m) = (f * f)(m)$$

↓ ↓
enakovredni



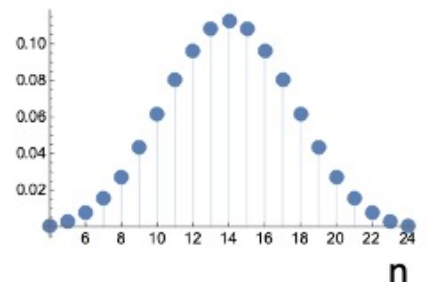
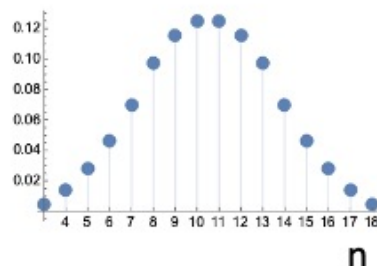
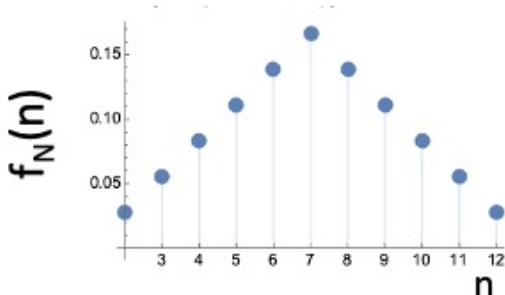
D = 3

$$f_N = f * f * f$$



D = 4

$$f_N = f * f * f * f$$



datoteke: konvolucija.nb

Najverjetnejše in povprečno število vsote pik pri metu D kock je vedno $D \cdot 3.5$, kar je v skladu z pričakovanjem glede na met ene koncke, kjer je povprečno število $\frac{1}{6} \sum_{i=1}^6 i = 3.5$

Zgled: Pokažimo, da je konvolucija dveh Poissonovih por. zpet Poiss.

Trditve: $f_{N_1}(m_1; \bar{N}_1) * f_{N_2}(m_2; \bar{N}_2) = f_N(m, \bar{N}_1 + \bar{N}_2)$ $m = m_1 + m_2$

Poisson Poisson Poisson

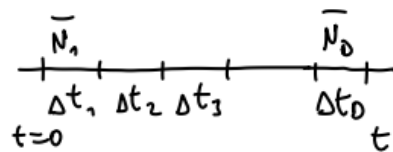
Doker: $f_N(m) = \sum_{m_1=0}^m f_{N_1}(m_1; \bar{N}_1) f_{N_2}(m-m_1; \bar{N}_2)$ zgoraj je meja
je m ker
 $m_{1,2} \geq 0$

$$f_N(m) = \sum_{m_1=0}^m \frac{\bar{N}_1^{m_1} e^{-\bar{N}_1}}{m_1!} \frac{\bar{N}_2^{(m-m_1)} e^{-\bar{N}_2}}{(m-m_1)!}$$

$$= e^{-(\bar{N}_1 + \bar{N}_2)} \frac{1}{m!} \sum_{m_1=0}^m \frac{m!}{m_1! (m-m_1)!} \bar{N}_1^{m_1} \bar{N}_2^{m-m_1} = f_N(m; \bar{N}_1 + \bar{N}_2)$$

Poisson

Primer uporabe:



če je porazdelitev v vseh časovnih intervalih $\Delta t_i, i=1..D$ Poissonovske z enakim povprečnim številom ugodnih dogodkov $\bar{N}_i = \bar{N}$ na vseh intervalih (aktivnost se tokom $[0, t]$ ne spreminja), potem je porazdelitev celotnega števila ugodnih dogodkov $m = m_1 + \dots + m_D$ podena zpet s Poissonovo porazdelitvijo $f_N(m)$ s povprečnim $\bar{N} = D \cdot \bar{N}_1 = \frac{t}{\Delta t} \bar{N}_1 = t \cdot r$, $r = \frac{\bar{N}_1}{\Delta t}$

$$f_N(m) = \frac{\bar{N}^m}{m!} e^{-\bar{N}} = \frac{(r \cdot t)^m}{m!} e^{-r \cdot t}$$

št. dogodkov
v $[0, t]$ pri dani
povprečni $\bar{N}$

št. dogodkov
v $[0, t]$ pri dani
aktivnosti $r = \frac{dN}{dt}$
celotnega vzorca

↑
upr: r je aktivnost celotnega vzorca
= št. razpadov vseh jeder v dt
z tem dt: $r = \frac{dN}{dt} = n_1 \cdot z$
aktivnost enega jedra

Če se aktivnost r v času med 0 in t ne spreminja zanesljivo, potem iz Poissonove por. v dt dobimo Poissonovo v času $[0, t]$.

6.1 KONVOLUCIJA ZVEŽNITI PORAZDELITEV

Imamo: Neodvisni zvezni in neodvisni spremenljivki X_1 in X_2
z danih verjetnostnih porazdelitvama

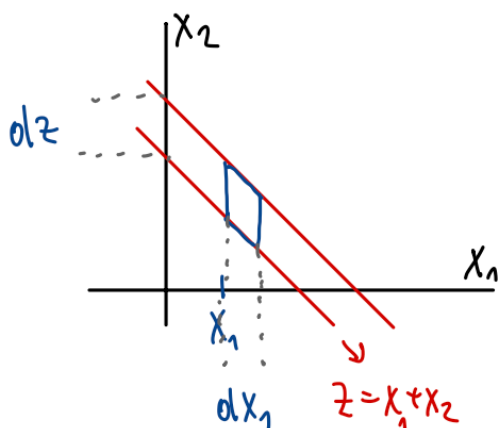
Zanima nas se za njuno vsoto: $Z = X_1 + X_2$

V preteklih poglavjih smo že spoznali kaj velja za
povprečje in varianco Z , torej za določene pričakovane vr.

$$\bar{Z} = \bar{X}_1 + \bar{X}_2, \quad \text{var}[Z] = \text{var}[X_1] + \text{var}[X_2] + \underbrace{2\text{cov}[X_1, X_2]}_{0 \text{ ker sta neodvisni}}$$

V tem poglavju pa nas ne zanimajo le pričakovane
vrednosti, temveč PORAZDELITEV $f_Z(z)$ same!

Določimo verjetnost da vsota $Z = X_1 + X_2$ pade na interval $[z, z+dz]$.
Vse točke, ki ustrezajo tej zahtevi ležijo znotraj rdečega pasu



verjetnost znotraj modrega lika,
izrežemo Z in

$$d^2P = f_{X_1}(x_1) f_{X_2}(z-x_1) dx_1 dz$$

z integracijo po celotnem X_1
dobimo iskano verjetnostno gostoto
v intervalu $[z, z+dz]$

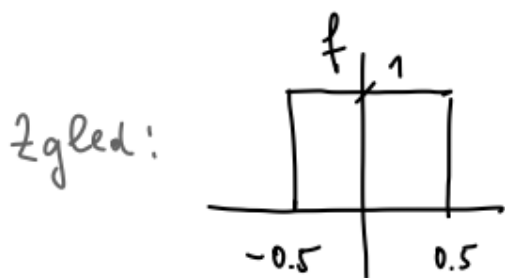
$$dP = f_Z(z) dz = \int_{-\infty}^{\infty} d^2P = dz \int_{-\infty}^{\infty} f_{X_1}(x_1) f_{X_2}(z-x_1) dx_1$$

To operacijo imenujemo konvolucija porazdelitev f_{X_1} in f_{X_2}
in jo označimo z simbolom $*$

$$f_Z(z) = (f_{X_1} * f_{X_2})(z) = \int_{-\infty}^{\infty} f_{X_1}(x_1) f_{X_2}(z-x_1) dx_1$$

Konvolucijo 3 porazdelitev izračunamo

$$f_z(z) = (f_{x_1} * f_{x_2} * f_{x_3})(z) = \iint f_{x_1}(x_1) f_{x_2}(x_2) f_{x_3}(z - x_1 - x_2) dx_1 dx_2$$



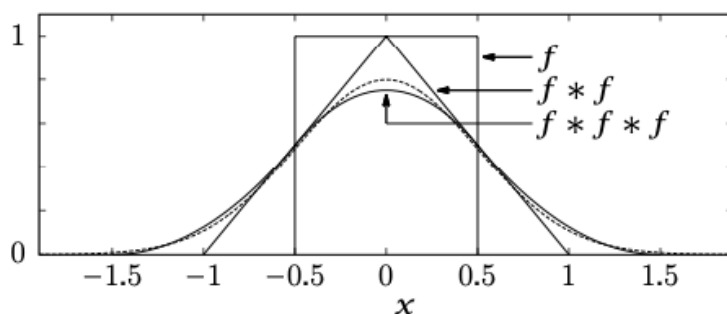
Določiti $(f * f)(z)$ ter $(f * f * f)(z)$
DN, str 126

$$(f * f)(z) = \int_{-1/2}^{1/2} f(x_1) f(z - x_1) dx_1 = \begin{cases} z \in [-1, 0]: \int_{-1/2}^{z+1/2} dx_1 = z + \frac{1}{2} + \frac{1}{2} = z + 1 \\ z \in [0, 1]: \int_{z-1/2}^{1/2} dx_1 = \frac{1}{2} - z + \frac{1}{2} = 1 - z \end{cases}$$

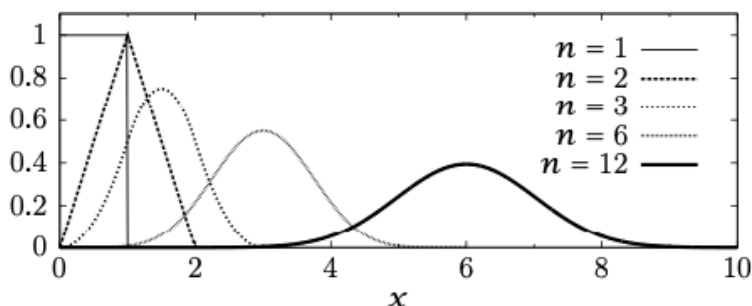
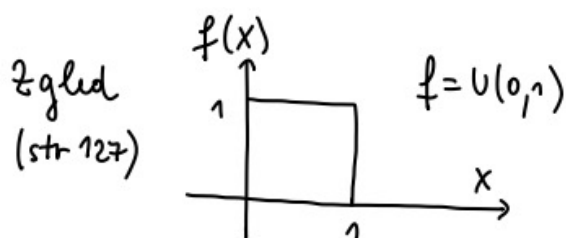
$$\begin{aligned} -\frac{1}{2} < z - x_1 < \frac{1}{2} \\ \downarrow \quad \downarrow \\ x_1 < z + \frac{1}{2} \quad z - \frac{1}{2} < x_1 \end{aligned}$$

možna pri $z < 0$ možna pri $z > 0$

ker $x_1 \in (-1/2, 1/2)$

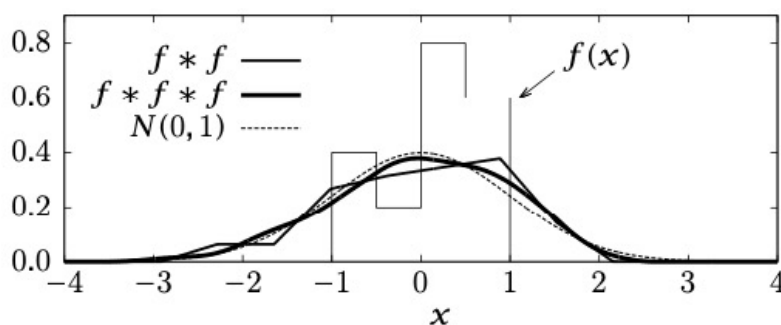


— Dvakratna konvolucija porazdelitve $U(-1/2, 1/2)$ same s seboj.



— Večkratne konvolucije porazdelitve $U(0, 1)$ same s seboj.

zglede (str 128)



6.3 CENTRALNI LIMITNI TEOREM (CLT)

tu sledim knjigi MaFi kušcer kodre : 301-303 (kk)
Slovarček označ je me koncu poglavje.

Iz preteklih zgledov smo lahko razbrali da je že konvolucija $f * f$ dveh razmeroma nepravilnih por. f nekega podobne normalni porazdelitvi, konvolucija večjega števila $f * f * \dots * f$ pa je še bolj podobna $N(\dots)$. Centralni limitni teorem govori o prid temu da to velja ne splošno, uvideli bomo pod kakšnimi pogoji se konvolucije mnogih porazdelitev poljubno približe normalni.

Naj bodo $X_1, \dots, X_N$ realne neodvisne naključne spremenljivke, vse porazdeljene po porazdelitvi f . Zanima nas porazdelitev naključne spremenljivke $Z = X_1 + \dots + X_N$ pri $N \rightarrow \infty$.

Vemo že: $\bar{Z} = \sum_{i=1}^N X_i = N \bar{X}_1 = N \bar{X}$, $\sigma_Z^2 = \sum_{i=1}^N \sigma_{X_i}^2 = N \sigma_{X_1}^2$, $\sigma_Z = \sqrt{N} \sigma_X$
 $\equiv \bar{Z}^{(N)}$ variance $X_i \equiv \sigma_Z^{(N)}$

vedli smo oznake za povprečje $\bar{z}$, standardno deviacijo σ (podobno) in porazdelitev z , ki poudarjajo da so vse odvisne od N . (kot KK)

Zanima nas $f_Z(z) = \underbrace{f * f * \dots * f}_N(z) \equiv \bar{f}_Z^{(N)}(z)$ za $N \rightarrow \infty$

Ob $N \rightarrow \infty$: $\bar{Z}$ in σ_Z divergirata, zato uvedemo premaknjeno in reskalirano naključno spr $S = (z - \bar{Z}^{(N)}) / \sigma_Z^{(N)}$.

Centralni limitni teorem pravi da je za $N \rightarrow \infty$ ob določenih pogojih S porazdeljen normalno $\rightarrow N(0,1)$, z pa $\rightarrow N(\bar{Z}^{(N)}, \sigma_Z^{(N)})$:

$$S = \frac{z - \bar{Z}^{(N)}}{\sigma_Z^{(N)}} : \lim_{N \rightarrow \infty} f_S(s) = N(0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}}$$

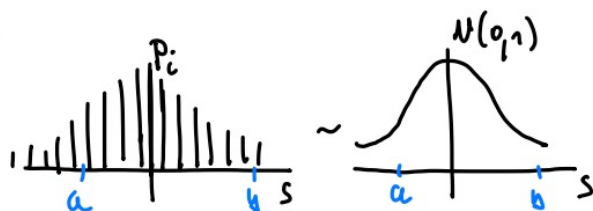
$$f_Z(z) \xrightarrow{N \rightarrow \infty} N(\bar{Z}^{(N)}, \sigma_Z^{(N)}) = \frac{1}{\sqrt{2\pi} \sigma_Z^{(N)}} e^{-\frac{(z - \bar{Z}^{(N)})^2}{2\sigma_Z^{(N)2}}}$$

$$Z : \lim_{N \rightarrow \infty} \sigma_Z^{(N)} \underbrace{f_Z^{(N)}(\sigma_Z^{(N)} S + \bar{Z}^{(N)})}_Z = \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}}$$

komentar glede pogojev veljavnosti: povprečje $\bar{x}$ in varianca σ_x^2 morete biti končne. Mi bomo v dokaz vgradili tudi, da je tretji moment končen $\overline{|x - \bar{x}|^3} < \infty$, to ni potreben pogoj. Spremenljivke morejo biti neodvisne in porazdeljene po identičnih porazdelitvah f. velja po analognem teorem tudi če so spremenljivke x_i porazdeljene po različnih f_{x_i} , a mi ga ne bomo obremenjevali.

CLT velja tudi za diskretne porazdelitve (konvoluciji n kock), a temu je $f_z(z) = f_1 \dots \ast f_1$ sestavljene iz serij delta funkcij, se pa približuj normalni porazdelitvi v smislu integrale po območjih ki vsebujejo več delta funkcij

$$\lim_{N \rightarrow \infty} \int_a^b \sigma_z^{(N)} f_z^{(N)}(\sigma_z^{(N)} s + \bar{z}^{(N)}) ds = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds$$

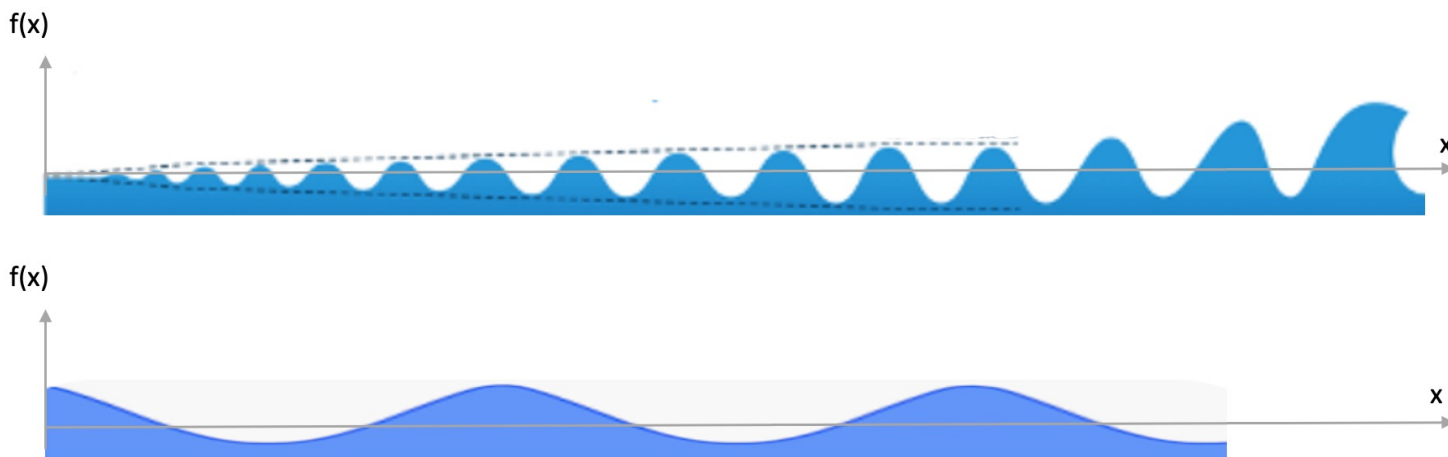


Sverilo: če velik n končen N meja CLT bolje v okolici $z \approx \bar{z}^{(N)}$ kjer je verjetnost največja, a v repih lahko precej slebo večja.

Fourierove transformacije in Fourierov integral

Pri dokazu Centralnega Limitnega Teorema (CLT) bomo potrebovali Fourierove transformacije in Fourierov integral. Ker boste to obravnavali pri matematiki sele v 2. semestru 2. letnika, se tej na splosno zelo pomembni temi posvetimo na tem mestu.

Zaradi boljše predstavljalivosti obravnavajmo $f(x)$, ki si jo predstavljamo kar kot obicajno funkcijo (ne verjetnostno gostoto). Naj $f(x)$ naj primer predstavlja odmik vala na vodi v odvisnosti od koordinate x v nekem trenutku, odmiki pa naj bodo odvisni le od x (in ne od y, z) – torej imamo v mislih eno dimenzionalni problem. Koordinata x naj ima enoto meter.



slike prikazujejo realni del $f(x)$

Fourierova analiza ponuja odgovor na vprašanje kateri valovni vektorji $k = 2\pi/\lambda$ so zastopani pri dani funkciji $f(x)$. Tu ima k enoto $[k] = 1/m$. Na spodnji sliki sklepamo da je zastopan le en valovni vektor, na zgornji sliki pa gotovo ni zastopan le en valovni vektor, temveč linearna kombinacija valovnih vektorjev. Kako močno je zastopan valovni vektor k naj označuje $\tilde{f}(k)$. Na splosno niso zastopane le diskretne vrednosti valovnega vektorja, temveč so lahko zastopane vrednosti porazdeljene zvezno. Spodaj $f(x)$ zapisemo kot linearno kombinacijo valovnih vektorjev k :

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(k) e^{ikx} dk$$

$\tilde{f}(k)$ pravimo Fourierova transformiranja funkcije $f(x)$ v prostor valovnih vektorjev k . Pricakujemo da bo $\tilde{f}(k)$ na spodnji sliki vseboval le en k , na zgornji pa vec k . Fourierova analiza nam pove, da se Fourierovo transformiranko $\tilde{f}(k)$ doloci na naslednji nacin

$$\tilde{f}(k) = \int_{-\infty}^{\infty} f(x) e^{-ikx} dx$$

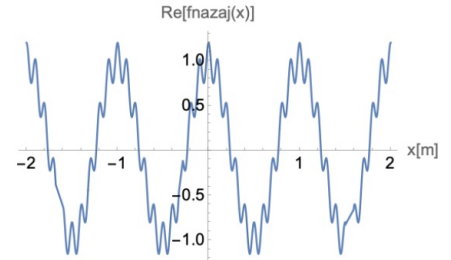
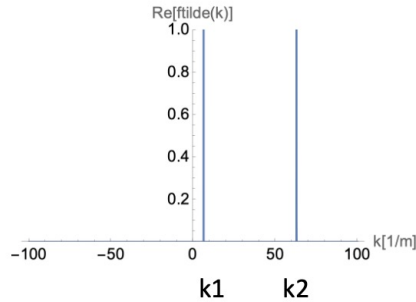
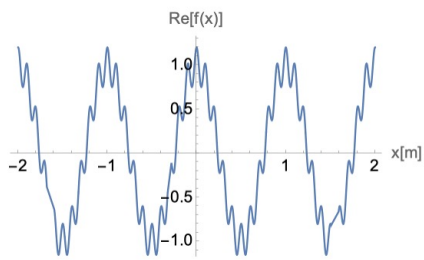
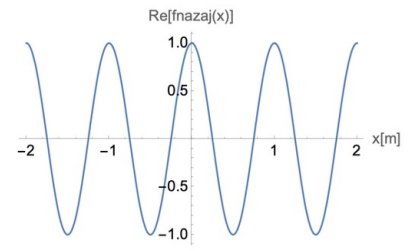
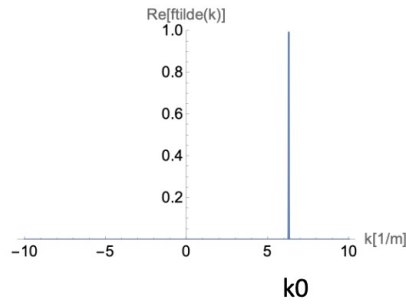
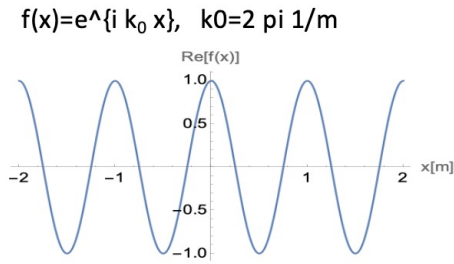
Spodaj si oglejmo najprej nekaj konkretnih primerov, potem pa jih z izpeljavami na podlagi zgornjih dveh zvez utemeljimo vsakega posebej.

$$f(x)$$

$$\tilde{f}(k) = \int_{-\infty}^{\infty} f(x) e^{-ikx} dx$$

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(k) e^{ikx} dk$$

$$f(x) = e^{i k_0 x}, \quad k_0 = 2 \pi / m$$



$$f(x) = A_1 e^{i k_1 x} + A_2 e^{i k_2 x},$$

$$A_1 = 1, k_1 = 2 \pi / m, A_2 = 0.2, k_2 = 10 * k_1$$

$$f(x)$$

$$\tilde{f}(k) = \int_{-\infty}^{\infty} f(x) e^{-ikx} dx$$

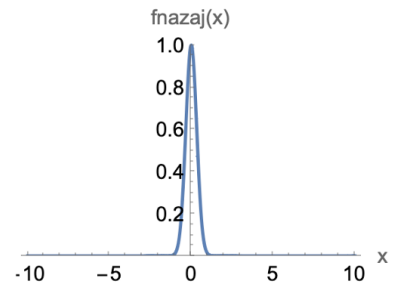
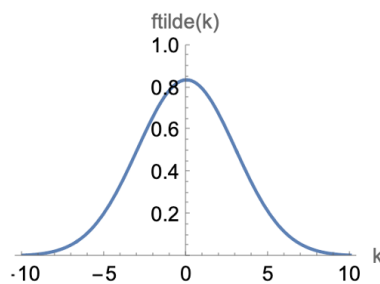
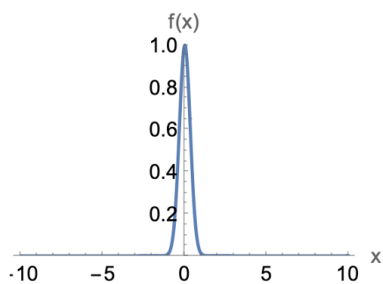
$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(k) e^{ikx} dk$$

$$f(x) = e^{-\frac{x^2}{2 s x^2}}$$

$$f_{\text{tilde}}(k) = e^{-\frac{1}{2} k^2 s x^2} \sqrt{2 \pi} s x$$

$$f_{\text{nazaj}}(x) = e^{-\frac{x^2}{2 s x^2}}$$

$$f(x) = N(0, s x = 1/3)$$

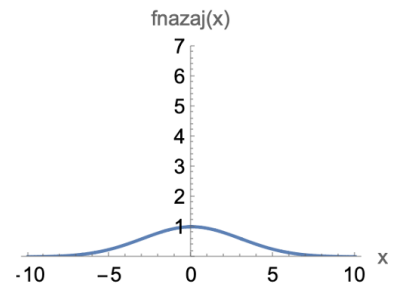
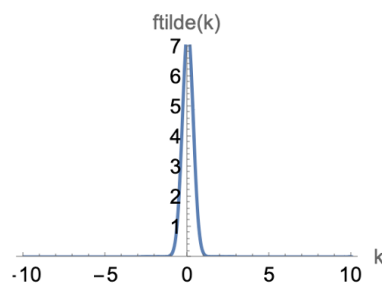
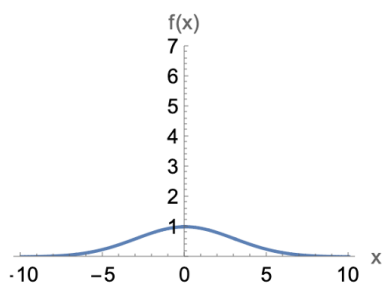


$$f(x) = e^{-\frac{x^2}{2 s x^2}}$$

$$f_{\text{tilde}}(k) = e^{-\frac{1}{2} k^2 s x^2} \sqrt{2 \pi} s x$$

$$f_{\text{nazaj}}(x) = e^{-\frac{x^2}{2 s x^2}}$$

$$f(x) = N(0, s x = 3)$$



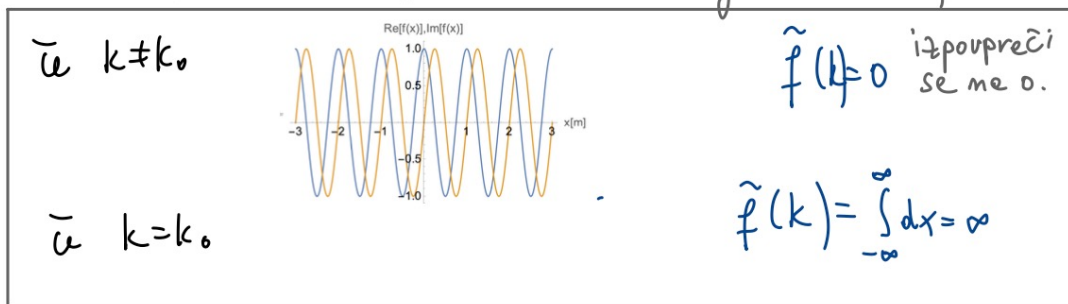
izpeljave k zgornji sliki

•) $f(x) = e^{ik_0 x}$ Pričakujemo da ta $f(x)$ vsebuje le velovni vektor $k=k_0$, to spodaj utemeljimo

$$\tilde{f}(k) = \int_{-\infty}^{\infty} f(x) e^{-ikx} dx = \int_{-\infty}^{\infty} e^{ik_0 x} e^{-ikx} dx = \int_{-\infty}^{\infty} e^{-i(k-k_0)x} dx$$

$$\int_{-\infty}^{\infty} \cos[(k-k_0)x] - i \sin[(k-k_0)x] dx = 2\pi \delta(k-k_0)$$

normalizirani nismo utemeljili



poenostavne zveze: $\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i(k-k_0)x} dx = \delta(k-k_0)$

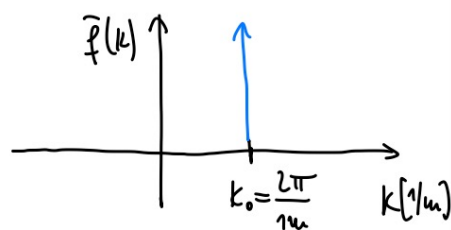
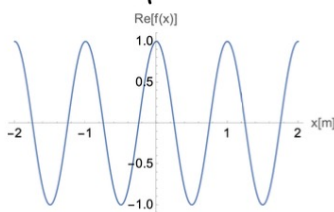
Torej lahko $f(x)$ pišemo tudi

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(k) e^{ikx} dk, \quad \tilde{f}(k) = \delta(k-k_0) 2\pi$$

Preverimo ali to drži

$$f(x) = \int_{-\infty}^{\infty} \delta(k-k_0) e^{ikx} dk = e^{ik_0 x} \Big|_{k=k_0} \text{ "fuzaj"}$$

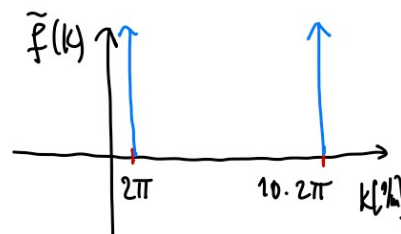
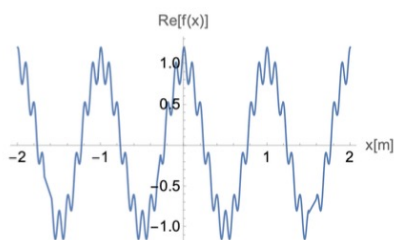
$$k_0 = 2\pi/\lambda_0, \quad \lambda_0 = 1 \mu$$



•) $f(x) = A_1 e^{ik_1 x} + A_2 e^{ik_2 x}$

$$\tilde{f}(k) = (A_1 \delta(k-k_1) + A_2 \delta(k-k_2)) \cdot 2\pi$$

$$A_1 = 1, \quad k_1 = \frac{2\pi}{\lambda_1} (\lambda_1 = 1 \mu), \quad A_2 = 0.2, \quad k_2 = \frac{2\pi}{0.1 \mu} = 10 \cdot 2\pi/\mu (\lambda_2 = 0.1 \mu)$$



a) $f(x) = e^{-\frac{x^2}{2\sigma_x^2}}$

$$\tilde{f}(k) = \int_{-\infty}^{\infty} f(x) e^{-ikx} dx = \int_{-\infty}^{\infty} e^{-\left(\frac{x^2}{2\sigma_x^2} - ikx\right)} dx = \int_{-\infty}^{\infty} e^{-\frac{x'^2}{2\sigma_x^2}} e^{-\frac{1}{2}k^2\sigma_x^2} dx'$$

$$\frac{1}{2\sigma_x^2} (x^2 - i \cdot 2k\sigma_x^2 x) = \frac{1}{2\sigma_x^2} \left[\underbrace{(x - ik\sigma_x^2)^2}_{x'} - k^2\sigma_x^4 \right]$$

$$\int_{-\infty}^{\infty} e^{-Ax^2} dx = \frac{\sqrt{\pi}}{\sqrt{A}} \quad \text{če } \operatorname{Re} A > 0$$

$$\tilde{f}(k) = \sqrt{2\pi} \sigma_x e^{-\frac{1}{2}k^2\sigma_x^2} = \frac{2\pi}{\sqrt{2\pi}\sigma_k} e^{-\frac{k^2}{2\sigma_k^2}} \quad \begin{array}{l} \text{PRI TEM SMO} \\ \text{DEFINIRALI} \\ \text{SILNO } \sigma_k \equiv \frac{1}{\sigma_x} \end{array}$$

$$\begin{aligned} f(x) &\stackrel{\text{DN}}{=} \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(k) e^{ikx} dk = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_k} e^{-k^2/2\sigma_k^2} e^{ikx} dk \\ &= \frac{1}{\sqrt{2\pi}\sigma_k} \sqrt{2\pi}\sigma_k e^{-\frac{1}{2}x^2\sigma_k^2} = e^{-\frac{x^2}{2\sigma_x^2}} \end{aligned}$$

b) Dokaz da sta $f(k)$ in $\tilde{f}(k)$ povezana preko

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(k) e^{ikx} dk, \quad \tilde{f}(k) = \int_{-\infty}^{\infty} f(x') e^{-ikx'} dx'$$

Vstavimo $\tilde{f}(k)$ v levo zvezo in preverimo ali integral res da $f(x)$

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x') e^{-ikx'} dx' \right] e^{ikx} dk = \frac{1}{2\pi} \int_{-\infty}^{\infty} dx' f(x') \left[\int_{-\infty}^{\infty} e^{ik(x-x')} dk \right]$$

$$= \int_{-\infty}^{\infty} dx' f(x') \delta(x-x') = f(x) \quad \checkmark$$

Integral z uporabo transformiramo $\tilde{f}(k)$ res da uresniči $f(x)$

ENA OD POMEMBNIH LASTNOSTI V ZVEZI S CONVOLUCIJO:
(BREZ DOKAZA)

$$\checkmark \text{ ce } f_z^{(N)}(z) = \underbrace{(f * \dots * f)}_N(z)$$

$$\downarrow$$
$$\boxed{\tilde{f}^{(N)}(k) = [\tilde{f}(k)]^N}$$

$$\tilde{f}^{(N)}(k) = \int f_z^{(N)}(z) e^{-ikz} dz$$

S Fourierovo transformacijo
preide konvolucijski integral
v produkt
(Kk, str 43)

$$\tilde{f}(k) = \int f(x) e^{-ikx} dx$$

Dokaz centralnega limitnega teorema

dve tipični poti

→ Fourierove transformacije: kújige KK, predevenja SS ←

→ Rodovne funkcije: kújiga SS

Izberemo Fourierove transformacije saj so ze fizike
me splošno mnogo bolj pomembne kot rodovne funkcij

zaradi enostavnosti privzemimo da je $\bar{X}=0 \rightarrow \bar{z}^{(N)}=0$

Poreadeltva bomo iz prostora X Fourierovo transformirali v
prostor k in konvolucijo določili preko zveze (kon) z6012A1.

$$\tilde{f}(k) = \int_{-\infty}^{\infty} f(x) e^{-ikx} dx$$

$$\tilde{f}^{(N)}(k) = [\tilde{f}(k)]^N, \text{ kjer } \tilde{f}^{(N)}(k) = \int_{-\infty}^{\infty} f_z^{(N)}(z) e^{-ikz} dz, f_z^{(N)}(z) = (f * \dots * f)(z)$$

e^{-ikx} razvijmo okoli $k=0$ saj pri uvedbi brezdimezijske
spremenljivke $x = \sigma_z^{(N)} k = \sqrt{N} \sigma_x k \rightarrow k = \frac{x}{\sqrt{N} \sigma_x}$ uvidimo da bode
členi z $N \rightarrow \infty$ višje potence k potisnjeni

$$\tilde{f}(k) = \int \left[1 - ikx \underset{1}{e^0} + \frac{1}{2} (-ikx)^2 \underset{1}{e^0} + \frac{1}{3!} (ikx)^3 \underset{1}{e^0} \right] f(x) dx$$

ostanek, načrt, kadičnik: I, str 145

$$\tilde{f}(k) = \underbrace{\int f(x) dx}_1 - i k \underbrace{\int f(x) x dx}_{\bar{X}=0} - \frac{1}{2} k^2 \underbrace{\int f(x) x^2 dx}_{\sigma_x^2} + R = 1 - \frac{1}{2} k^2 \sigma_x^2 + R$$

$$R \leq \frac{1}{6} |k|^3 \int |f(x)| |x|^3 dx = \frac{1}{6} |k|^3 \overline{|x|^3} \quad k = \frac{x}{\sqrt{N} \sigma_x}$$

$$\tilde{f}^{(N)}(k) = [\tilde{f}(k)]^N = \left[1 - \frac{1}{2} k^2 \sigma_x^2 + R \right]^N = \left[1 - \frac{1}{2} \frac{x^2}{N} + o\left(\frac{|x|^3 \overline{|x|^3}}{6 \sigma_x^3 N^{3/2}} \right) \right]^N$$

V dokaz vgradimo predpostavko da $\overline{|x|^3} < \infty$, ze $N \rightarrow \infty$ je
tedaj tretji člen z ostankom zanemareljiv proti drugemu

$$\lim_{N \rightarrow \infty} \tilde{f}^{(N)}(x) = \lim_{N \rightarrow \infty} \left[1 - \frac{1}{2} x^2 \frac{1}{N} \right]^N = e^{-\frac{1}{2} x^2}$$

Fourierovo transformiramo bresdimenzijski δ v bresdimen. S

$$\lim_{N \rightarrow \infty} f_s^{(N)}(s) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}^{(N)}(k) e^{iks} dk = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{s^2}{2}} e^{iks} dk = \frac{1}{2\pi} \sqrt{2\pi} e^{-\frac{s^2}{2}} = \frac{e^{-\frac{s^2}{2}}}{\sqrt{2\pi}}$$

N-KRATNA KONVOLUCIJA V OBČAJNEM PROSTORU PODAJA VERJETNOSTNO PORAZDELITEV NAKLJUČNIH SPREMEV. S OZIROMA $Z = \sum X_i$.

VIDIMO, DA JE V LIMITI $N \rightarrow \infty$ SPREMEVNIŠKAT S PORAZDELJENA

SCLOVAREK
OZNAK

TA PREDAVANJA

KUŠČER KODRČ (301-303)

$$f(x)$$

$$u_i(x) = u(x)$$

$$\tilde{f}(k) = FT f(x)$$

$$U(k) = FT u(x)$$

$$f_z^{(N)}(z)$$

$$w^{(N)}(z)$$

$$\tilde{f}^{(N)}(k) = FT f_z^{(N)}(z)$$

$$W^{(N)}(k) = FT w^{(N)}(z)$$

$$\sigma_x$$

$$\sigma_1$$

Zgled (Prirejeno po [6].) Masa M zrnč granulata farmacevtske učinkovine je naključna spremenljivka, porazdeljena v skladu z verjetnostno gostoto

$$f_M(m) = \frac{1}{24m_0^5} m^4 e^{-m/m_0}, \quad m \geq 0, \quad m_0 = 40 \text{ mg}. \quad (6.7)$$

Za analizo granulata vzamemo naključen vzorec 30 zrnč. Kolikšna je verjetnost, da skupna masa zajetih zrnč za več kot 10% presega svojo povprečno vrednost?

Povprečna masa zrnca in njena varianca sta

$$\bar{M} = \int_0^{\infty} m f_M(m) dm = 5m_0, \quad \sigma_M^2 = \int_0^{\infty} (m - \bar{M})^2 f_M(m) dm = 5m_0^2.$$

Verjetnostna gostota f_Z skupne mase vzorca Z , ki je prav tako naključna spremenljivka, je konvolucija tridesetih gostot oblike (6.7); to število je dovolj veliko za veljavnost centralnega limitnega izreka, zato je gostota f_Z skoraj normalna, s povprečjem $\underline{Z} = 30\bar{M} = 150 m_0$ in varianco $\sigma_Z^2 = 30\sigma_M^2 = 150 m_0^2$:

$$f_Z(z) = N(z, 150 m_0, \sigma_Z^2 = 150 m_0^2)$$

$$P(Z > 1.1 \times \bar{Z}) = P(Z > 165 m_0) = \int_{-\infty}^{165 m_0} N(z) dz = F(165 m_0) \doteq 11\%$$

Svarilni zgled

Zgled Konvolucija vendarle ne dela čudežev. Izračunajmo n -kratno konvolucijo Cauchyjeve porazdelitve same s sabo! Dobimo

$$\begin{aligned}f^{(1)}(x) &= f(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \\f^{(2)}(x) &= (f * f)(x) = \frac{1}{\pi} \frac{2}{4+x^2}, \\f^{(3)}(x) &= (f * f * f)(x) = \frac{1}{\pi} \frac{3}{9+x^2}, \\&\vdots \\f^{(n)}(x) &= \underbrace{(f * f * \dots * f)}_n(x) = \frac{1}{\pi} \frac{n}{n^2+x^2}.\end{aligned}\quad (6.3)$$

Ni videti, da bi se funkcija $f^{(n)}$ približevala normalni porazdelitvi, temveč ostaja zvesta svojemu rodu! Z zaporednimi konvolucijami dobivamo same Cauchyjeve porazdelitve! Pravimo, da je Cauchyjeva porazdelitev *stabilna glede na konvolucijo*. Vzroke za takšno obnašanje bomo spoznali v nadaljevanju. $\triangleleft$

Katera predpostavka centralnega limitnega teorema tu ni bila upravičena? Varianca te porazdelitve je neskončna! Končna varianca in povprečna vrednost izhodiscne $f(x)$ pa sta pogoj za veljavnost CLT. Izkaze se, da so tudi vsi višji momenti Cauchyjeve porazdelitve f neskončni.

CENTRALNI LIMITNI TEOREM ZA POVPREČJE

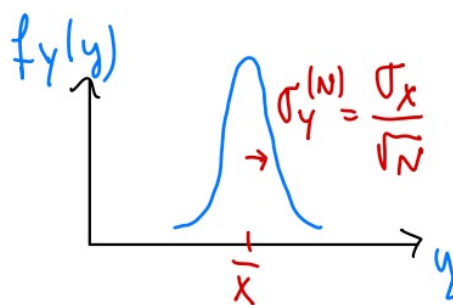
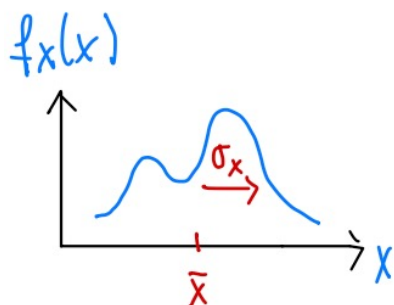
Zgorej smo obremenovali $f_Z(z)$ z vsoto $z = \sum_{i=1}^N X_i$.

Zdej me podlagi tege določimo $f_Y(y)$ z povprečjem $y = \frac{1}{N} \sum_{i=1}^N X_i$.

$$\bar{z}^{(N)} = N\bar{X} \rightarrow \bar{y}^{(N)} = \bar{X}$$

$$\sigma_z^{(N)} = \sqrt{N} \sigma_X \rightarrow \sigma_y^{(N)} = \frac{1}{\sqrt{N}} \sigma_X$$

$$f_Y(y) \xrightarrow{N \rightarrow \infty} N(\bar{y}^{(N)}, \sigma_y^{(N)}) = \frac{1}{\sqrt{2\pi} \sigma_y^{(N)}} e^{-\frac{(y - \bar{y}^{(N)})^2}{2\sigma_y^{(N)2}}}$$



porazdelitev enega izmerka

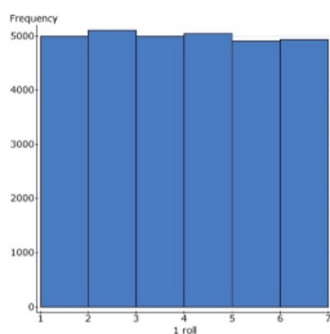
porazdelitev povprecja Y za N izmerkov

Mecemo N kock in za vsak met določimo povprečje števila pik za N kock:

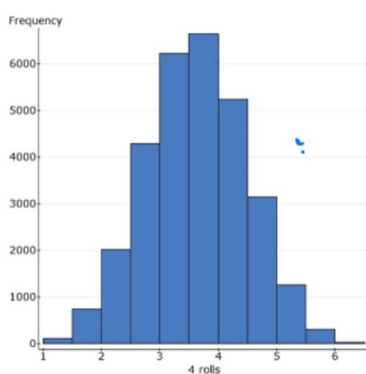
$$Y = \frac{1}{N} \sum_{i=1}^N X_i$$

x_i je število pik na eni kocki
N: številko kock

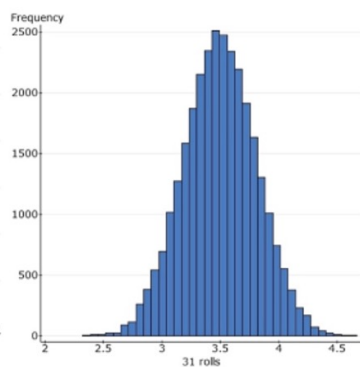
N=1



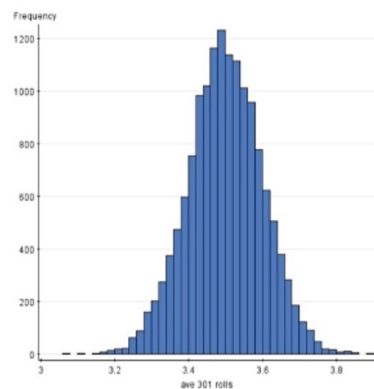
N=4



N=31



N=301



7. Statistično sklepanje na podlagi vzorcev

Poglavja 7 do 10 so namenjena osnovam statistike. Osnovna naloga statistike je empirično določanje verjetnostnih porazdelitev in njihovih parametrov.

V meso in kri nam morata najprej preiti pojma *populacije* in *vzorca*. Populacija (angl. *population*) je končna ali neskončna množica elementov, iz katere zajemamo *vzorke* (angl. *samples*). S statističnimi metodami poskušamo iz analize vzorcev sklepati na lastnosti celotne populacije, četudi so vzorci največkrat bistveno manjši kot populacija. Če količino, ki jo predstavlja naključna spremenljivka X , merimo (štejemo, ugotavljamo, beležimo) n -krat, dobimo nabor vrednosti $\{x_i\}_{i=1}^n$, obremenjenih z *negotovostjo* oziroma *napako*. Del te negotovosti je naključne (*statistične*) narave: vrednosti x_i so raztresene, ker v vzorec v splošnem vsakič zajamemo druge elemente. Ta del negotovosti zmanjšamo, če povečamo vzorec. Drugi del napake je *sistematičnega* izvora in je ne moremo odpraviti s povečevanjem vzorca.

Zgled Iz populacije $N = 2 \cdot 10^6$ Slovencev zajamemo vzorec $n = 1000$ ljudi in izmerimo njihovo telesno višino. Iz izmerjenih n vrednosti bi radi določili povprečno višino in njeno razpršenost ter podali neke vrste izjavo, kaj ti količini pomenita za celotno populacijo. Če izmerimo višino 1000 naključno izbranih ljudi danes in višino 1000 znova naključno izbranih ljudi jutri, bomo v splošnem dobili različni povprečni višin in različni raztresenosti (statistična negotovost). Če ves čas uporabljamo pokvarjeno napravo, ki kaže 1 cm premalo, bomo merili napačne višine ne glede na velikost vzorca in ne glede na to, ali vzorec zajemamo večkrat zapored (sistematska negotovost). $\triangleleft$

Populacija je lahko končna ali neskončna ($N = \infty$). Z mnogimi meti kovanca na primer pridemo do ocene za verjetnost, da pade cifra (dobili bomo približno $1/2$, glej razdelek 1.3), vendar celotno populacijo v tem primeru sestavlja množica vseh možnih metov kovanca, ki je neskončna.

Primer neskončne populacije je tudi $N=\infty$ metov kocke, da bi določili verjetnostno porazdelitev po številu pik za to kocko. Ali pa $N=\infty$ eksperimentov, da bi določili verjetnostno porazdelitev lege elektrona v vodikovem atomu v okviru kvantne mehanike.

Primer končne populacije je na primer populacija $N=2 \cdot 10^6$ Slovencev. Kasneje bomo napravili zgled za populacijo z $N=5$ elementi $x_{j=1,\dots,5}=\{2,3,6,8,11\}$

Povzetek: vrednosti naključne spremenljivke X

populacija: N elementov, $x_{j=1,\dots,N}$, zanima nas populacijsko povprečje μ
populacijska varianca σ^2
verjetnostna porazdelitev populacije $f_X(x)=?$

vzorec: n elementov, $x_{i=1,\dots,n}$, določimo lahko "le" vzorcno povprečje $\bar{x}$
vzorcno varianco s_X^2

V tem poglavju imamo v mislih **neodvisne naključne spremenljivke** X_i , kar bomo uporabili pri izpeljavah, razen kadar bomo eksplicitno poudarili, da spremenljivke niso neodvisne.

7.1 Statistike in cenilke

Naše izhodišče je naključni vzorec $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ iz populacije, ki jo karakterizira neznan parameter θ . Ta parameter bi radi ocenili iz vzorca. Funkcijo

cenilka za parameter
populacije theta

$$\hat{\theta} = T(\mathbf{X}) = T(X_1, X_2, \dots, X_n) \quad (7.1)$$

naključnih spremenljivk X_i z vrednostmi x_i , iz katerih dobimo oceno za θ , imenujemo *cenilka* (angl. *estimator*) parametra θ . Dogovorimo se, da uporabljamo isto oznako za cenilko kot predpis za spremenljivke, na primer $\hat{\theta} = (X_1 + X_2)/2$, kakor tudi za njeno konkretno vrednost ali oceno (angl. *estimate*), na primer $\hat{\theta} = (x_1 + x_2)/2$. Funkcijam, kot je (7.1), pravimo *vzorčne statistike* (angl. *sample statistic*), verjetnostnim porazdelitvam teh statistik pa *vzorčne porazdelitve* (angl. *sample distributions*).

Funkcijska oblika cenilke seveda ni poljubna: cenilko poskušamo zasnovati predvsem tako, da spostuje dve načeli: nepristranskost (unbiased) in doslednost (consistent), ki sta opredeljeni v nadaljevanju:

- Želimo, da **je cenilka nepristranska** (unbiased). To pomeni, da je za poljubno velikost vzorca n , ne le neskončno, pričakovana vrednost $\hat{\theta}$ enaka iskanemu parametru,

$$E[\hat{\theta}] = \theta, \quad \forall n. \quad (7.3)$$

Če je, nasprotno,

$$E[\hat{\theta}] = \theta + b(\hat{\theta}), \quad (7.4)$$

kjer je $b \neq 0$, pravimo, da je cenilka pristranska (angl. *biased*). Pri smiselnih cenilkah pričakujemo $b(\hat{\theta}) \ll \theta$ in, denimo, $b(\hat{\theta}) \sim 1/n$, ko gre $n \rightarrow \infty$.

$E[*]$ pomeni pričakovano vrednost pri povprečenju preko vseh vzorcev velikosti n iz populacije z N elementi: število vzorcev označimo z N_{vzorcev} (sample): to število običajno ni enako številu elementov v populaciji (N) ali številu elementov v vzorcu (n).

- Želimo tudi, da je **cenilka dosledna** (consistent). To za nepristranske cenilke v praksi pomeni, da ocena $\hat{\theta}$ z večanjem števila meritev n konvergira k pravi vrednosti θ in gre varianca cenilke za beskončno velike vzorce proti nič:

$$\lim_{n \rightarrow \infty} \text{var}[\hat{\theta}_n] = 0.$$

Tu $\text{var}[*]$ pomeni varianco izračunano preko vseh N_{vzorcev} vzorcev velikosti n iz populacije z N elementi.

7.1.1 Vzorčno povprečje in vzorčna varianca

Parametra, ki nas slej ko prej zanimata pri verjetnostni porazdelitvi sleherne naključne spremenljivke, sta povprečje in varianca. Vrednosti obeh teh količin za poljuben vzorec se v splošnem razlikujeta od njunih vrednosti za populacijo.

Zato najprej uvedemo oznake da bomo oboje razlikovali

$$\begin{aligned}\mu, \sigma^2 &\Leftrightarrow \text{populacija,} \\ \bar{X} \text{ (ali } \bar{x}), s_X^2 \text{ (ali } s_x^2) &\Leftrightarrow \text{vzorec,} \\ E[*] &: \text{pričakovana vrednost}\end{aligned}$$

Pozor: črta nad simbolom zdaj pomeni vzorčno povprečje, medtem ko smo jo v prejšnjih poglavjih uporabljali kot alternativno oznako za pričakovano vrednost.

Doslej so $\underline{X}$, μ in $E[X]$ označevali isto rec.

- Mislamo si, da iz populacije z N elementi zajamemo vzorec n vrednosti, in sicer tako, da se lahko vsaka vrednost ponovi večkrat. V primeru s telesnimi višinami ljudi to pomeni, da naključno izberemo osebo, ji izmerimo višino — to je vrednost spremenljivke X — in jo “vrnemo” v populacijo, od koder jo lahko po naključju znova “izberemo” v isti vzorec. Pravimo, da **vzorec nabiramo z nadomeščanjem oziroma vračanjem v populacijo**; tako je v danem vzorcu lahko večkrat ista oseba oziroma isti element. Tedaj lahko iz končne populacije z N elementi nabereмо tudi vzorce z n =neskončno elementi (saj lahko vzorec na primer nabereмо tudi neskončno enakih elementov). Prednost takega nabiranja vzorcev je, da so posamezni elementi vzorca lahko tedaj neodvisni od drugih (saj ni potrebno paziti da se isti element ne ponovi)
- Pri nabiranju vzorcev **brez vračanja v populacijo (brez nadomesčanja)** pa elementi niso neodvisni saj moramo paziti da drugi element ne more biti enak prvemu. Takega primera nabiranja ne bomo obravnavali.

Sicer pa pričakujemo, da za $N \gg n$ ni bistvene razlike ali elemente vzorca vracamo v populacijo ali ne

- Pokazali bomo da je **nepristranska in dosledna cenilka za populacijsko povprečje** (to velja tudi za spremenljivke ki niso neodvisne)

$$\hat{\theta} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

- Pokazali bomo, da je **nepristranska in dosledna cenilka za populacijsko varianco** za primer neodvisnih spremenljivk X_i (torej ta velja za vzorce nabrane z vračanjem v populacijo, ne pa za vzorce brez vračanja v populacijo)

$$s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

- Cenilka za katero bi naivno pričakovali da bo cenilka za populacijsko varianco pa ni nepristranska, zato slednja ni optimalna cenilka (je pa $\tilde{s}_X^2$ praktično enaka s_X^2 za velik n , zato je v tem primeru pogosto uporabljena)

$$\tilde{s}_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Preden dokažemo doslednost in nepristranskost cenilk za populacijsko povprečje in varianco, napravimo konkreten zgled za populacijo z $N=5$ elementi, kjer bomo lahko dejansko izvednotili pričakovane vrednosti in variance po vseh vzorcih, kar je potrebno za dokaz doslednosti in nepristranskosti.

Zgled (str 160, razširjen): populacija eksotičnega plazilca na nekem otoku je sestavljena iz **$N=5$** plazilcev, in ti imajo visine $x_i=\{2,3,6,8,11\}$ (merjetno v decimetrih).

- Doloci populacijsko povprečje in varianco za naključno spremenljivko visine
- Preuci primere za vzorce s stevili plazilcev **$n=2,3,4$** .

Obravnavaj primer ko pri nabiranju vzorca plazilca vsakic sproti plazilca vrnes v populacijo, tako je v danem vzorcu lahko večkrat isti plazilec, npr v vzorcu $[11,11]$ z $n=2$ je dvakrat največji plazilec. Tako nabrani vzorci imajo neodvisne elemente.

Doloci statistike teh cenilk in razisci ali so nepristranske in dosledne.

S tem preucis kaj manjsi vzorci povedo o populaciji in dobis informacijo o verjetnostni porazdelitvi visine v populaciji: verjetnost je enaka $1/5$ za vseh pet visin $x_i=\{2,3,6,8,11\}$

Zgled je obravnavan na prosojnicah.

V knjigi je obravnavan tudi primer ko pri nabiranju vzorcev plazilcev ne vračas v populacijo, torej vzorec $(11,11)$ ni zastopan. Elementi v tako nabranem vzorcu niso neodvisni in nekateri koraki pri izpeljavi cenilke za vzorcno varianco niso vec upraviceni, zato mora biti cenilka za vzorce brez vračanja drugačna kot za vzorce z vračanjem; cenilka za povprečje pa je ista za vzorce z vračanjem. (glej knjigo)

- **Cenilka za populacijsko povprečje**

$$\hat{\theta} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Pri dokazu za nepristranskost in doslednost $E[*]$ in $\text{var}[*]$ določimo po vseh N vzorcih.

Cenilka (7.5) je očitno nepristranska, saj je v skladu z zahtevo (7.3) njena pričakovana vrednost enaka populacijskemu povprečju,

$$E[\bar{X}] = \frac{1}{n} (E[X_1] + E[X_2] + \dots + E[X_n]) = \frac{1}{n} n\mu = \mu. \quad (7.6) \text{ populaciji,}$$

zato je $E[X_1]$ po vseh vzorcih enaka populacijski vrednosti μ , kar smo uporabili zgoraj.

Poleg tega je tudi dosledna, saj gre njena varianca proti nič, ko gre $n \rightarrow \infty$:

$$\text{var}[\bar{X}] = \text{var} \left[\frac{1}{n} (X_1 + X_2 + \dots + X_n) \right] = n \frac{1}{n^2} \text{var}[X] = \frac{\sigma^2}{n}. \quad (7.7)$$

To velja le za vzorce kjer so X_i neodvisni, kar je naša predpostavka za primer nabiranja vzorcev z vračanjem v populacijo.

- Določitev cenilk za populacijsko varianco

$$\tilde{s}_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

videli bomo da je to pristranska cenilka za populacijsko varianco

$$s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

končna (nepristranska) cenilka za populacijsko varianco

$\tilde{s}_X^2$ je gotovo nekakšna cenilka za σ^2 , toda vprašanje je, ali je tudi nepristranska. Če bi bila nepristranska, bi morala biti njena pričakovana vrednost enaka populacijski varianci,

$$E[\tilde{s}_X^2] \stackrel{?}{=} \sigma^2.$$

Ali to drži? Obravnavajmo le prvi člen vsote in ga malce preoblikujmo,

$$\begin{aligned} X_1 - \bar{X} &= X_1 - \frac{1}{n}(X_1 + X_2 + \dots + X_n) \\ &= \frac{1}{n}[(n-1)X_1 - X_2 - \dots - X_n] \\ &= \frac{1}{n}[(n-1)(X_1 - \mu) - (X_2 - \mu) - \dots - (X_n - \mu)], \end{aligned}$$

nato pa ga kvadrirajmo,

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

Spremenljivke X_i in X_j ($i \neq j$) so med seboj neodvisne, torej $E[(X_i - \mu)(X_j - \mu)] = 0$, zato mešani členi v pričakovani vrednosti odpadejo:

$$\begin{aligned} E[(X_1 - \bar{X})^2] &= \frac{1}{n^2} \{ (n-1)^2 E[(X_1 - \mu)^2] + E[(X_2 - \mu)^2] + \dots + E[(X_n - \mu)^2] \} \\ &= \frac{1}{n^2} \{ (n-1)^2 \sigma^2 + \underbrace{\sigma^2 + \dots + \sigma^2}_{(n-1)\sigma^2} \} = \frac{n-1}{n} \sigma^2. \end{aligned}$$

V vsoti (7.8) imamo n takšnih prispevkov, pred vsoto pa je še faktor $1/n$, torej

$$E[\tilde{s}_X^2] = \frac{n-1}{n} \sigma^2. \quad (7.9)$$

Potemtakem je $\tilde{s}_X^2$ pristranska cenilka za populacijsko varianco: nepristransko cenilko dobimo šele, če (7.8) na desni pomnožimo z $n/(n-1)$,

$$\text{končna (nepristranska) cenilka za populacijsko varianco} \quad s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (7.10)$$

Ta korak je upravičen le za vzorce kjer so X_i neodvisni, torej za vzorce nabrane z vračanjem v populacijo. Za vzorce, ki temu ne zadoscajo, moramo uporabiti drugačno cenilko.

- **cenilka za varianco vzorčnega povprečja** (estimator for variance of sample mean)

Želimo zapisati se cenilko za varianco vzorčnega povprečja. To je tista varianca, ki se z vecanjem vzorca manjša, in je njena prava vrednost (ko seštejemo po vseh vzorcih) enaka

želimo cenilko za to pravo vrednost
$$\text{var}[\hat{\theta}_n] = \text{var}[\bar{X}] = \frac{\sigma^2}{n}$$

A ker prave vrednosti populacijske variance σ^2 ne vemo, tvorimo cenilko za zgornjo količino tako, da σ^2 nadomestimo z njeno cenilko s_x^2 , torej cenilko za pop. varianco

**cenilka za varianco
vzorčnega povprečja**

$$s_{\bar{X}}^2 = \frac{s_X^2}{n} = \frac{1}{n(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2$$

od tod pa s korenjenjem še negotovost ali "napako" povprečja (error of the mean).

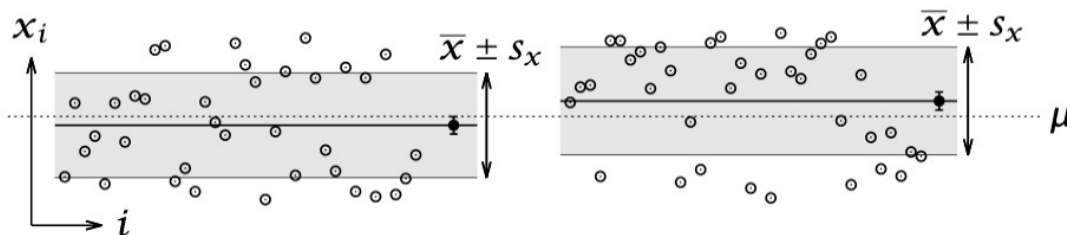
Iz tega rezultata sledi pomemben nauk, ki ga tako rekoč ne moremo dovolj poudariti. Vsak vzorec, ki ga izberemo iz populacije, ima v splošnem drugačno vzorčno povprečje in drugačno varianco. Raztresenost izmerkov bo vselej približno $\pm s_x$ (senčeni območji na sliki 7.1), vendar to ne pomeni, da bo tudi povprečje raztreseno za $\pm s_x$ — raztreseno bo le za $\pm s_x / \sqrt{n}$! Zato radi pišemo kar

$$\bar{x} = \mu \pm \frac{\sigma}{\sqrt{n}} \quad \text{ali} \quad \mu = \bar{x} \pm \frac{s_x}{\sqrt{n}}, \quad (7.11)$$

kar pomeni skoraj isto. Z izrazom na levi imamo v mislih, da je vzorčno povprečje približno enako pravemu, populacijskemu, njegova negotovost pa je $\pm \sigma / \sqrt{n}$. Z zapisom na desni izražamo, da je **$\bar{x}$ dober približek za μ , zmotili smo se nemara za $\pm s_x / \sqrt{n}$.** Ne glede na njihovo interpretacijo iz teh formul spoznamo tudi, kako velik vzorec potrebujemo za želeno natančnost: če na primer želimo določiti μ na 0.01σ natančno, potrebujemo $n = 10^4$ izmerkov.

V poglavju o intervalih zaupanja bomo zgornjo grobo formulacijo "zmotili smo se nemara za .." kvantitativno opredelili na z izjavo ki se nanasa na verjetnosti.

Primer vzorcev z različnima povprečjema $\bar{x}$ in skoraj enakima variancama s_x^2 prikazuje slika 7.1. Glej tudi nalogo 7.6.1.



Slika 7.1 — Vzorca velikosti $n = 36$ z različnima povprečjema in skoraj enakima efektivnima odklikoma $\bar{x} = -0.052$, $s_x = 0.313$ (levo) in $\bar{x} = 0.092$, $s_x = 0.322$ (desno), zajeta iz populacije s povprečjem $\mu = 0$ in efektivnim odklikom $\sigma = 0.288$. Polna krožca označujeta vzorčni povprečji $\bar{x}$ in njuni negotovosti $s_x / \sqrt{n}$.

- raztresenost (varianca) vzorčne variance

Vprašajmo se še, kolikšna je raztresenost vzorčne variance s_X^2 okrog populacijske variance σ^2 ob predpostavki, da je porazdelitev posameznih X_i normalna. Pri velikih n , kjer je razlika med formulama (7.8) in (7.10) zanemarljiva, je ta "varianca variance" enaka

$$E \left[(s_X^2 - \sigma^2)^2 \right] = E \left[\left(\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n} - \sigma^2 \right)^2 \right].$$

koj dobimo pri kvadriranju ()²

- $E((x_i - \bar{x})^4) = 3\sigma^4$ za normalno porazdelitev; n takih členov
- $E(2(x_i - \bar{x})^2(x_j - \bar{x})^2) = 2 \overbrace{(x_i - \bar{x})^2 (x_j - \bar{x})^2}^{i \neq j} \approx 2 \cdot \sigma^2 \cdot \sigma^2 = 2\sigma^4$, $\frac{n(n-1)}{2}$ členov
ker x_i in x_j neodvisne
- miseln člen $-2s_X^2\sigma^2 = -2\sigma^4$

$$E \left[(s_X^2 - \sigma^2)^2 \right] \approx \frac{1}{n^2} \left(3n\sigma^4 + 2 \frac{n(n-1)}{2} \sigma^4 - 2n^2\sigma^4 \right) + \sigma^4 = \frac{2\sigma^4}{n}.$$

Ugotovili smo

$$s_X^2 = \sigma^2 \left(1 \pm \sqrt{\frac{2}{n}} \right),$$

kar v primeru $n \gg 1$ pomeni

$$s_X = \sigma \left(1 \pm \frac{1}{\sqrt{2n}} \right) \quad \text{oziroma} \quad \sigma = s_X \left(1 \pm \frac{1}{\sqrt{2n}} \right).$$

Če želimo določiti σ na 1 % natančno, potrebujemo torej $n = 5000$ izmerkov.

Primerjava te približne ocene $2\sigma^4/n$ za varianco vzorčne variance z dejansko varianco vzorčne variance za zgled z $N=5$ plazilci je na prosojnicah. Dejanska varianca pada z n , ni pa povsem enaka $2\sigma^4/n$, saj tudi približki niso upravičeni: $n=2,3,4$ ni velik in porazdelitev X_i za populacijo ni normalna.

7.2 Primeri pomembnih vzorčnih porazdelitev

- 7.2.1 vzorčna porazdelitev vsot in razlik
- 7.2.2 vzorčna porazdelitev varianc $\tilde{s}_X^2$

7.2.1 Vzorčna porazdelitev vsot in razlik

Denimo, da imamo opravka z dvema neskončnima populacijama s povprečjema μ_1 in μ_2 ter variancama σ_1 in σ_2 . Iz prve populacije izberemo vzorec velikosti n_1 , iz druge pa neodvisno od prve še vzorec velikosti n_2 ter izračunamo vzorčni povprečji $\bar{X}_1$ in $\bar{X}_2$. Zanima nas porazdelitev vsote vzorčnih povprečij $\bar{X}_1 + \bar{X}_2$ za velika vzorca n_1 in n_2 .

za velik n_1 je porazdelitev povprečja $\bar{X}_1$ normalna: $f_{\bar{X}_1}(\bar{X}_1) = N(\mu_1, \sigma_1^2/n_1)$
za velik n_2 je porazdelitev povprečja $\bar{X}_2$ normalna: $f_{\bar{X}_2}(\bar{X}_2) = N(\mu_2, \sigma_2^2/n_2)$

Kaj vemo o vsoti ali razliki $\bar{X}_1 + \bar{X}_2$, kjer sta X_1 in X_2 dve neodvisni spremenljivki?

$$E[\bar{X}_1 \pm \bar{X}_2] = E[\bar{X}_1] \pm E[\bar{X}_2] = \mu_1 \pm \mu_2$$

$$\text{var}[\bar{X}_1 \pm \bar{X}_2] = \text{var}[\bar{X}_1] + \text{var}[\bar{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Poleg tega vemo, da je vsota $X_1 + X_2$ dveh neodvisnih spremenljivk podana s konvolucijo obeh porazdelitev. Konvolucija dveh normalnih porazdelitev pa je zopet normalna (tega ne bomo dokazali). Zato je vsota porazdeljena

$$f_{\bar{X}_1 + \bar{X}_2}(\bar{X}) = (f_{\bar{X}_1} * f_{\bar{X}_2})(\bar{X}) = N(\mu_1 + \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)$$

analogno pa velja za razliko $f_{\bar{X}_1 - \bar{X}_2}(\bar{X}) = N(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)$

Reskalirana naključna spremenljivka Z pa je porazdeljena po standardizirani normalni porazdelitvi $N(0,1)$.

$$Z = \frac{(\bar{X}_1 \pm \bar{X}_2) - (\mu_1 \pm \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Zgled Življenjska doba vezja tipa A je porazdeljena normalno s povprečjem $\mu_A = 7.0$ let in standardno deviacijo $\sigma_A = 1.1$ leta. Vezja tipa B imajo povprečno življenjsko dobo $\mu_B = 5.8$ leta s standardno deviacijo $\sigma_B = 0.9$ leta. Testiramo $n_A = 40$ vezij tipa A in $n_B = 40$ vezij tipa B. Kolikšna je verjetnost, da bodo vezja A delovala leto dlje kot vezja B?

Po formulah (7.13) in (7.14) dobimo

$$\mu_A - \mu_B = 1.2 \text{ leta}, \quad \sqrt{\text{var}[\bar{X}_A - \bar{X}_B]} = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = 0.225 \text{ leta}.$$

Zanima nas verjetnost, da je razlika pričakovanih življenjskih časov večja od enega leta, $\bar{X}_A - \bar{X}_B > 1$ leto. Ustrezna standardizirana spremenljivka (7.15) za mejno vrednost $\bar{X}_A - \bar{X}_B = 1$ leto je

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\text{var}[\bar{X}_A - \bar{X}_B]}} = \frac{1 - 1.2}{0.225} \approx -0.89$$

in je porazdeljena normalno. Iskana verjetnost je tedaj

$$P(\bar{X}_A - \bar{X}_B > 1 \text{ leto}) = P(Z > -0.89) = \frac{1}{2} + P(0 \leq Z \leq 0.89) \approx 81.3\%,$$

kjer smo si pomagali s tabelo D.1.

7.2.2 vzorčna porazdelitev varianc : porazdelitev χ^2

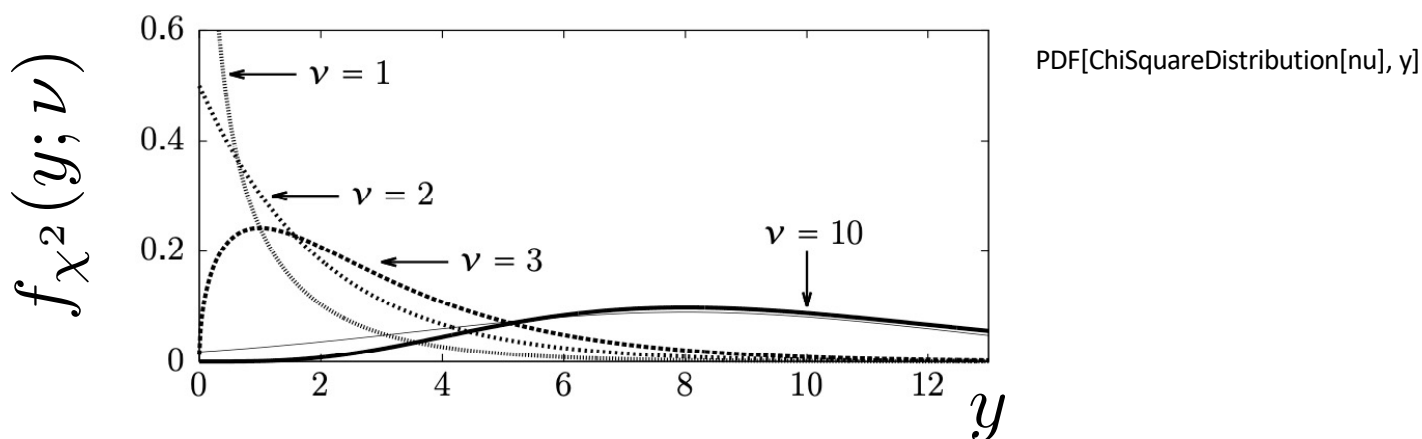
Vzorčne porazdelitve varianc dobimo, ko zajamemo iz populacije vse možne naključne vzorce velikosti n in izračunamo varianco za vsak vzorec. Iz variance populacije, σ^2 , in vzorčne variance $\tilde{s}_X^2$ (v pristranski obliki) tvorimo naključno spremenljivko

$$\chi^2 \stackrel{\text{def}}{=} \frac{n \tilde{s}_X^2}{\sigma^2} = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2}$$

χ^2 je naključna spremenljivka, ki zavzame različne vrednosti za različne vzorce. Imenujmo jo $Y = \chi^2$, torej naključna spremenljivka Y zavzame različne vrednosti y .

Velja izrek, ki ga bomo dokazali/nakazali: **Če naključne vzorce velikosti n zajemamo iz normalno porazdeljene populacije, je porazdelitev $f_Y(y)$ naključne spremenljivke $Y = \chi^2$ po vrednosti y podana s porazdelitvijo "hi—kvardrat" (chi square) z $\nu = n - 1$ prostostnimi stopnjami, torej**

$Y = \chi^2$	$f_Y(y) = f_{\chi^2}(y; \nu = n - 1)$	$\nu = \text{število prostostnih stopenj}$
	$f_{\chi^2}(y; \nu) \stackrel{\text{def}}{=} \frac{1}{2^{\nu/2}} \frac{1}{\Gamma(\nu/2)} y^{\nu/2-1} e^{-y/2}, \quad y > 0$	



Lastnosti pomembne porazdelitve χ^2 :

$$E[Y] = \int_0^{\infty} f_{\chi^2}(y; \nu) y dy = \nu$$

$$\text{var}[Y] = \int_0^{\infty} f_{\chi^2}(y; \nu) (y - \bar{y})^2 dy = 2\nu$$

$$\text{mod}[Y] = \max[f_{\chi^2}(y; \nu)] = \max(0, \nu - 2)$$

za velik ν : porazdelitev postane podobna normalni porazdelitvi

Preverimo ali razumemo njeno povprečje te porazdelitve: v ta namen določimo pricakovano vrednost χ^2 po vseh vzorcih

$$E[Y] = E\left[\frac{n\tilde{s}_X^2}{\sigma^2}\right] = \frac{n}{\sigma^2} E[\tilde{s}_X^2] = \frac{n}{\sigma^2} \frac{(n-1)\sigma^2}{n} = n-1 = \nu$$

Kjer smo uporabili pricakovano vrednost variance v pristranski obliki. Ugotovimo da je pricakovana vrednost Y res enaka pricakovani vrednosti χ^2 porazdelitve: to je $\nu=n-1$ za vzorec velikosti n .

Zgled: V veliki populaciji ($N=10\,000$) lisic je teža porazdeljena približno normalno z povprečjem $\mu=10$ kg in standardnim odklonom $\sigma=2$ kg. Iz nje zajemaj vzorce za različnimi števili elementov $n=3,8,100$ in razisci statistike oziroma porazdelitve vzorčnih varianc. Za $n=3$ in 8 obravnavaj $N_{vz}=100$ vzorcev, za $n=100$ pa $N_{vz}=1000$ vzorcev. Zgled je obravnavan na prosojnicah.

Opozorilo: Zgled iz knjige (str 162) zajema iz populacije $N=5$ eksotičnih zivali, kjer njihove visine $=\{2,3,6,8,11\}$ niso porazdeljene po normalni porazdelitvi. Zato χ^2 porazdelitev ni upravičena. Pri zgledu z lisicami pa je teža ⁺porazdeljena normalno in je χ^2 porazdelitev upravičena.

V grobem je raztresenost vsakega elementa v vzorcu velikostnega reda σ^2 , vsak člen v vsoti v grobem prispeva 1, zato je χ^2 velikostnega reda n oziroma bolj natančno $\nu=n-1$ (kar pa pokazuje sele natančen račun zgoraj)

Uporabne porazdelitve χ^2 in določitev relevantnega števila prostostnih stopenj

Porazdelitev χ^2 je pomembna iz različnih razlogov. Z njo smo/bomo:

- določiti porazdelitev vzorčnih varianc, kvadratov residuelo
kjer je residual $r_i = X_i - \mu, X_i - \bar{X}, X_i - g(\theta, t_i)$
residual = odstopanje pri prilagajanju

- interval zaupanja za populacijsko varianco σ^2
- pri prilagajanju funkcij in parametrov θ podatkom t_i, x_i
 $x_i = g(\theta, t_i)$: presoje ali so prilagojeni parametri
in predpostavljene verjetnostne porazdelitve skladne
s meritveni oziroma podatki (t_i, x_i) : t_i neodvisne, x_i odvisne
- V tovrstne merene se χ^2 definira v različnih različicah, tipično pa je vsota kvadratov residuelov r_i^2 .
- "Dokazati" ali v grobem mekerali bomo naslednje trditve

$$(1) \quad Y = \chi^2 \equiv \sum_{i=1}^n \frac{[X_i - \mu]^2}{\sigma^2} \rightarrow f_Y(y) = f_{\chi^2}(y; \nu = n)$$

$$(2) \quad Y = \chi^2 \equiv \sum_{i=1}^n \frac{[X_i - \bar{X}]^2}{\sigma^2} \rightarrow f_Y(y) = f_{\chi^2}(y; \nu = n - 1)$$

$$(3) \quad Y = \chi^2 \equiv \sum_{i=1}^n \frac{[X_i - g(t_i; \vec{\theta})]^2}{\sigma_i^2} \rightarrow f_Y(y) = f_{\chi^2}(y; \nu = n - p)$$

prilagajanje funkcije $x = g(t; \vec{\theta})$ s p parametri $\vec{\theta} = (\theta_1, \dots, \theta_p)$ na podatke $(t_i, x_i \pm \sigma_i)$, $i = 1, \dots, n$

1) Izpeljemo, da je $Y = \chi^2 = \sum_{i=1}^m \frac{(x_i - \mu)^2}{\sigma^2}$ porazdeljeva $f_Y(y) = f_{\chi^2}(y, \nu=m)$

uvvedemo novo naključno spr. $z_i = \frac{x_i - \mu}{\sigma}$ ki je normalna porazdeljena, nos pa zapišemo porazdelitev $Y = \sum z_i^2$ torej moremo le uvesti novo spremenljivko $Y = r^2 = \sum z_i^2$, ker so x_i neodvisni $\rightarrow z_i$ neodvisni $\rightarrow$

$$f_{\vec{z}}(z) dz_1 \dots dz_m = \frac{1}{\sqrt{2\pi}^m} e^{-\frac{1}{2}(z_1^2 + z_2^2 + \dots + z_m^2)} \underbrace{dz_1 \dots dz_m}_{dr r^{m-1} d\Omega_m}$$

preidemo na cilindrične koordinate v m-dimenzijah

$d\Omega_m$ je prostorski kot v m-dim: $d\Omega_3 = \sin\theta d\theta d\varphi$
 $d\Omega_2 = d\varphi$
 r je polmer sfere

Izberimo vse z $y = r^2$ in integriramo (marginaliziramo) po $m-1$ kotih, saj nas zanima le porazdelitev po $y = r^2$

$$f_Y(y) dy = \int \frac{1}{\sqrt{2\pi}^m} e^{-\frac{1}{2}y} \frac{1}{2\sqrt{y}} dy y^{\frac{m-1}{2}} d\Omega_m$$

$$r = y^{1/2} \quad dr = \frac{1}{2\sqrt{y}} dy$$

$$\int_{m-1 \text{ kotih}} d\Omega_m = \frac{2\pi^{m/2}}{\Gamma(m/2)}, \quad \int d\Omega_3 = \frac{2\pi^{3/2}}{\Gamma(3/2)} = \frac{2\pi\sqrt{\pi}}{\sqrt{\pi}/2} = 4\pi$$

↓
ker izpeljemo

$$f_Y(y) = \frac{1}{\sqrt{2\pi}^m} e^{-\frac{1}{2}y} y^{\frac{m}{2}-1} \frac{\pi^{m/2}}{\Gamma(m/2)} = f_{\chi^2}(y, \nu=m)$$

2)

zelimo
izpeljati

$$(2) \quad Y = \chi^2 \equiv \sum_{i=1}^n \frac{[X_i - \bar{X}]^2}{\sigma^2} \rightarrow f_Y(y) = f_{\chi^2}(y; \nu = n-1)$$

pokazimo na primeru: $\sigma = 1, n=3$

$$\bar{X} = \frac{1}{3}(X_1 + X_2 + X_3) \quad \text{vzorčno povprečje}$$

$$Y = \sum_{i=1}^3 [X_i - \bar{X}]^2 = [X_1 - \bar{X}]^2 + [X_2 - \bar{X}]^2 + [X_3 - \bar{X}]^2 = Z_1^2 + Z_2^2 = \sum_{i=1}^2 Z_i^2$$

↑
vedemo nove spremenljivke

originalen prostor X_i : 3D

prostor $X_i - \bar{X}$: 2D

od treh linearno neodvisnih vektorjev

ostaneta dva linearno neodvisna

kar smo eksplicitno uporabili v novi bazi Z_i

$$Z_1 = \frac{1}{\sqrt{2}}(X_1 - X_2)$$

$$Z_2 = \frac{1}{\sqrt{6}}(X_1 + X_2 - 2X_3)$$

$$Z_3 = \frac{1}{\sqrt{3}}(X_1 + X_2 + X_3)$$

$$f_Y(y) = f_{\chi^2}(y; \nu = 2 = n-1)$$

3)

$$(3) \quad Y = \chi^2 \equiv \sum_{i=1}^n \frac{[X_i - g(t_i; \vec{\theta})]^2}{\sigma_i^2} \rightarrow f_Y(y) = f_{\chi^2}(y; \nu = n-p)$$

nakazimo na primeru: $\sigma_i = 1, n=3, p=2$:

$$x = g(t; \theta_1, \theta_2) = \theta_1 + \theta_2 t^2$$

meritve: $(t_1, t_2, t_3) = (1, 2, 3)$
 (x_1, x_2, x_3)

odvisnost od theta je tu linearna,
odvisnost od t je poljubna

minimizacija χ^2 pri danih
podatkih da vedno parametre,
ki so linearno odvisni od x_i

$$\left\{ \text{th1} \rightarrow \frac{6 \times x_1}{7} + \frac{3 \times x_2}{7} - \frac{2 \times x_3}{7}, \text{th2} \rightarrow -\frac{11 \times x_1}{98} - \frac{x_2}{49} + \frac{13 \times x_3}{98} \right\}$$

$$Y = \chi^2 = \sum_{i=1}^{n=3} [X_i - \theta_1 - \theta_2 t_i^2]^2 = \dots = Z_1^2 = \sum_{i=1}^{\nu=n-p=1} Z_i^2$$

originalen prostor X_i : 3 dimenzionalen

prostor residualov r_i : 3-2 dimenzionalen (pravokoten na dva "vektorja" th1 in th2)

7.3 Intervali zaupanja (confidence intervals)

$$\mu = \bar{x} \pm \frac{s_x}{\sqrt{n}}$$

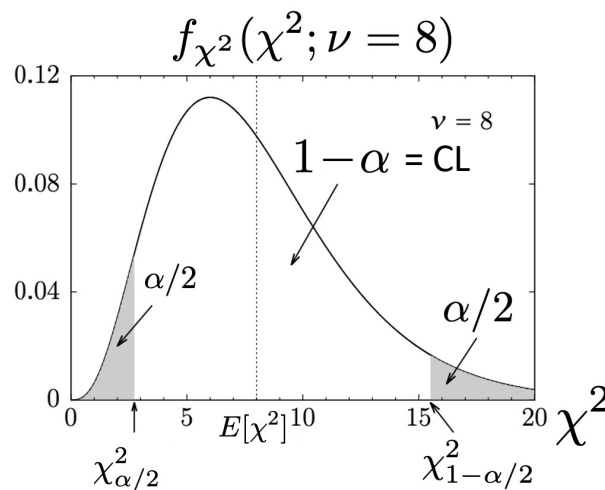
Ob (7.11) smo zapisali: "Zmotili smo se nemara za $\pm s_x/\sqrt{n}$." Kaj natančno to pomeni? Polna krožca na sliki 7.1, ki označujeta vzorčni povprečji, sta razmaknjena za več, kot znaša njuna negotovost, torej za več kot $\pm s_x/\sqrt{n}$! Očitno potrebujemo kvantitativno merilo za to, "koliko se utegnemo zmotiti". Najdemo ga v pripomočku, ki mu pravimo interval zaupanja (angl. *confidence interval*).

$$\sigma = s_x \left(1 \pm \frac{1}{\sqrt{2n}}\right)$$

Zelimo določiti kvantitativno merilo za koliko smo se utegnili zmotiti pri določitvi populacijske variance na podlagi danega vzorca.

7.3.2 Interval zaupanja za populacijsko varianco

Začnimo s populacijsko varianco, saj je ta neposredno povezana z porazdelitvijo χ^2 , ki smo jo pravkar spoznali.



Zanima nas interval, na katerem bo vzorčni χ^2 zastopan z verjetnostjo $1-\alpha$ če meritev vzorca velikokrat ponovimo. Pri tem izvzamemo delež $\frac{\alpha}{2}$ najmanjših χ^2 in delež $\frac{\alpha}{2}$ največjih χ^2 .

Verjetnosti **1- α pravimo stopnja zaupanja (confidence level = CL)**

Iz podrazdelka 7.2.2 že vemo, da ima spremenljivka $\chi^2 = n\tilde{s}_X^2/\sigma^2$ po- razdelitev χ^2 z $\nu = n - 1$ prostostnimi stopnjami

$$\chi_{\alpha/2}^2 \leq \frac{n\tilde{s}_X^2}{\sigma^2} \leq \chi_{1-\alpha/2}^2$$

Od tod dobimo **interval zaupanja, na katerem se populacijska varianca nahaja z vnaprej predpisano verjetnostjo 1- α** (ki ji pravimo stopnja zaupanja oz **confidence level = CL**)

$$\frac{\sqrt{n} \tilde{s}_X}{\sqrt{\chi_{1-\alpha/2}^2}} \leq \sigma \leq \frac{\sqrt{n} \tilde{s}_X}{\sqrt{\chi_{\alpha/2}^2}}$$

interval zaupanja za populacijski standardni odklon sigma pri stopnji zaupanja $1 - \alpha = CL$

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

Zgled Vzemimo spet zgled z vzorcem $n=8$ lisic in določimo interval zaupanja za varianco pri stopnji zaupanja $1-\alpha=0.8$. Pri prejsnjem zgledu smo že določili da bodo vzorčne variance $\tilde{s}_X^2$ oz. pripadajoč χ^2 na intervalu

$$\chi_{0.1}^2 \leq \frac{n\tilde{s}_X^2}{\sigma^2} \leq \chi_{0.9}^2$$

interval zaupanja za vzorčno varianco pri stopnji zaupanja (confidence level) $1-\alpha=0.8$

Nas pa iz izračunanega

$\tilde{s}_X^2$ za dan vzorec bolj zanima kakšen je interval zaupanja za populacijsko varianco σ^2

$$\frac{\sqrt{n \tilde{s}_X^2}}{\sqrt{\chi_{0.9}^2}} \leq \sigma \leq \frac{\sqrt{n \tilde{s}_X^2}}{\sqrt{\chi_{0.1}^2}}$$

interval zaupanja za populacijsko varianco σ^2 pri stopnji zaupanja (confidence level) $1-\alpha=0.8$

Primer za vzorce z $n=8$ lisicami je na prosojnicah. Za vsak od 100 vzorcev določimo interval zaupanja za sigma, in za vsak vzorec potem preverimo ali je pravi populacijski sigma res na tem intervalu. Izkaze se, da to drži v približno deležu $1-\alpha$, kot bi pričakovali.

Zgled: Pričakujemo da pri večjih vzorcih dobimo manjše območje intervala zaupanja, torej da je populacijska varianca bolj natančno določena. To preverimo za vzorce $n=100$ lisic na prosojnicah.

Zgled: 7.6.4 Doziranje aktivne učinkovine

+

V tovarni zdravil v tablete dozirajo aktivno farmacevtsko učinkovino, vendar je njena masa v tabletah raztresena okrog znane povprečne vrednosti: v vzorcu $n = 20$ tablet ugotovimo varianco $\tilde{s}_X^2 = 0.12 \text{ mg}^2$ ($\tilde{s}_X = 0.346 \text{ mg}$). Določi 80% interval zaupanja za populacijsko standardno deviacijo mase učinkovine (10% verjetnost, da je premajhna in 10% verjetnost, da je prevelika)!

 Interval zaupanja za populacijsko varianco določimo po formuli (7.20). Iz tabele D.3 pri $\nu = n - 1 = 19$ razberemo kritični vrednosti porazdelitve χ^2 , ki sta $\chi_{0.10}^2 = 11.7$ in $\chi_{0.90}^2 = 27.2$, torej lahko σ omejimo z

$$\sqrt{n\tilde{s}_X^2/\chi_{0.90}^2} \leq \sigma \leq \sqrt{n\tilde{s}_X^2/\chi_{0.10}^2}$$

oziroma $0.297 \text{ mg} \leq \sigma \leq 0.453 \text{ mg}$. Bodi pozoren na to, da s_X ne leži na sredini tega intervala, kot vemo že iz podrazdelka 7.3.2 (slika 7.2).

Z verjetnostjo 0.8 je populacijski standardni odklon sigma za učinkovino na intervalu, s čimer imamo v mislih: pri mnogokratni meritvi vzorcev bo v 80% vzorcev dejanska populacijska varianca znotraj vzorčnega intervala zaupanja. Pri tem se seveda populacijska varianca od vzorca do vzorca ne spreminja, interval zaupanja pa se od vzorca do vzorca spreminja.

7.3.1. Interval zaupanja za populacijsko povprečje

Naj bosta μ_T in σ_T^2 povprečje in varianca vzorčne porazdelitve neke statistike T , na primer $T = X$ ali $T = \sum_i X_i$. Če je vzorčna statistika približno normalna — kar velja za mnoge statistike, če je velikost vzorca vsaj nekaj 10 — pričakujemo, da bo vrednost T ležala na intervalu $[\mu_T - \sigma_T, \mu_T + \sigma_T]$ v približno 68.3 % primerov, na $[\mu_T - 2\sigma_T, \mu_T + 2\sigma_T]$ v približno 95.5 % primerov, na $[\mu_T - 3\sigma_T, \mu_T + 3\sigma_T]$ v približno 99.7 % primerov in tako dalje (glej (3.13)). Zato rečemo: s stopnjo zaupanja (angl. *confidence level*, CL) 68.3 % *zaupamo* (verjamemo, predvidevamo), da bomo T našli na intervalu $[\mu_T - \sigma_T, \mu_T + \sigma_T]$ in analogno za ostale. Takšnemu intervalu pravimo interval zaupanja (angl. *confidence interval*).

V tem odstavku govorimo o vzorcnem povprecju, v naslednjem pa zelimo nasloviti populacijsko povprecje.

Vzemimo vzorec $\{x_i\}_{i=1}^n$ oziroma n neodvisnih izmerkov, ki smo jim že določili vzorčno povprečje $\bar{x}$ in varianco s_x^2 v nepristranski obliki (7.10). Da bi ugotovili, kako dobro $\bar{x}$ ocenjuje pravo populacijsko povprečje μ , najprej tvorimo statistiko

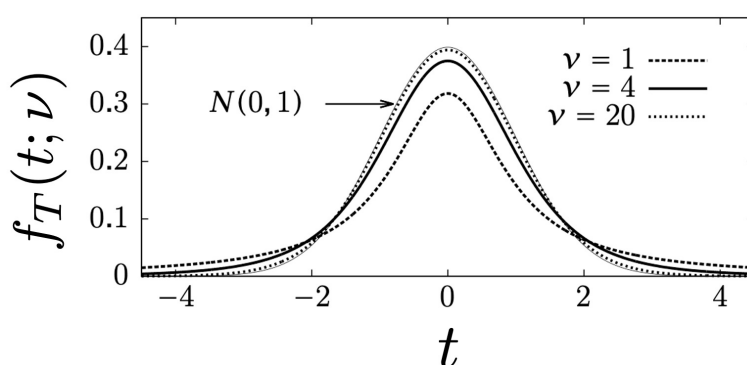
naključna spremenljivka: T
njena vrednost: t

$$T = \frac{\bar{X} - \mu}{s_X} \sqrt{n} \quad : \quad f_T(t; \nu = n-1) \quad (7.18)$$

Če so X_i porazdeljeni normalno po $N(\mu, \sigma^2)$, je statistika T porazdeljena po Studentovi porazdelitvi (3.22) z $\nu = n - 1$ prostostnimi stopnjami. $f_T(t; \nu = n-1)$

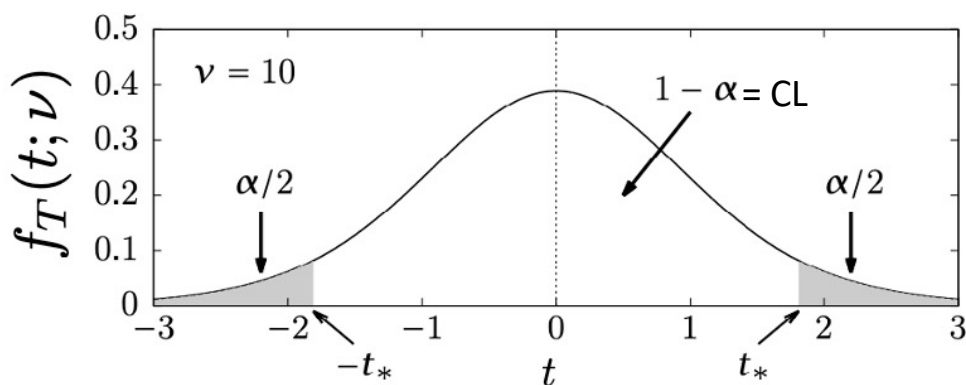
Studentovo porazdelitev navedemo zopet brez dokaza, za velik n je zelo podobna $N(0,1)$:

$$f_T(t; \nu) = \frac{1}{\sqrt{\nu} B\left(\frac{\nu}{2}, \frac{1}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$



Integral ver-

jetnostne gostote $f_T(t; \nu)$ določa meji intervala $[-t_*, t_*]$, na katerem z neko vnaprej izbrano verjetnostjo $1 - \alpha$ (stopnjo zaupanja) pričakujemo vrednosti t oziroma ustrezno povprečje μ , medtem ko z verjetnostjo (stopnjo tveganja) α pričakujemo, da bo vrednost t izven tega intervala: glej sliko 7.4.



Slika 7.4 — Ponazoritev zveze med stopnjo zaupanja $1 - \alpha$ in mejnima vrednostma $\pm t_*$ za določitev intervala zaupanja $[-t_*, t_*]$ za vzorčno povprečje. (Zgled z $\nu = 10$.)

Tedaj velja

$$-t_* \leq \frac{\bar{x} - \mu}{s_x} \sqrt{n} \leq t_*$$

oziroma

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

interval zaupanja za
populacijsko povprečje μ
pri dani stopnji zaupanja $1 - \alpha$

kar preberemo kot: "Pravo povprečje velike populacije, iz katere smo vzeli vzorec $\{x_i\}_{i=1}^n$, ocenimo z $\mu = \bar{x}$, povprečje μ pa je z verjetnostjo $1 - \alpha$ na intervalu

zaupanja (7.19).” Za velike vzorce ($n \geq 30$) je Studentova porazdelitev praktično enaka standardizirani normalni in ustrezne meje t_* so kar meje v Gaussovi krivulji (tabela 7.1). Z večanjem stopnje zaupanja se t_* povečuje, kar je razumljivo: širši interval pomeni manj “tveganja”.

Zgled Pri enajstkratni ($n = 11$) meritvi mase nekega delca smo dobili povprečno vrednost $\bar{m} = 4.180 \text{ GeV}/c^2$ in nepristransko oceno za standardno deviacijo $s_m = 0.060 \text{ GeV}/c^2$. Določimo interval zaupanja, na katerem s stopnjo zaupanja $1 - \alpha = 0.90$ pričakujemo dejansko maso delca μ .

Če so izmerki porazdeljeni normalno, je spremenljivka $T = (\bar{M} - \mu)\sqrt{n}/s_M$ porazdeljena po Studentovi porazdelitvi z $\nu = n - 1 = 10$ prostostnimi stopnjami. Mejna vrednost t_* je tista pod katero se t realizira z 0.95 odstotno verjetnostjo (5% spodaj, 5% zgoraj), torej

$$t_* = F^{-1}(0.95, \nu=10) = \text{InverseCDF}[\text{StudentTDistribution}[10], 0.95] = 1.812.$$

Iskani interval zaupanja za populacijsko povprečje μ je torej

$$\left[\bar{m} - \frac{t_* s_m}{\sqrt{n}}, \bar{m} + \frac{t_* s_m}{\sqrt{n}} \right] = [4.147, 4.213] \text{ GeV}/c^2.$$

Z verjetnostjo $CL=0.9$ je populacijsko povprečje mase delca na zgornjem intervalu s cimer imamo v mislih: pri mnogokratni meritvi vzorcev bo v 90% vzorcev dejanska masa delca znotraj vzorcnega intervala zaupanja. Pri tem je seveda dejanska masa od vzorca do vzorca nespremenjena, interval zaupanja pa se od vzorca do vzorca spreminja.

8. METODA MAKSIMALNEGA VERJETJA (Maximum likelihood method)

Tudi v tem poglavju bomo iz podatkov o vzorcih želeli sklepeti na lastnosti populacije. Porezdelitve vzorcev bodo odvisne od parametrov $\vec{\theta} = (\theta_1, \dots, \theta_p)$. Zasnovani bomo cenilke za parameter $\vec{\theta}$, obsevnaveki povprečij in varianc te cenilke po vzorcih.

8.1 Funkcija verjetja (likelihood function L) in njen logaritem (log-likelihood function ℓ)

Vzorec $(x_1, \dots, x_m)$ ima m elementov in vsak element je v populaciji porezdeljen z verjetnostno gostoto $f_{X_i}(x_i; \vec{\theta})$ kjer lahko verjetnostne gostote zavisijo od elementa i . Obenem bomo primer ko so mekljune spremenljivke X_i neodvisne. $\vec{\theta} = (\theta_1, \dots, \theta_p)$ so parametri porezdelitve f_{X_i} verjetnost da smo ob danih parametrih $\vec{\theta}$ izmerili prav $\vec{x} = (x_1, \dots, x_m)$ na intervalu $d\vec{x} = dx_1 \dots dx_m$ je že neodvisne spr. X_i luke

$$dP[\vec{x}, \vec{x} + d\vec{x}] = \underbrace{\prod_{i=1}^m f_{X_i}(x_i; \vec{\theta})}_{=}$$

funkcija
verjetja
(likelihood
function)

$$L(\vec{x} | \vec{\theta}) \stackrel{\text{def}}{=} L(x_1, \dots, x_m | \vec{\theta}) = \prod_{i=1}^m f_{X_i}(x_i; \vec{\theta})$$

$L(\vec{x}|\vec{\theta})$ je verjetnostne gostote za meritev vzorce da pogoju da je preva vrednost parametrov enake $\vec{\theta}$, na pogojno verjetnost pa nas opozarja znak " $|$ ", torej vzorec $\vec{x}$ ob pogoju $\vec{\theta}$ (for given)
 Prikladno je definirati še merevni logaritma L

logaritma funkcij verjetja
 (log-likelihood function)

$$l(\vec{x}|\vec{\theta}) \equiv \log L(\vec{x}|\vec{\theta}) = \sum_i \log f_{X_i}(x_i; \vec{\theta})$$

saj bomo v nadaljevanju odviseli po θ ker je bolj prikladno za l kot L .

8.2 Načelo maksimalnega verjetja

$\vec{\theta}$: parametri verjetnostne por.

$\hat{\vec{\theta}}$: cenilke za parametre $\vec{\theta}$ po metodi maksimalnega verjetja ($\hat{\vec{\theta}} = \text{maximum likelihood estimator} = \text{MLE}$)
 vrednost cenilke je odvisna od vzorca, torej je naključna spremenljivka ki jo določimo za vsak vzorec posebej

$\vec{\theta}_0 = E[\hat{\vec{\theta}}]$: iskane prave vrednosti parametrov v populaciji; tu je $E[...]$ pričakovane vrednosti po vseh vzorcih

Najčelo maksimalne verjetje pravi, da optimalno vrednost parametrov $\vec{\theta}$ najdemo z maksimizacijo funkcije verjetje L oziroma njenega logaritma l . Torej je realizirane meritve vzorce maksimalno verjetne. To je v skladu s frekventističnim pristopom kjer nos pri danih vrednostih parametrov (oziroma dani hipotezi) zeirne verjetnost da smo izmerili $\vec{x}$.

Opomba: V Bayes pristopu bi obravnavali kako verjeten je parameter $\vec{\theta}$ pri danem izmerjenem vzorcu in bi maksimizirali $P(\vec{\theta}|\vec{x})$: $P(\vec{\theta}|\vec{x}) = \frac{P(\vec{x}|\vec{\theta})P(\vec{\theta})}{P(\vec{x})}$, pozneti bi moreli $P(\vec{\theta})$.

Pri frekventističnem pristopu išemo $\vec{\theta}$ ki maksimizira verjetnost

za en parameter θ : $\frac{d l(\vec{x}|\theta)}{d\theta} \bigg|_{\theta=\hat{\theta}} = 0$

dobljenih podatkov $\vec{x}$: $l(\vec{x}|\hat{\theta}) \big|_{\hat{\theta}=\hat{\theta}} = \max_{\hat{\theta}=\hat{\theta}}$

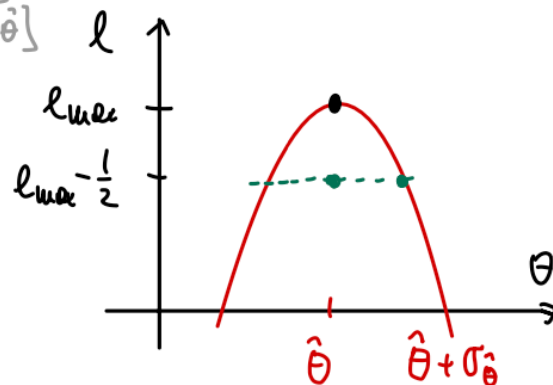
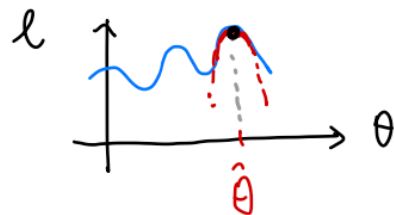
$$l'' = \frac{d^2 l}{d\theta^2} \bigg|_{\theta=\hat{\theta}} = -A = -\frac{1}{\text{var}[\hat{\theta}]} = -\frac{1}{\sigma_{\hat{\theta}}^2} < 0$$

$\theta=\hat{\theta}$ $\downarrow$ $\begin{matrix} \text{veke} \\ \text{pozitivne} \\ \text{kost} \end{matrix}$ $\downarrow$ $\begin{matrix} \text{poketeli bomo v} \\ \text{tem poglavju de } l'' = -\frac{1}{\text{var}[\hat{\theta}]} \\ \text{(poglavju 8.3.2)} \end{matrix}$

$$l = l_{\max} - \frac{1}{2} \frac{(\theta - \hat{\theta})^2}{\sigma_{\hat{\theta}}^2} + o((\theta - \hat{\theta})^3)$$

$l(\vec{x}|\hat{\theta})$ v svojem drugem členu vsebuje info o optimalni vrednosti cenilke $\hat{\theta}$ in njem standardnem odklonu $\sigma_{\hat{\theta}}$.

$$l(\theta = \hat{\theta} + \sigma_{\hat{\theta}}) = l_{\max} - \frac{1}{2} \frac{\sigma_{\hat{\theta}}^2}{\sigma_{\hat{\theta}}^2} = l_{\max} - \frac{1}{2}$$



•) za več parametrov $\vec{\theta}$

$$\frac{\partial \ell(\vec{x}|\vec{\theta})}{\partial \theta_j} \Big|_{\vec{\theta}=\hat{\vec{\theta}}} = \sum_i \frac{1}{\partial \theta_j} \log f_{x_i}(x_i, \vec{\theta}) \Big|_{\vec{\theta}=\hat{\vec{\theta}}} = 0$$

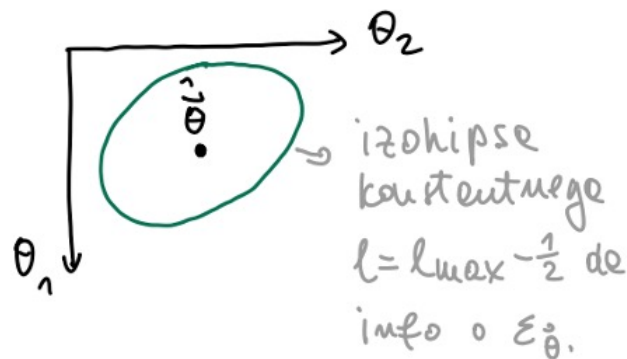
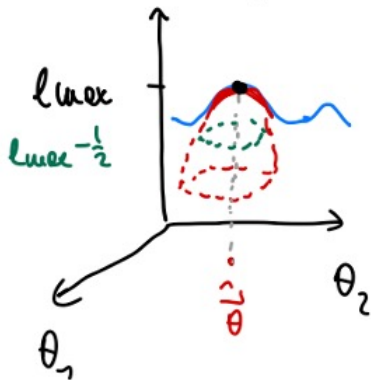
$$\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \Big|_{\vec{\theta}=\hat{\vec{\theta}}} = -A_{ij} = -\left(\frac{1}{\Sigma_{\vec{\theta}}} \right)_{ij}$$

$$\ell = \ell_{\max} - \frac{1}{2} (\vec{\theta} - \hat{\vec{\theta}})^T \frac{1}{\Sigma_{\vec{\theta}}} (\vec{\theta} - \hat{\vec{\theta}}) + \dots$$

neke pozitivno
definitne metrike

✓ meketeli bomo da ji drugi odvod po ℓ (poglejaj) teko povezan z korelacijsko metriko za $\hat{\vec{\theta}}$ (8.3.3)

$\ell(\vec{x}|\vec{\theta})$ torej v svojem drugem členu vsebuje info o $\hat{\vec{\theta}}$ in $\Sigma_{\vec{\theta}}$.



Velja: MLE cenilke $\hat{\theta}$ so nepristranske v limiti velikih n , drugače pa so v splošnem pristranske (brez dokaza, a bomo to osvetlili v 8.3)

$$E[\hat{\theta}] = \theta_0 \text{ za } n \rightarrow \infty$$

$E[...]$ je pričakovana vrednost po vseh vzorcih

$$E[\hat{\theta}] = \theta_0 + O(1/n) \text{ za velik } n$$

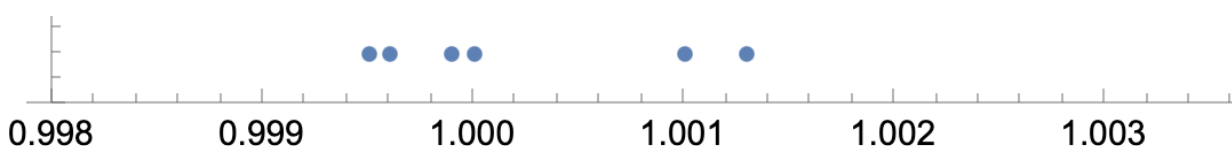
$$\lim_{n \rightarrow \infty} \text{var}[\theta_n] = 0$$

pokazali bomo da so dosledne

Zgled Tocke na sliki ponazarjajo izmerke

$\{x_1, x_2, \dots, x_6\} = \{0.9995, 0.9996, 0.9999, 1, 1.001, 1.0013\}$

za katere domnevamo, da izvirajo iz populacije, porazdeljene po Cauchyjevi porazdelitvi širino $s = 0.0001$ in neznanim povprečjem μ . Nal: Izračunajmo na podlagi tega vzorca oceno za povprečje $\hat{\mu}$ po metodi maksimalnega verjetja!



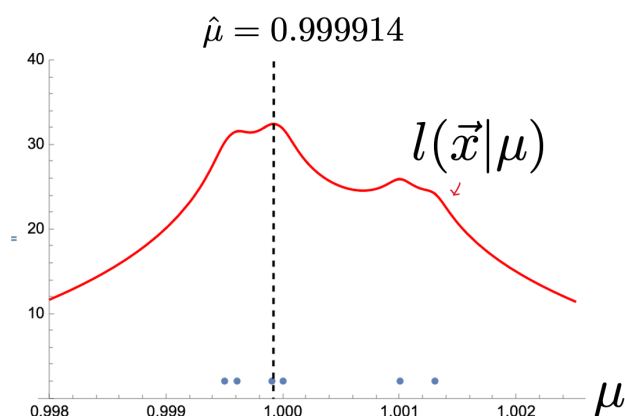
Logaritem funkcije verjetja za vzorec velikosti n je

$$\ell(\mathbf{x}|\mu) = \log \left(\prod_{i=1}^n \frac{1}{\pi} \frac{s}{s^2 + (x_i - \mu)^2} \right) = \sum_{i=1}^n \log \left(\frac{1}{\pi} \frac{s}{s^2 + (x_i - \mu)^2} \right),$$

enačba verjetja pa

$$\frac{\partial \ell}{\partial \mu} \propto 2 \sum_{i=1}^n \frac{x_i - \mu}{s^2 + (x_i - \mu)^2} = 0.$$

To lahko zapišemo v obliki $p(\mu) = 0$, kjer je p polinom stopnje $2n - 1$. Enačba verjetja ima torej v splošnem $2n - 1$ rešitev in nekatere od njih ustrezajo maksimumom funkcije ℓ . Optimalna vrednost $\hat{\mu}$ ustreza globalnemu maksimumu, ki ga poiščemo numerično in je običajno blizu vzorčne mediane. V našem primeru imamo štiri lokalne maksimume in globalnega pri $\hat{\mu} \approx 0.999914$.



Zgled Denimo, da z različnimi merskimi napravami (z različnimi natančnostmi oziroma negotovostmi σ_i) merimo isto količino. Izmerki $\{x_1, x_2, \dots, x_n\} = \mathbf{x}$ so torej raztreseni okrog prave vrednosti μ . Predpostavimo, da so odstopanja od povprečja porazdeljena normalno. Verjetnost, da je ob vrednosti parametra μ izmerek x_i na intervalu $[x_i, x_i + dx]$, je tedaj

$$f_X(x_i; \mu) dx = \frac{1}{\sqrt{2\pi} \sigma_i} \exp \left(-\frac{(x_i - \mu)^2}{2\sigma_i^2} \right) dx.$$

Ustrezna funkcija verjetja je

$$L(\mathbf{x}|\mu) = \prod_{i=1}^n f_X(x_i; \mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi} \sigma_i} \exp \left(-\frac{(x_i - \mu)^2}{2\sigma_i^2} \right),$$

njen logaritem pa

$$\ell = \log L(\mathbf{x}|\mu) = -\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_i^2} - \sum_{i=1}^n \log \sigma_i + \text{konst.} \quad (8.6)$$

Z rešitvijo enačbe verjetja

$$\frac{d\ell}{d\mu} = \sum_{i=1}^n \frac{x_i - \mu}{\sigma_i^2} = 0$$

dobimo oceno $\hat{\mu}$ za parameter μ :

$$\hat{\mu} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}, \quad w_i = \frac{1}{\sigma_i^2},$$

utez (weight w_i) za i -ti izmerek je obratno sorazmerna z njegovo varianco. Tisti z navecjo varianco bodo prispevali najmanj.

kar je standardna formula za uteženo povprečje izmerkov x_i z normalno porazdeljenimi negotovostmi σ_i . Drugi odvod ℓ po μ nam razodene tudi, da $\hat{\mu}$ res ustreza maksimumu ℓ , saj je $d^2\ell/d\mu^2 = -\sum_{i=1}^n w_i < 0$.

8.3 Pričakovana vrednost in varianca cenilke $\hat{\theta}$ za parameter θ

Doslednost in nepristranskost (glej definicije (7.2), (7.3) in (7.4)) nista edini lastnosti, ki ju pričakujemo od statistične cenilke. Želimo si, da ima tudi čim manjšo varianco. Različne cenilke iste količine imajo v splošnem različne variance: tako vzorčno povprečje kot tudi vzorčna mediana sta na primer dosledni in nepristranski cenilki za "težišče" normalne populacije z znano varianco. Toda — kot bomo pokazali v nadaljevanju — je varianca povprečja manjša kot varianca mediane, zato je vzorčno povprečje "boljša" cenilka.

Varianca cenilke $\hat{\theta}$ v limiti velikih vzorcev

8.3.1 varianca cenilke $\hat{\theta}$ za en parameter theta, vsi $f_{x_i}=f_X$ enaki

Izpeljimo izraz za varianco cenilke za en parameter za primer ko so vsi elementi vzorca porazdeljeni z enako verjetnostno gostoto $f_{x_i}=f_X$ (na koncu pa jo bomo posplošili) Želimo izpeljati $I''=-1/\text{var}[\theta]$ kot smo navedli v podpoglavju "načelo maksimalnega verjetja".

V primeru velikih vzorcev ($n \gg 1$) pričakujemo, da se bo ocena $\hat{\theta}$ le malo razlikovala od prave vrednosti parametra θ_0 . Zato enačbo verjetja (8.2), deljeno z n , razvijemo po Taylorju in obdržimo le prva člena:

$$\frac{1}{n} \frac{d\ell}{d\theta} \approx \underbrace{\frac{1}{n} \sum_{i=1}^n \left[\frac{\partial}{\partial \theta} \log f_X(x_i; \theta) \right]_{\theta_0}}_{\alpha(\mathbf{x}; \theta_0)} + (\theta - \theta_0) \underbrace{\frac{1}{n} \sum_{i=1}^n \left[\frac{\partial^2}{\partial \theta^2} \log f_X(x_i; \theta) \right]_{\theta_0}}_{\beta(\mathbf{x}; \theta_0)} = 0$$

oziroma na kratko

$$\alpha + (\theta - \theta_0)\beta = 0.$$

(8.8)

$$\hat{\theta} = \theta_0 - \frac{\alpha}{\beta}$$

Količini α in β imata naključno naravo, saj sta odvisni od vsakokratnega vzorca $\mathbf{x}$, θ_0 pa je vrednost parametra v populaciji. Zanima nas pričakovana vrednost cenilke in varianca. Pokazali bomo da za velike n velja

$$E[\hat{\theta}] = \theta_0, \quad \text{var}[\hat{\theta}] \propto 1/n$$

V ta namen moramo sešteti po vseh vzorcih ($x_1, \dots, x_n$). Ker so x_i med seboj v skladu z obravnavano predpostavko neodvisni to pomeni, da moramo sešteti po vseh vrednostih x_i neodvisno : $\sum_i x_i$ oziroma v zvezni limiti integrirati po vseh vrednostih x . Torej, za velike n lahko vsoti nadomestimo z integraloma: vlogo uteži $1/n$ pred vsoto prevzame verjetnostna gostota. Tedaj lahko zapišemo

$$E[\alpha] = \left[\int \frac{d \log f_X}{d\theta} f_X dx \right]_{\theta_0} = \left(\int \frac{1}{f_X} \frac{\partial f_X}{\partial \theta} f_X dx \right)_{\theta_0} = \left(\frac{\partial}{\partial \theta} \int f_X(x; \theta) dx \right)_{\theta_0} = 0$$

kjer smo v zadnjem koraku upoštevali, da je verjetnostna gostota f_X normirana ne glede na vrednost njenega parametra, $\int f_X(x; \theta) dx = 1$, odvod konstante pa je nič. Podobno ravnamo tudi z β :

$$\begin{aligned} E[\beta] &= E\left[\frac{\partial^2 \log f_X}{\partial \theta^2}\right]_{\theta_0} = \left(\int \left[\frac{1}{f_X} \frac{\partial^2 f_X}{\partial \theta^2} - \frac{1}{f_X^2} \left(\frac{\partial f_X}{\partial \theta}\right)^2\right] f_X dx\right)_{\theta_0} \\ &= \left(\frac{\partial^2}{\partial \theta^2} \int f_X(x; \theta) dx\right)_{\theta_0} - \left(\int \left(\frac{\partial \log f_X}{\partial \theta}\right)^2 f_X(x; \theta) dx\right)_{\theta_0}. \end{aligned}$$

Prvi člen je enak nič iz istega razloga kot pri računu $E[\alpha]$, v drugem členu pa uzremo (negativno) pričakovano vrednost količine $(\partial \log f_X / \partial \theta)^2$, torej

$$E[\beta] = E\left[\frac{\partial^2 \log f_X}{\partial \theta^2}\right]_{\theta_0} = -E\left[\left(\frac{\partial \log f_X}{\partial \theta}\right)^2\right]_{\theta_0} \neq 0 \quad (8.9)$$

za vse neizrojene primere. Iz enačbe (8.8) tedaj razberemo, da se $\hat{\theta}$ pri velikih n poljubno približa pravi vrednosti θ_0 , saj je $\lim_{n \rightarrow \infty} (\hat{\theta} - \theta_0) = -\lim_{n \rightarrow \infty} \alpha / \beta = -E[\alpha] / E[\beta] = 0$. Ocena za parameter θ_0 je torej

$$\hat{\theta} \approx \theta_0 - \frac{\alpha}{E[\beta]}.$$

Kolikšna je varianca te ocene? Dobimo jo kot pričakovano vrednost

$$\text{var}[\hat{\theta}] = E[(\hat{\theta} - \theta_0)^2] = E\left[\left(\frac{\alpha}{E[\beta]}\right)^2\right] = \frac{E[\alpha^2]}{E[\beta]^2}.$$

Imenovalci smo že izračunali v (8.9), števec pa je

$$E[\alpha^2] = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n E\left[\frac{\partial \log f_X(X_i; \theta)}{\partial \theta} \frac{\partial \log f_X(X_j; \theta)}{\partial \theta}\right]_{\theta_0}.$$

Posamezne naključne spremenljivke X_i so med seboj neodvisne, zato odpadejo vsi mešani členi, preživi le n kvadratičnih:

$$E[\alpha^2] = \frac{1}{n^2} \sum_{i=1}^n E\left[\left(\frac{\partial \log f_X(X_i; \theta)}{\partial \theta}\right)^2\right]_{\theta_0} = -\frac{E[\beta]}{n}.$$

Iskana varianca je torej

$$\text{var}[\hat{\theta}] = -\frac{1}{nE[\beta]} = -\frac{1}{n E\left[\frac{d^2}{d\theta^2} \log f_X\right]_{\theta_0}}$$

kjer smo v zadnjem enačaju uporabili (8.9). To lahko izrazimo z $E[l'']$, ker l v našem primeru vsebuje n enakih verjetnostnih gostot in ker so x_i neodvisni

$$E[l''] = E\left[\sum_{i=1}^n (\log f_X(x_i))''\right] = n E[(\log f_X)'']$$

Tako dobimo koncen izraz za varianco cenilke

odvodi
so po
theta

$$(\text{var}[\hat{\theta}])^{-1} = -E[l''] = -nE[(\log f_X)'']$$

Varianca cenilke po metodi maksimalnega verjetja res pada z n !

S tem smo utemeljili $E[l''] = -1/\text{var}[\theta]$ kar smo zapisali v podpoglavju "načelo maksimalnega verjetja"

Zgled: Določiti varianco cenilke za povprečje ($\hat{\mu}$) pri vzorcu velikosti n , kjer so vsi elementi vzorca porazdeljeni normalno z varianco σ . Uporabi metodo maksimalnega verjetja.

$$L(\vec{x}|\mu) = \prod_{i=1}^n \underbrace{\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)}_{f_X(x_i, \mu)}$$

$$\underline{\text{var}[\hat{\mu}]} = -\frac{1}{n E[(\log f_X)^2]} = -\frac{1}{n \left(-\frac{1}{\sigma^2}\right)} = \frac{\sigma^2}{n} = \underline{\left(\frac{\sigma}{\sqrt{n}}\right)^2}$$

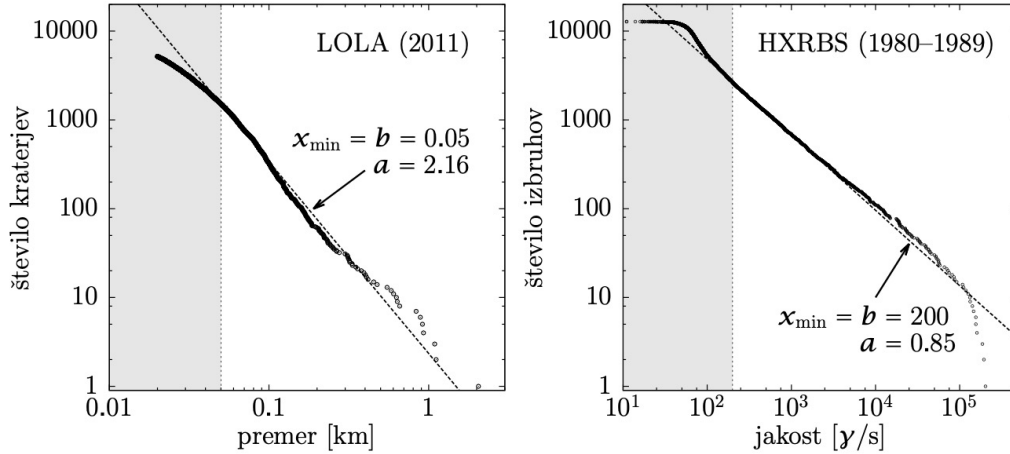
$$\log f_X = \text{const} - \frac{(x_i-\mu)^2}{2\sigma^2}, \quad \frac{d}{d\mu} \log f_X = \frac{2(x-\mu)}{2\sigma^2}, \quad \frac{d^2 \log f_X}{d\mu^2} = -\frac{1}{\sigma^2}$$

Varianca cenilke μ iz vzorca velikosti n je res n -krat manjše od variance σ^2 enega elementa.

Uvod v zgled s Paretovo porazdelitvijo (3.5), ki je pogosto prikladna za kolicine:

- katerih vrednosti se razprostirajo prek več redov velikosti,
- z veliko verjetnostjo za majhne vrednosti in majhno verjetnostjo za velike vrednosti ustreznih naključnih spremenljivk

$$f_X(x) = \frac{ab^a}{x^{a+1}} = \frac{a}{b} \left(\frac{b}{x} \right)^{a+1}, \quad 0 < b \leq x$$



Slika 3.5 — [LEVO] Porazdelitev luninih kraterjev po njihovem premeru, kot so jo določili raziskovalci projekta Lunar Orbiter Laser Altimeter (LOLA) [12, 13] do leta 2011. [DESNO] Porazdelitev izbruhov trdih rentgenskih žarkov po njihovi jakosti, izmerjenih s spektrometrom HXRBS (Hard X-Ray Burst Spectrometer) [14] med letoma 1980 in 1989. Ravni črti predstavljata približni potenčni odvisnosti, zaradi nazornosti vrisani tudi v senčenem območju, čeprav se Paretovi porazdelitvi začneta šele na njunih desnih robovih ($x \geq x_{\min}$).

+

Zgled Določimo parameter a v Paretovi porazdelitvi (3.16) na podlagi izmerjenega vzorca $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ ob predpostavki, da je parameter b znan. Funkcija verjetja je

$$L(\mathbf{x}|a) = \prod_{i=1}^n f_X(x_i; a) = \prod_{i=1}^n \frac{a}{b} \left(\frac{b}{x_i} \right)^{a+1},$$

njen logaritem pa

$$\ell = \sum_{i=1}^n \left[\log \frac{a}{b} + (a+1) \log \frac{b}{x_i} \right] = n \log \frac{a}{b} + (a+1) \sum_{i=1}^n \log \frac{b}{x_i}.$$

Iz enačbe verjetja $\partial \ell / \partial a = 0$ sledi $n/a + \sum_{i=1}^n \log(b/x_i) = 0$, od tod pa ocena

$$\hat{a} = n \left[\sum_{i=1}^n \log \frac{x_i}{b} \right]^{-1}. \quad (8.11)$$

Po formuli (8.10) izračunamo še njeno varianco:

$$\begin{aligned} (\text{var}[\hat{a}])^{-1} &= -n E \left[\left(\log f_X \right)'' \right] = -n E \left[\left(\log \frac{a}{b} + (a+1) \log \frac{b}{X} \right)'' \right] \\ &= -n E \left[\left(\frac{1}{a} + \log \frac{b}{X} \right)' \right] = -n E \left[-\frac{1}{a^2} \right] = \frac{n}{a^2}, \end{aligned}$$

torej $\text{var}[\hat{a}] = a^2/n$.

odvod tega clena po a je nic, zato ta clen ne prispeva
(v nasprotnem primeru bi morali izracunati pricakovano vrednost po porazdelitvi)

◀

8.3.2 Varianca cenilke $\hat{\theta}$ za en parameter θ , posplošitev na različne f_{x_i}

Zgornja izvepljava se po analoginih korakih posploši tudi na primer različnih verjetnostnih gostot, kjer pač pričakovane vrednosti $E[.]$ po vseh vzorcih dobimo z integracijami po $f_{x_i}(x_i)$. Izkaze se, da dobimo analogen rezultat za varianco cenilke theta

odvodi
so po
theta

$$(var[\hat{\theta}])^{-1} = -E[l''] = -E\left[\sum_{i=1}^n (\log f_{X_i})''\right]$$

ki očitno preide v prejšnji izraz ko je za vseh n elementov f_{x_i} enak.

8.3.3 Varianca cenilke $\hat{\theta}$ za več parametrov θ_k , različni f_{x_i} .

V primeru p -parametrov theta nas zanima celotna $p \times p$ kovariančna matrika, ki je posplošitev variance $var[\theta]$ za en parameter. Ta je po metodi maksimalnega verjetja in v skladu z enakimi približki enaka (brez dokaza)

$p \times p$ covariančna matrika za p parametrov theta

$$(\Sigma_{\hat{\theta}}^{-1})_{kl} = (cov[\hat{\theta}]^{-1})_{kl} = -E\left[\frac{\partial}{\partial \theta_k} \frac{\partial}{\partial \theta_l} l\right]$$

Izraz za en sam parameter očitno preide v prejšnji izraz iz poglavij 8.3.1 in 8.3.2.

S tem smo nakazali zakaj je drugi odvod po l povezan z kovariančno matriko parametrov, kar smo zapisali v podpoglavju "nacelo maksimalnega verjetja"

9. METODA NAJMANJŠIH KVADRATOV

V merenostavju se domela dnevno srečujemo z problemom prilagajanja izbrane funkcije (g) parom $(t_i, x_i \pm \sigma_i)$, $i=1, \dots, m$. Če je vsake izmed meritev x_i dobljene kot povprečje večjega števila meritev, je meključne spremenljivke x_i porazdeljene Normalno z standardnim odklonom σ_i (ki ustrezata standardnemu odklonu povprečja x_i teh meritev) - to sledi iz centralnega limitnega teorema. Torej iz metode maksimalnega verjetja sledi metode najmanjših kvadratov kot bomo izpeljali v naslednjem - in slednja je memere najpogostejši uporebljene pri prilagajanju.

9.1, 9.6 Splošna melinearna regresija in ustrezen χ^2

komentar glede notacij

$\vec{y} = f(\vec{x}, \vec{\theta})$	$\vec{x} = g(t, \vec{\theta})$
--------------------------------------	--------------------------------

• ker je χ^2 verjetnostna gostota, funkcijo rej označimo g
 • Privadili smo se da so meključne spremenljivke x_i , zato to okrevamo

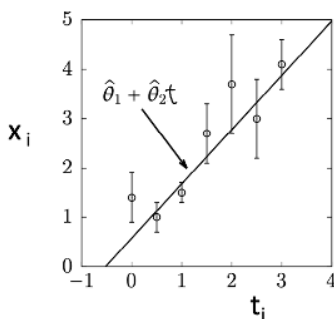
• medvisne meključne spremenljivke x_i

m meritev $(t_i, x_i \pm \sigma_i)$, $i=1, \dots, m$:

prčekovene vrednosti $E[x_i] = \mu_i$ bi želeli opisati $E[x_i] = g(t_i; \vec{\theta})$

ker pe so x_i normalno porazdeljene meključne spremenljivke z varianco σ_i^2 . Tu je $\vec{\theta} = (\theta_1, \dots, \theta_p)$ vektor merenih parametrov.

$$L(\vec{x} | \vec{\theta}) = \prod_i \frac{1}{\sqrt{2\pi} \sigma_i} \exp \left[- \frac{(x_i - g(t_i, \vec{\theta}))^2}{2 \sigma_i^2} \right], \quad l = \log L$$



Zgled za kaj bomo to na primer uporabili;
 funkcija g je lahko tudi nelinearna funkcija parametrov theta

$$l(\vec{x}|\vec{\theta}) = \log L = \text{konst} - \frac{1}{2} \chi^2, \quad \chi^2 = \sum_i \frac{(x_i - g(t_i, \vec{\theta}))^2}{\sigma_i^2}$$

$$l(\vec{x}|\vec{\theta}) \Big|_{\vec{\theta} = \hat{\vec{\theta}}} \rightarrow \chi^2 \Big|_{\vec{\theta} = \hat{\vec{\theta}}} = \min$$

Maksimalno verjetj $\acute{e}$ torej dose $\acute{e}$ mo z minimizacijo χ^2 v odvisnosti od parametrov $\vec{\theta}$.

- odvisne meklju $\acute{c}$ ne spremenljivke x_i
za katere poznamo kovariacijo metriko $\Sigma_{\vec{x}}: m \times m$

V tem primeru je funkcija verjetj $\acute{e}$ podana z multivariacno normalno porazdelitvijo ki smo jo izpeljali v poglavju 4.8.1

$$\underbrace{\vec{x} = (x_1, \dots, x_m), \quad t = (t_1, \dots, t_m), \quad \vec{g}(\vec{t}, \vec{\theta}) = (g(t_1, \vec{\theta}), \dots, g(t_m, \vec{\theta}))}_{\text{vektorji dol $\acute{z}$ ine } m}$$

$$L(\vec{x}|\vec{\theta}) = \frac{1}{\sqrt{2\pi}^m} (\det(\Sigma_{\vec{x}}))^{-\frac{1}{2}} \exp \left(-\frac{1}{2} [\vec{x} - \vec{g}(\vec{t}, \vec{\theta})]^T \Sigma_{\vec{x}}^{-1} [\vec{x} - \vec{g}(\vec{t}, \vec{\theta})] \right)$$

$$l(\vec{x}|\vec{\theta}) = \log L = \text{konst} - \frac{1}{2} \chi^2, \quad \chi^2 = [\vec{x} - \vec{g}(\vec{t}, \vec{\theta})]^T \Sigma_{\vec{x}}^{-1} [\vec{x} - \vec{g}(\vec{t}, \vec{\theta})]$$

$$l(\vec{x}|\vec{\theta}) \Big|_{\vec{\theta} = \hat{\vec{\theta}}} \rightarrow \chi^2 \Big|_{\vec{\theta} = \hat{\vec{\theta}}} = \min$$

Ze $\Sigma_{\vec{x}}^{\textcircled{*}} = \text{diag}(\sigma_1^2, \dots, \sigma_m^2)$ sledi zveze ze neodvisne x_i tgoraj.

V modeljevanju izpeljave naprevimo ze splošnejši primer koreliranih meritev, ne neodvisne pa zleeka preidemo preko $\textcircled{*}$.

Napake me parametrik $\vec{\theta}$:

variance $\text{var}(\vec{\theta})$ in korelacijske metrike $\Sigma_{\vec{\theta}}$

V najnižjem netrivialnem redu Taylorjeve razvoja $\vec{\theta} - \vec{\theta}_0$ bomo izpeljali me dve načini

$$\Sigma_{\vec{\theta}} = (\mathbf{J}_0^T \Sigma_{\vec{x}}^{-1} \mathbf{J}_0)^{-1}$$

$$(\mathbf{J}_0)_{ik} = \left. \frac{\partial g_i}{\partial \theta_k} \right|_{\vec{\theta} = \vec{\theta}_0} ; i=1, \dots, m ; k=1, \dots, p$$

Jacobijeva metrike
odvodov funkcije g

Obe izpeljavi zgoraj zveze temeljita me $\vec{g}_0$

$$\ell = -\frac{1}{2} \chi^2 = -\frac{1}{2} \vec{\pi}^T \Sigma_{\vec{x}}^{-1} \vec{\pi}, \quad \vec{\pi} = \vec{x} - \vec{g}(\vec{t}, \vec{\theta}) \doteq \vec{x} - \vec{g}(\vec{t}, \vec{\theta}_0) - \mathbf{J}_0 (\vec{\theta} - \vec{\theta}_0)$$

$\vec{\pi}$ = vektor
residualov

$$\underbrace{\begin{bmatrix} \square & \square & \square \end{bmatrix}}_m \underbrace{\begin{bmatrix} \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \end{bmatrix}}_{m \times m} \underbrace{\begin{bmatrix} \square \\ \square \\ \square \end{bmatrix}}_m$$

$$m \left\{ \begin{bmatrix} \square \\ \square \\ \square \end{bmatrix} \right\}$$

$$m \left\{ \underbrace{\begin{bmatrix} \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \end{bmatrix}}_p \underbrace{\begin{bmatrix} \square & \square \\ \square & \square \\ \square & \square \end{bmatrix}}_p \right\}$$

1. vertikalna izpeljava:

Uporebimo $(\Sigma_{\vec{\theta}}^{-1})_{ke} = -E \left[\frac{\partial}{\partial \theta_k} \frac{\partial}{\partial \theta_e} \ell \right]$ ker smo "meketoli"

v pogledu o metodi maksimalne verjetje: $\ell = -\frac{1}{2} \pi_i (\Sigma_{\vec{x}}^{-1})_{ij} \pi_j$

$$\frac{\partial \ell}{\partial \theta_k} = -\frac{1}{2} \frac{\partial \pi_i}{\partial \theta_k} (\Sigma_{\vec{x}}^{-1})_{ij} \pi_j - \frac{1}{2} \pi_i (\Sigma_{\vec{x}}^{-1})_{ij} \frac{\partial \pi_j}{\partial \theta_k} \doteq -\frac{\partial \pi_i}{\partial \theta_k} (\Sigma_{\vec{x}}^{-1})_{ij} \pi_j = -(\mathbf{J}_0^T)_{ki} (\Sigma_{\vec{x}}^{-1})_{ij} \pi_j$$

dve enaki člene ker je $\Sigma_{\vec{x}}$ simetrična

(*)

$$(\Sigma_{\vec{\theta}}^{-1})_{ke} = -E \left[\frac{\partial}{\partial \theta_k} \frac{\partial}{\partial \theta_e} \ell \right] = +E \left[(\mathbf{J}_0^T)_{ki} (\Sigma_{\vec{x}}^{-1})_{ij} \frac{\partial \pi_j}{\partial \theta_e} \right] = E \left[(\mathbf{J}_0^T)_{ki} (\Sigma_{\vec{x}}^{-1})_{ij} (\mathbf{J}_0)_{je} \right]$$

$$\Sigma_{\vec{\theta}}^{-1} = \mathbf{J}_0^T \Sigma_{\vec{x}}^{-1} \mathbf{J}_0$$

2. vertija izpeljave Prednost druge izpeljave je, da me

slowi me zvezi $(\Sigma_{\vec{\theta}}^{-1})_{kl} = -E[\frac{1}{\partial \theta_k} \frac{1}{\partial \theta_l} l]$ ki jo misimo v splošnem izpeljati. Vse korake druge izpeljave dejansko utemeljimo. Načrt dela: določimo $\vec{\theta} - \vec{\theta}_0$ ki sledi iz minimizacije χ^2 , nato določimo koveriančno metriko za $\vec{\theta} - \vec{\theta}_0$.

Podobna izpeljave je v knjigi (poglavje 9.1) kjer imamo mi v mislih splošno nelinearno funkcijo $\vec{g}$ od $\vec{\theta}$, ki pa jo razvijemo v bližini $\vec{\theta} = \vec{\theta}_0$.

Ekstremizacija $l = -\frac{1}{2} \chi^2$ da

$$\frac{\partial l}{\partial \theta_k} = 0 \xrightarrow[\text{pri prvi izpeljavi}]{\text{glej } (*)} \mathbf{J}_0^T \Sigma_{\vec{x}}^{-1} \vec{r} = 0$$

$$\mathbf{J}_0^T \Sigma_{\vec{x}}^{-1} (\vec{x} - \vec{g}_0 - \mathbf{J}_0 (\vec{\theta} - \vec{\theta}_0)) = 0$$

Izrežimo $\vec{\theta} - \vec{\theta}_0$ z $\vec{x} - \vec{g}_0$:

$$\vec{\theta} - \vec{\theta}_0 = \underbrace{(\mathbf{J}_0^T \Sigma_{\vec{x}}^{-1} \mathbf{J}_0)^{-1} \mathbf{J}_0^T \Sigma_{\vec{x}}^{-1}}_{\equiv} (\vec{x} - \vec{g}_0)$$

Zdej se spomnimo da smo izpeljali, da če sta dve količini povezani $\vec{y}(\vec{x}) - \vec{y}(\vec{\mu}) = \mathbf{D}(\vec{x} - \vec{\mu}) \rightarrow \Sigma_{\vec{y}} = \mathbf{D} \Sigma_{\vec{x}} \mathbf{D}^T$ (formule 4.27 iz knjige) torej če $\vec{y} \rightarrow \vec{\theta}$, $\mathbf{D} \rightarrow \equiv$ dobimo

$$\Sigma_{\vec{\theta}} = \Sigma_{\vec{\theta} - \vec{\theta}_0} = \equiv \Sigma_{\vec{x} - \vec{g}_0} \equiv^T = \equiv \Sigma_{\vec{x}} \equiv^T = \dots = (\mathbf{J}_0^T \Sigma_{\vec{x}}^{-1} \mathbf{J}_0)^{-1}$$

Zgled:

Prilagajanje splosne funkcije s koreliranimi podatki za x

n=9

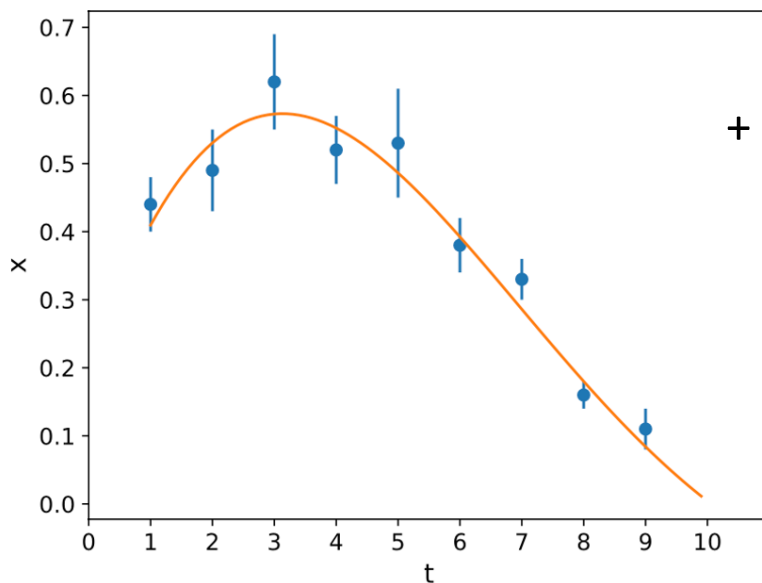
t= [1, 2, 3, 4, 5, 6, 7, 8, 9]

x=[0.44, 0.49, 0.62, 0.52, 0.53, 0.38, 0.33, 0.16, 0.11]

Sigma_x=

```
[[0.00160, -0.00054, 0.00050, -0.00009, 0.00070, -0.00044, -0.00052, 0.00045, -0.00005],
[-0.00054, 0.00250, -0.00060, -0.00014, 0.00089, 0.00049, -0.00063, 0.00016, -0.00016],
[0.00050, -0.00060, 0.00360, 0.00058, -0.00086, -0.00012, -0.00039, 0.00062, -0.00040],
[-0.00009, -0.00014, 0.00058, 0.00250, 0.00082, -0.00049, 0.00044, 0.00036, -0.00053],
[0.00070, 0.00089, -0.00086, 0.00082, 0.00360, 0.00042, 0.00006, -0.00064, 0.00002],
[-0.00044, 0.00049, -0.00012, -0.00049, 0.00042, 0.00160, 0.00018, 0.00014, 0.00039],
[-0.00052, -0.00063, -0.00039, 0.00044, 0.00006, 0.00018, 0.00250, -0.00001, 0.00012],
[0.00045, 0.00016, 0.00062, 0.00036, -0.00064, 0.00014, -0.00001, 0.00160, -0.00023],
[-0.00005, -0.00016, -0.00040, -0.00053, 0.00002, 0.00039, 0.00012, -0.00023, 0.002]]
```

$$x(t) = g(t; \theta_1, \theta_2) = e^{-\theta_1 t} \sin(\theta_2 t + 0.2)$$



rezultati:

- minimizacija(chi2), $\hat{\theta} = (J^T \hat{\Sigma}_x^{-1} J)^{-1}$
- curve_fit
- minuit

$$\theta_1 = 0.1433 \pm 0.009$$

$$\theta_2 = 0.284 \pm 0.006$$

$$\Sigma_{\hat{\theta}} = \begin{bmatrix} 8.4e-05 & 3.6e-06 \\ 3.6e-06 & 3.8e-05 \end{bmatrix}$$

$$\frac{\chi^2}{9-2} = 0.84$$

glej: gradiva, DN: naredi primerjavo za en primer fita tudi sam !

fit-compare-curvefit-iminuit-correlated-VF.py

data-compare-curvefit-iminuit-correlated-VF.txt podatki

plot-compare-curvefit-iminuit-correlated-VF.pdf slika

result-compare-curvefit-iminuit-correlated-VF.txt rezultati

Zgled

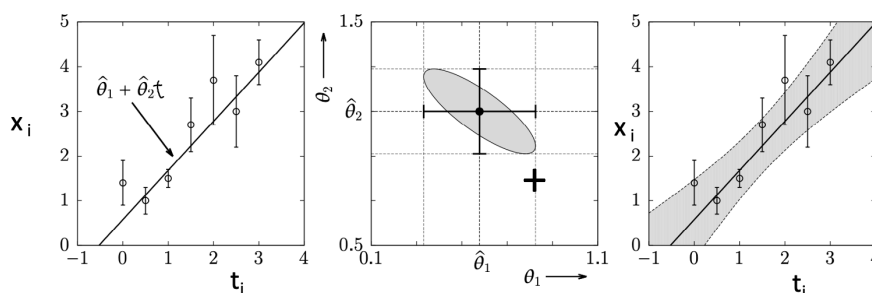
Prilagajanje premice

t_i = 0, 0.5, 1, 1.5, 2, 2.5, 3

(nekorelirani) izmerki $x_i = 1.4, 1.0, 1.5, 2.7, 3.7, 3.0, 4.1$

$$x = g(t, \vec{\theta}) = \theta_1 + \theta_2 t$$

z napakami $\sigma_i = 0.5, 0.3, 0.2, 0.6, 1.0, 0.8, 0.5$



$$\hat{\theta}_1 = 0.578$$

$$\pm \sigma(\hat{\theta}_1) = \pm 0.247$$

$$\hat{\theta}_2 = 1.100$$

$$\pm \sigma(\hat{\theta}_2) = \pm 0.190$$

$$\chi^2 = 1.1$$

manjka rezultat za korelacijo koncnih parametrov.

[Desno]. Premica, ki ustreza pravima vrednostma θ_1 in θ_2 , z verjetnostjo $1 - e^{-1/2}$ leži znotraj te pahljače.

V naslednjem podpoglavju bomo opredelili kako to določimo.

Veliko zgledov za prilagajanje; naloge iz poglavja 7 v knjigi (a v vecini primerov kovariančne matrike začetnih podatkov in koncnih parametrov niso podane)

Kako dobro je prilagajanje? Za ta namen je relevantna porazdelitev naključne spremenljivke χ^2 (po različnih vzorcih $\vec{x}$)

Za vsak vzorec lahko določimo parametre $\vec{\theta} = \hat{\vec{\theta}}$, in posledično bo vrednost naključne spremenljivke χ^2 varirala od vzorca do vzorca. Pričakujemo, da bo ta naključna spremenljivka porazdeljena v skladu z porazdelitvijo χ^2 z $\nu = n - p$ prostostnimi stopnjami :

$$f_{\chi^2}(\chi^2, \nu = n - p)$$

To pričakujemo v skladu z izpeljavami iz poglavja 7

- če so izmerki x_i res porazdeljeni v skladu z normalno porazdelitvijo z danimi $\text{var}[x_i]$
- če funkcijska odvisnost $g(t; \theta)$ res dobro opisuje pričakovano vrednost x_i
- če smo s prilagajanjem res določili prave parametre
- če so x_i neodvisni in smo uporabili $\chi^2 = \sum_i (x_i - g)^2 / \sigma_i^2$:
(za ta primer smo utemeljili χ^2 porazdelitev v poglavju 7)
- zgornja porazdelitev z $\nu = n - p$ je relevantna tudi za primere koreliranih x_i
ce uporabimo koreliran $\chi^2 = (x - g)^T \Sigma_x^{-1} (x - g)$ kjer je Σ kovarianca matrika

Pričakovana vrednost porazdelitve χ^2 in njena varianca sta

$\nu = n - p = \text{dof}$: število prostostnih stopenj (number of degrees of freedom=dof)

$$E[\chi^2] = \nu, \quad \frac{E[\chi^2]}{\nu} = \frac{E[\chi^2]}{\text{dof}} = 1, \quad \text{var}[\chi^2] = 2\nu$$

reduciran χ^2

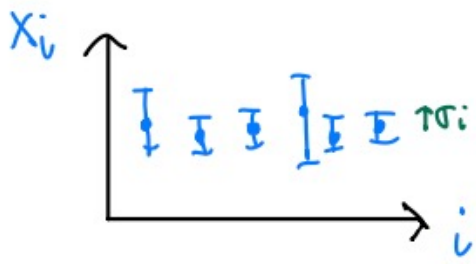
V primerih, kjer so upravičene predpostavke iz zgornjih alinej, pričakujemo, da bo reduciran χ -kvadrat (χ^2/ν) v povprečju enak 1. Ali obratno: če pri prilagajanju na vzorcu dobimo χ^2/ν približno enak 1 v grobem sklepamo, da so bile zgornje predpostavke upravičene.

Če pa dobimo χ^2/ν veliko večji od 1, običajno določimo **kolikšna je verjetnost (to verjetnost imenujemo vrednost-p = p-value)** da bo χ^2 tako ekstremen kot ga je dala minimizacija na vzorcu (imenujmo tega χ^2_{vz}) ali se bolj ekstremen

$$\text{vrednost-p} = P(\chi^2 \geq \chi^2_{vz}) = \int_{\chi^2_{vz}}^{\infty} f_{\chi^2}(\chi^2; n - p) d\chi^2 \quad (\text{p-value})$$

Velik χ^2/ν ali zelo majhna vrednost p nakazujeta na slabo kakovost prilagajanja. S zadnjo zvezo torej ocenimo **kakovost prilagajanja (goodness of fit)**. Več o tem bomo spregovorili v poglavju o testiranju hipotez.

Prilagajanje konstante (9.1.8)



vzemimo nekorrelirane podatke

$$x_i \pm \sigma_i, \quad i=1, \dots, m$$

$$g(t, \theta) = g(\theta) = \theta \rightarrow \chi^2 = \sum_i \frac{(x_i - \theta)^2}{\sigma_i^2}$$

ni odvisnosti od t

$$\bullet) \hat{\theta}: \left. \frac{d\chi^2}{d\theta} \right|_{\theta=\hat{\theta}} = -2 \sum_i \frac{(x_i - \theta)}{\sigma_i^2} \Big|_{\theta=\hat{\theta}} = 0 \rightarrow \hat{\theta} = \frac{\sum_i \frac{x_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}}$$

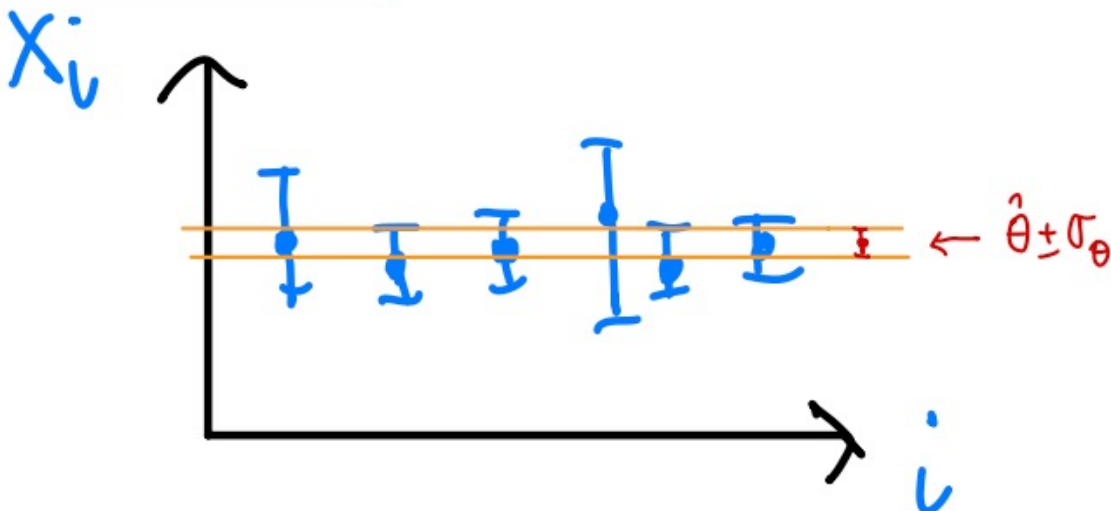
↓

smislen rezultat saj so meritve z manjšo napako σ_i^2 bolj utežene

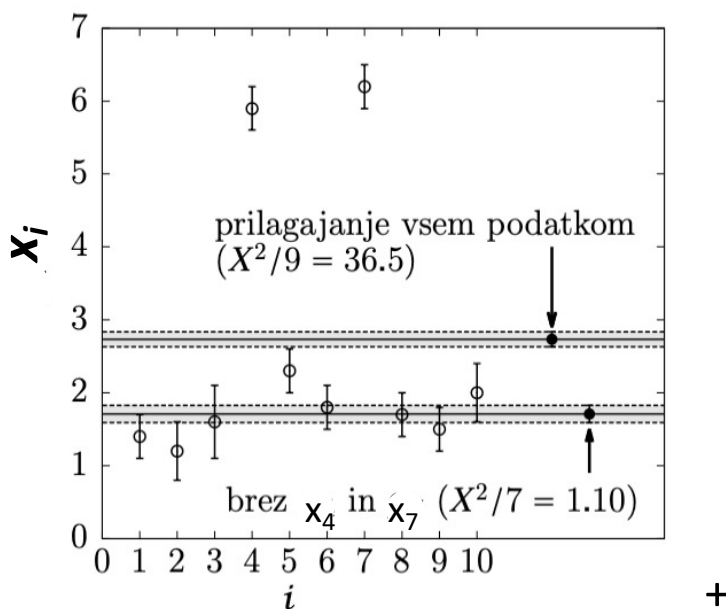
$$\bullet) \text{var}[\hat{\theta}]: \underset{\substack{\uparrow \\ 1 \times 1}}{\Sigma_{\hat{\theta}}} = \left[\underset{\substack{\uparrow \\ 1 \times 1}}{J_0^T} \underset{\substack{\uparrow \\ m \times m}}{\Sigma_{\tilde{x}}} \underset{\substack{\uparrow \\ 1 \times 1}}{J_0} \right]^{-1} = \frac{1}{\left| \begin{matrix} 1/\sigma_1^2 & \dots & 1/\sigma_m^2 \end{matrix} \right| \left| \begin{matrix} 1 \\ \vdots \\ 1 \end{matrix} \right|^2} = \frac{1}{\sum_i \frac{1}{\sigma_i^2}}$$

$$J_0 = \frac{dg_i}{d\theta} = \left\{ \begin{matrix} 1 \\ \vdots \\ 1 \end{matrix} \right\}_m$$

$$\frac{1}{\sigma_{\theta}^2} = \sum_i \frac{1}{\sigma_i^2}$$



Zgled Za vso to preprostostjo se skrivajo številne pasti. Slika 9.4 (levo) prikazuje $n = 10$ izmerkov, ki jim prilagajamo konstanto θ . Z opisanim postopkom dobimo $\hat{\theta} = 2.73 \pm 0.10$ in $X^2/(n - 1) = 36.5$,



Slika 9.4 — Prilagajanje konstante meritvam $\{x_i \pm \sigma_i\}_{i=1}^n$, $n = 10$. [LEVO] Uteženo povprečje v prisotnosti dveh ubežnih vrednosti (x_4 in x_7) in brez njiju. [DESNO] Uteženo povprečenje podatkov z nepojasnjeno sistematično napako, kjer dobimo nesmiselno majhno negotovost parametra, in povprečje z reskaliranimi napakami (9.23).

Ogromen χ^2/ν že sam po sebi nakazuje na to, da je bila ena od predpostavk (iz alinej na prejsnji strani) napačna. Verjetnost (p-value) za tako velik ali se večji χ^2 je izjemno majhna, tako majhna da Mathematica vrne 0

$$\begin{aligned} \text{p-value} &= P(\chi^2 \geq 36.5 \cdot 9) = \int_{328.5}^{\infty} f_{\chi^2}(\chi^2; 9) d\chi^2 = \\ &= 1 - \text{CDF}[\text{ChiSquareDistribution}[9], 36.5 \cdot 9] = 0.0000* \end{aligned}$$

Če opustimo ubežni vrednosti x_4 in x_7 , dobimo $\hat{\theta} = 1.71 \pm 0.12$ in $X^2/7 = 1.10$, in zmerno vrednost p, kar kaže na ustrezno prilagajanje

$$\text{p-value} = P(\chi^2 \geq 1.1 \cdot 7) = \int_{7.7}^{\infty} f_{\chi^2}(\chi^2; 7) d\chi^2 = 0.36$$

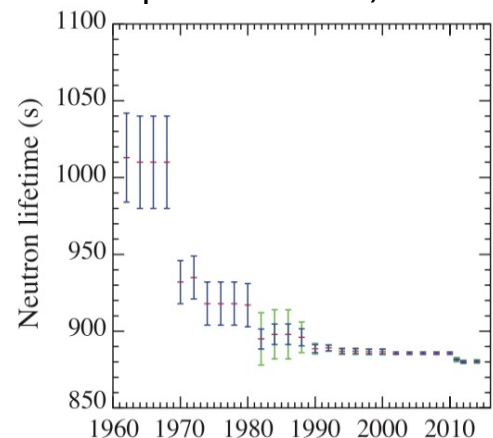
a seveda le, če obstaja utemeljen razlog, da smo dva ubežna podatka zavrgli.

9.1.9 Ali smemo kakšen podatek preprosto zavreči?

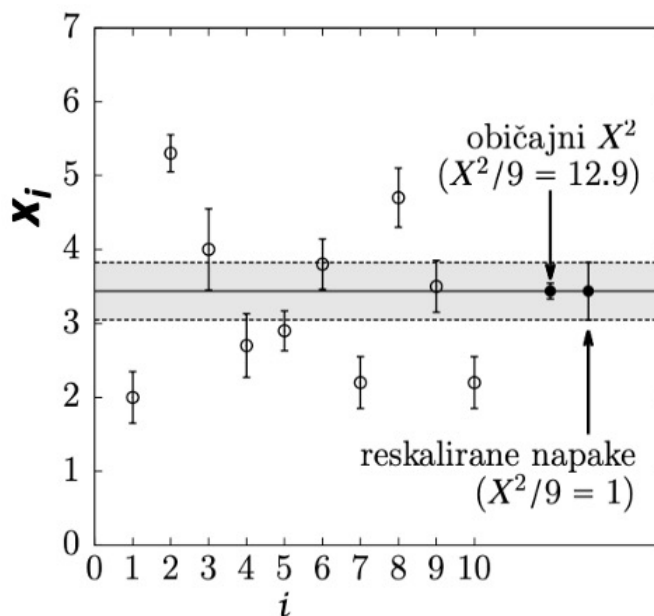
Med razloge, zaradi katerih smemo določeno meritev ali posamezen izmerek zavreči, po merilih skupine Particle Data Group sodijo [5]:

- obstaja kasnejša meritev, ki je boljša od prejšnje ali vključuje prejšnjo kot svojo podmnožico;
- negotovost izmerka ni podana;
- meritev temelji na vprašljivih predpostavkah;
- meritev ima nizko razmerje med signalom in šumom, ima majhno statistično pomembnost ali je tudi sicer slabše kakovosti kot druge meritve;
- izmerek je očitno neskladen z drugimi rezultati, ki so videti zanesljivejši;
- meritev ni neodvisna od ostalih rezultatov.

Slika prikazuje povprečne vrednosti razpadnega časa nevtrona, kakršne smo poznali v letih med 1960 in 2015. (Vsako leto je bilo narejenih več neodvisnih meritev; prikazana so letna povprečja.) Razmisli, katero od navedenih vrednosti je smiselno uporabiti danes, če upoštevaš našeta merila!



Zgled Drugačno zadrego prikazuje slika 9.4 (desno). V tem primeru dobimo $X^2/9 = 12.9$. Če smo prepričani, da naj bi količino dobro opisovala konstanta, za velik izračunani X^2 niso krivi posamezni ubežniki, temveč izmerki očitno vsebujejo neko neznano, podcenjeno sistematično napako.



Slika 9.4 — Prilagajanje konstante meritvam $\{x_i \pm \sigma_i\}_{i=1}^n$, $n = 10$. [LEVO] Uteženo povprečje v prisotnosti dveh ubežnih vrednosti (x_4 in x_7) in brez njiju. [DESNO] Uteženo povprečenje podatkov z nepojasnjeno sistematično napako, kjer dobimo nesmiselno majhno negotovost parametra in povprečje z reskaliranimi napakami (9.23).

V takšnih primerih **včasih** mersko napako reskalirajo (če so utemeljeno izključili druge možnosti za velik χ^2):

$$\sigma'_i = \sigma_i \sqrt{\frac{X^2}{n-1}}, \quad i = 2, \dots, n. \quad (9.23)$$

Uteženo povprečje za izračun iskanega parametra še vedno tvorimo po (9.22), le da njegova negotovost zdaj postane

$$\sigma_\theta = \sqrt{\text{var}[\hat{\theta}]} = \sqrt{\frac{X^2}{n-1}} \left(\sum_{i=1}^n \frac{1}{\sigma_i^2} \right)^{-1/2}.$$

To sicer vodi do rezultata z reduciranim $\chi^2/(n-1)=1$, a sama imam pri tem postopku pomisleke.

Pozor: **curve_fit** v pythonu z opcijo **absolute_sigma=False** (default !!) sam reskalira napake na xdata tako da bi reduciran χ^2 enak 1 in potem vrne napake na parametrih v skladu z reskaliranimi xdata!

`curve_fit(g,tdata,xdata, ... absolute_sigma=False)`

absolute_sigma : bool, optional

If True, *sigma* is used in an absolute sense and the estimated parameter covariance *pcov* reflects these absolute values.

If False (default), only the relative magnitudes of the *sigma* values matter. The returned parameter covariance matrix *pcov* is based on scaling *sigma* by a constant factor. This constant is set by demanding that the reduced *chisq* for the optimal parameters *popt* when using the *scaled sigma* equals unity. In

https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html

Nauk:

- raje sami napisite funkcijo χ^2 ki jo je treba minimizirati
- ce uporabljate crne skatle dobro preberite in preverite kaj pocno na preprostem primeru

9.1.3 Intervali zaupanja za optimalne parametre

populacija : θ_j^0 , $\sigma_{\theta_j}^0$

vzorec : $\hat{\theta}_j$, $\sigma_{\hat{\theta}_j} = \sqrt{\text{var}[\hat{\theta}_j]}$

Zelimo določiti interval zaupanja za parameter θ_j^0 v populaciji. Poznamo pa vzorčno vrednost cenilke za θ_j in pripadajočo vzorčno varianco. Določili smo ju iz izmerkov (t_i, x_i) , kjer so x_i porazdeljeni normalno. Naključna spremenljivka T je zopet porazdeljena po Studentovi porazdelitvi (brez dokaza)

$$T = \frac{\hat{\theta}_j - \theta_j^0}{\sigma_{\hat{\theta}_j}} : f_T(t, \nu = n - p)$$

Intervale zaupanja za parametre θ izračunamo po analogiji z intervali zaupanja za vzorčno povprečje (podrazdelek 7.3.1). Če napak σ_i izmerkov y_i ne poznamo — predpostavimo le, da so med seboj neodvisne in identično porazdeljene — najprej izračunamo kovariančno matriko (9.18) in iz nje izluščimo variance

$$\text{var}[\hat{\theta}_j] = (\Sigma_{\hat{\theta}})_{jj} , \quad j = 1, 2, \dots, p .$$

Parametru θ_j tedaj pri izbrani stopnji zaupanja $1 - \alpha$ ustreza interval zaupanja

$$\left[\hat{\theta}_j - t_* \sqrt{\text{var}[\hat{\theta}_j]} , \hat{\theta}_j + t_* \sqrt{\text{var}[\hat{\theta}_j]} \right] , \quad (9.19)$$

na katerem leži prava (neznana) vrednost parametra θ_j . Mejno vrednost t_* razberemo iz tabele 7.1, pri čemer upoštevamo število prostostnih stopenj $\nu = n - p$. Z drugimi besedami,

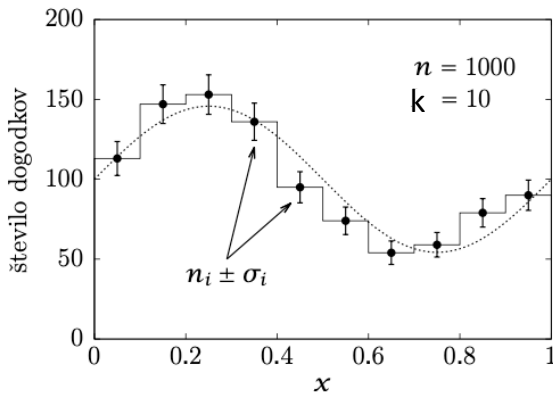
$$P \left(\hat{\theta}_j - t_* \sqrt{\text{var}[\hat{\theta}_j]} \leq \theta_j \leq \hat{\theta}_j + t_* \sqrt{\text{var}[\hat{\theta}_j]} \right) = 1 - \alpha = \text{CL}$$

Če je $n - p \gg 1$, porazdelitev t preide v standardizirano normalno porazdelitev in namesto t_* uporabimo z_* iz zadnje vrstice tabele 7.1.

Z verjetnostjo $\text{CL} = 1 - \alpha$ se prava vrednost populacijskega parametra θ_j nahaja na zgornjem intervalu. S tem imamo v mislih: pri mnogokratni meritvi vzorcev bo v deležu $\text{CL} = 1 - \alpha$ vzorcev dejanska populacijska vrednost parametra znotraj vzorčnega intervala zaupanja. Pri tem se seveda dejanska populacijska vrednost parametra od vzorca do vzorca ne spreminja, interval zaupanja pa se od vzorca do vzorca spreminja.

9.2 Metoda maksimalnega verjetja in približek z metodo najmanjših kvadratov za histogramirane podatke

Marsikdaj izmerjene podatke *histogramiramo* po spremenljivki x . To pomeni, da n izmerkov $\{x_1, x_2, \dots, x_n\}$ razvrstimo v k med seboj izključujočih se razredov ali predalčkov (angl. *bins*). V i -tem razredu pristane $y_i = n_i$ izmerkov. Primer prikazuje slika 9.5 (levo): v prvi razred $x \in [0.0, 0.1]$ je prišlo $n_1 = 113$ vrednosti, v drugem razredu $x \in [0.1, 0.2]$ jih je $n_2 = 147$, in tako dalje.



Slika 9.5 — [LEVO] Histogram $n = 1000$ dogodkov, zbranih v $k = 10$ razredov. [DESNO] Enaki dogodki, vendar s podanimi negotovostmi v spremenljivki x , ki se mestoma tudi prekrivajo (razredi se ne izključujejo). Prikaza nista enakovredna!

Verjetnost, da določena vrednost pristane v i -tem razredu, označimo s $p_i(\theta)$. Tu je θ nabor parametrov, ki določajo modelsko verjetnostno porazdelitev izmerkov. Pričakovano število izmerkov v i -tem razredu je torej $E[n_i] = np_i(\theta)$. Histogram vsebuje vse izmerke, zato velja

$$\sum_{i=1}^k n_i = \sum_{i=1}^k np_i(\theta) = n. \quad (9.24)$$

V razdelku 5.2 smo pokazali, da je porazdelitev števil n_i v k razredov multinomska, s kovariančno matriko

$$\Sigma = \begin{pmatrix} np_1(1-p_1) & -np_1p_2 & \dots & -np_1p_k \\ -np_2p_1 & np_2(1-p_2) & \dots & -np_2p_k \\ \dots & \dots & \dots & \dots \\ -np_kp_1 & -np_kp_2 & \dots & np_k(1-p_k) \end{pmatrix}$$

Funkcija verjetja je podana z Multinomsko funkcijo

$$L(\text{podatki} \mid \vec{\theta}) = P(X_1 = n_1, \dots, X_k = n_k) = \binom{n}{n_1, \dots, n_k} p_1^{n_1} p_2^{n_2} \cdot \dots \cdot p_k^{n_k}$$

V primeru, ko je v vsakem od N razredov veliko število izmerkov ($n_i \gg 1, i=1, \dots, N$), pa multinomska porazdelitev po centralnem limitnem primeru preide v multivariantno Normalno porazdelitev z zgornjo kovariančno matriko. Določanje optimalnih parametrov v tem primeru lahko torej izvedemo z minimizacijo χ^2 .

Zaradi normalizacijskega pogoja (9.24) je matrika Σ singularna ($\det \Sigma = 0$). Vendar lahko ravno zaradi tega pogoja nalogo najmanjših kvadratov vseeno formuliramo. En razred — na primer k -tega — lahko namreč izločimo, saj je $n_k = n - n_1 - n_2 - \dots - n_{k-1}$. Tako dobimo nesingularno matriko Σ' dimenzije $(k-1) \times (k-1)$, s katero iščemo minimum mere odstopanja

$$\chi^2(\vec{\theta}) = [\vec{n} - n\vec{p}(\vec{\theta})]^T \frac{1}{\Sigma'} [\vec{n} - n\vec{p}(\vec{\theta})], \quad \vec{n} = (n_1, \dots, n_{k-1})$$

Knjiga obravnava tudi kako se ta primer nadalje rešuje (mi tega ne bomo nadalje obravnavali). V knjigi je število razredov označeno z N (tu je označeno z k)

10. Preizkušanje hipotez in statistični testi

V poglavjih 7, 8 in 9 smo opisali metode, s katerimi na podlagi naključnih vzorcev sklepamo na lastnosti populacij in ocenimo parametre njihovih porazdelitev. V tem poglavju spoznamo orodja za presojo o tem, ali je s statističnega vidika dobljeni model ali hipoteza o populaciji s statističnega vidika sprejemljiva ali ne. O tem presojamo glede na izmerjene podatke v vzorcu.

10.1 Ničelna hipoteza H_0 in alternativna hipoteza H_1

Za presojo o veljavnosti modela uporabljamo hipoteze o lastnostih populacije oziroma njene verjetnostne porazdelitve.

Osnovno hipotezo, ki jo preverjamo, imenujemo ničelna (angl. null hypothesis) in jo označimo s H_0 . **Glede na izid statističnega testa lahko ničelno hipotezo sprejmemo ali zavrremo** — vendar je namesto “hipotezo sprejmemo” bolje reči “s statističnega vidika nimamo dovolj podatkov v vzorcu, da bi jo zavrnil”; v nadaljevanju obdržimo to subtilno razliko v mislih. **Glavna naloga statističnega testa je torej presoja o tem ali glede na dane vzorčne podake lahko hipotezo H_0 zavržemo ali ne.**

Strogo rečeno nikoli ne presojamo o ničelni hipotezi sami po sebi, temveč vselej glede na alternativno hipotezo (angl. alternative hypothesis), ki jo označimo s H_1 .

Alternativne hipoteze so lahko v različnih povezavah z ničelnimi hipotezami. A bomo tu zaradi preprostosti imeli v mislih le H_1 , ki se izključuje s H_0 , in kjer unija H_0 in H_1 vsebujeta vse možnosti. Torej bo v naših primerih H_1 komplement H_0 .

Vrednosti, ki jih testiramo pri hipotezah, se seveda nanašajo na celotno populacijo možnih meritev (in ne le na merjen vzorec)

Primeri:

a) vzorec: večkratni met določene kocke; populacija: neskončno metov te kocke

H_0 : kocka je poštena

H_1 : kocka ni poštena

b) vzorec: večkratna meritev temperature sredstva (ob predpostavki, da se temperatura sredstva ne spreminja); populacija: dejanska temperatura po neskončno meritvah

H_0 : $T=15C$ (temperatura sredstva) +

H_1 : $T \neq 15C$

c) **H_0 : $T \geq 15C$ (temperatura)**

H_1 : $T < 15C$

d) vzorec: vrednost določenega parametra pri krvnem testu

H_0 : pacient je zdrav (ni okužen z danim virusom)

H_1 : pacient je bolan (je okužen z danim virusom)

e) vzorec: večkratna meritev kvantnomehanske opazljivke O ; populacija: neskončno meritev te opazljivke bi dalo pričakovano vrednost μ_0

H0: $\mu = \mu_0$: opazljivka O ima vrednost $\mu = \mu_0$, kjer je μ_0 vrednost kot jo napoveduje obstoječa veljavna teorija (na primer Standardni Model osnovnih delcev). Glavna naloga testa je presoja o tem ali lahko z danimi podatki veljavno teorijo (torej H_0) ovrzemo s statističnega vidika – to je ena glavnih nalog eksperimentov.

H1: $\mu \neq \mu_0$: opazljivka O ima vrednost $\mu \neq \mu_0$, ki ni v skladu z napovedjo obstoječe veljavne teorije. Šele če smo iz statističnega vidika utemeljeno zavrgli veljavno teorijo, lahko sklepamo na upravičenost ene od novih teorij.

f) vzorec: večkratna meritev hitrosti (v) neznanega delca v vakuumu

H0: $v \leq c_0$ (hitrost neznanega delca v vakuumu, kjer je c_0 svetlobna hitrost)

Tu je seveda ključno vprašanje kako zanesljivo lahko iz statističnega (in sistematičnega!) vidika zavrnemo to dobro utemeljeno hipotezo. Krsitev le te bi bila seveda nepričakovana znanstvena revolucija.

H1: $v > c_0$

g) vzorec: meritev števila dogodkov N pri nekem procesu in energiji E v trkalniku, kjer iščemo signal za nov delec, npr $p \rightarrow p + \gamma$

H0: $N = N(b)$ število dogodkov je skladno z ozadjem ($b = \text{ozadje} = \text{background}$)

$N(b) = \text{napoved za število dogodkov v okviru teorije, ki ne vključuje Higgsovega bozona}$. Glavno vprašanje je ali lahko statistično signifikantno to ničelno hipotezo zavrnemo.

H1: $N = N(b+s)$: število dogodkov je skladno z ozadjem (b) + signalom (s); po tej hipotezi dotlej veljavna teorija brez Higgsovega bozona ni pravilna; prava teorija mora vsebovati Higgsov bozon ali kako drugo "novo fiziko"; šele ko z dovolj veliko zanesljivostjo zavrnemo H_0 lahko utemeljeno presojamo o novih delcih in pojavih

10.2 Stopnja zaupanja statističnega testa (confidence level) $CL = 1 - \alpha$

Osredotočimo se na vprašanje ali lahko na podlagi podatkov zavrnemo ničelno hipotezo H_0 ali je ne moremo zavreči.

Naj α označuje verjetnost, da zavrnemo H_0 v primeru, ko bi bila H_0 resnična.

- $\alpha = \text{statistična pomembnost testa (statistical significance)}$
- $1 - \alpha = \text{stopnja zaupanja} = \text{confidence level} = CL$

Vrednost α si vnaprej izberemo in potem presojamo o tem ali pri tej statistični pomembnosti testa H_0 zavrnemo ali ne.

Bolj podrobno:

- v primerih, ko H_0 predstavlja določeno možnost (npr $T = 15^\circ\text{C}$, $\mu = \mu_0$, $N = b$): α označuje verjetnost, da zavrnemo H_0 v primeru, ko bi bila H_0 resnična.
- v primerih, ko H_0 predstavlja več možnosti (npr območje $T \geq 15^\circ\text{C}$, $v \leq c_0$): α označuje verjetnost, da ob predpostavki veljavne H_0 zavrnemo H_0 v najbolj ekstremnem primeru (na primer zavrnemo $T = 15^\circ\text{C}$ kar pomeni da so $T > 15^\circ\text{C}$ še manj verjetne od $T = 15^\circ\text{C}$: ogledali si bomo zglede).

Opomba: Spomnimo se, da je pri intervalih zaupanja α označeval verjetnost, da je prava populacijska vrednost neke količine (npr populacijsko povprečje, varianca ali parameter) izven intervala zaupanja. Torej sta definiciji α tu in tam kar se da skladni.

- Pogosto pri testih določimo še "vrednost-p" ("p-value"). **"vrednost-p" je verjetnost, da je vrednost naključne spremenljivke tako ekstremna ali bolj kot je bila ob meritvi, kjer te verjetnosti določimo ob predpostavki da je H_0 pravilna.** Vrednosti-p, ki je manjša od izbranega α botruje zavrnitvi hipoteze H_0 . Pogosto uporabljeni količini α in p-value sta tako tesno povezani. Opomba: p-value ne smemo mesati z indeksom p ki je uporabljen pri t_p in z_p spodaj, res pa oba p predstavljata določeni verjetnosti).
- **Dobljeno vrednost p-value se pogosto pretvori v ekvivalentno številsko vrednost standardnih deviacij normalne porazdelitve $N(0,1)$.** Na primer, dobljeni vrednosti 3.5 standardnih deviacij pri zavrnitvi H_0 , pravimo zavrnitev s signifikanco 3.5 gaussovih sigma (ker se sigma nanasa na gausovo porazdelitev). Namen slednje pretvorbe je lažja primerjava signifikanc pri različnih porazdelitvah. **Odkritje novega delca, na primer, lahko trdimo, če nicelno hipotezo H_0 (kjer novega delca ni) zavrnemo s signifikanco 5 gaussovih sigma**, kar ustreza $p\text{-value}=1-F[5]=2.8 \cdot 10^{-7}$ ali dvakrat toliko (odvisno od tega ali je območje zavrnite na enem ali obeh koncih porazdelitve – s to podrobnostjo si ne bomo preveč belili glave)

10.3 Test za vrednost populacijskega povprečja v primeru normalno porazdeljenih spremenljivk (10.2.1 v knjigi)

Posvetimo se hipotezam ki govorijo o vrednosti populacijskega povprečja μ_0

Če poznamo vzorčno povprečje $\bar{X}$ in populacijsko varianco σ , bomo presojali po statistiki spremenljivke Z , ki je porazdeljena normalno $N(0,1)$

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} : f_Z(Z) = N(0,1)$$

Če poznamo vzorčno povprečje $\bar{X}$ in nepristransko vzorčno varianco s_X , bomo presojali po statistiki spremenljivke T , ki je porazdeljena po Studentovi porazdelitvi

$$T = \frac{\bar{X} - \mu_0}{s_X/\sqrt{n}} : f_T(t; \nu = n-1)$$

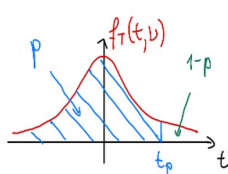
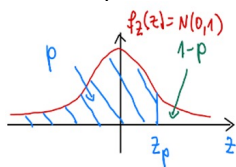
Opomba: za primere, ko je n dovolj velik (npr $n \geq 30$), je Studentova porazdelitev praktično enaka $N(0,1)$ in lahko uporabimo tudi normalno.

Dogovorimo se, da z_p označuje vrednost, pod katero se naključna spremenljivka z nahaja z verjetnostjo p , torej p -ti kvantil porazdelitve:

$$P(z \leq z_p) = p \text{ oziroma drugimi besedami } z_p = F^{-1}(p) = \text{InverseCDF}[\text{NormalDistribution}[0,1], p]$$

Podobno naj t_p označuje vrednost pod katero se t nahaja z verjetnostjo p , torej p -ti kvantil:

$$P(t \leq t_p) = p \text{ oziroma } t_p = F^{-1}(p) = \text{InverseCDF}[\text{StudentTDistribution}[n-1], p]$$



glej tabele
kvantilov na uclnici

Zgled: Z dvanajstimi enakimi termometri ($n=12$) izmerimo temperature

$x = \{13.6, 13.8, 13.9, 13.3, 13.9, 14.7, 13.7, 13.5, 14.5, 12.7, 12.2, 16.3\}$

Glede na razprsenost rezultatov so termometri precej slabe kakovosti ...

Preden se lotimo hipotez, določimo

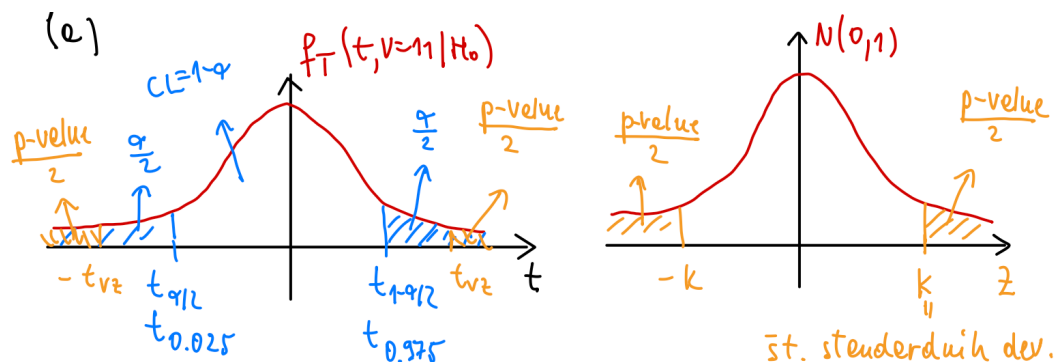
$\bar{x} = 13.84$ C (vzorcno povprečje), $sX = 1.03$ (vzorcna varianca), $sX/\sqrt{n} = 0.30$ (var. vz. povprečja)

Populacijsko povprečje mu torej pričakujemo v grobem na intervalu 13.84 ± 0.30

Zdaj testirajmo razne hipoteze pri stopnji zaupanja $= 1 - \alpha = CL = 0.95$,

a) $H_0: \mu = \mu_0 = 13$ C

$H_1: \mu \neq \mu_0$



Ker poznamo vzorcno povprečje, uporabimo za test hipoteze H_0 Studentovo porazdelitev z $\nu = 12 - 1 = 11$ prostostnimi stopnjami. O zavrnitvi ali sprejetju hipoteze H_0 sklepamo ob predpostavki, da je ta pravilna, torej da je temperatura dejansko μ_0 .

Tedaj z verjetnostjo α zavrnilo H_0 če se naključna spremenljivka t_{vz} za merjen vzorec nahaja izven območja $[t_{\alpha/2}, t_{1-\alpha/2}] = [t_{0.025}, t_{0.975}] = [-2.2, 2.2]$. Dejanska izmerjena temperatura vzorca da vrednost $t_{vz} = (\bar{x} - \mu_0) / (sX / \sqrt{n}) = (13.84 - 13) / (1.03 / \sqrt{12}) = 2.8$ kar je izven zgornjega območja, zato hipotezo H_0 zavrnilo ob stopnji zaupanja 0.95. p-vrednost je verjetnost da je t bolj ekstremna kot izmerjena

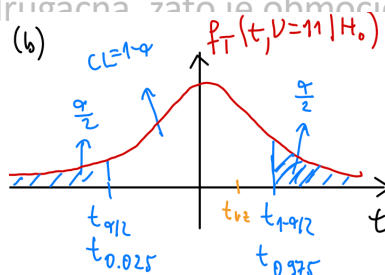
vrednost, torej $\text{vrednost-p} = 2 \int_{t_{vz}}^{\infty} f_T(t, \nu) dt = 0.016$, $\text{vrednost-p} = 2 \int_k^{\infty} N(0, 1) dz \rightarrow k = 2.4$

Ker je p-value pod privzeto vrednostjo $\alpha = 0.05$ tudi ta sklep vodi v zavrnitev H_0 ; oba sklepa sta pravzaprav enakovredna. Ta vrednost-p ustreza signifikanci 2.4 gausovih sigma. Pravimo, da smo H_0 zavrnilo z signifikanco 2.4 gausovih sigma ($k = 2.4$).

Opomba: V knjigi na str 230 je H_1 nekoliko drugačna, zato je območje zavrnitve drugačno.

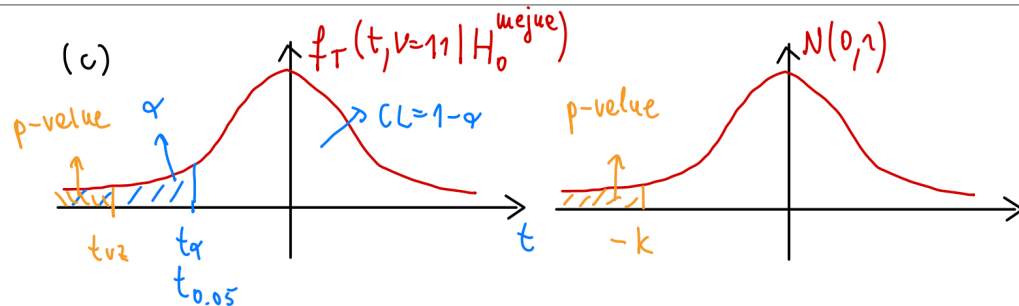
b) $H_0: \mu = \mu_0 = 13.6$ C

$H_1: \mu \neq \mu_0$



Ista vzorcna temperatura da pri tej ničelni hipotezi $t_{vz} = (\bar{x} - \mu_0) / (sX / \sqrt{n}) = (13.84 - 13.6) / (1.03 / \sqrt{12}) = 0.8$ kar je tokrat znotraj območja $[t_{\alpha/2}, t_{1-\alpha/2}] = [-2.2, 2.2]$ in hipotezo H_0 ob stopnji zaupanja $CL = 0.96$ sprejmemo, kar pravzaprav ne pomeni da je temperatura res 13.6C temveč da dane meritve niso dovolj natančne da bi hipotezo H_0 zavrnilo.

c) $H_0: \mu \geq \mu_0 = 14.5 \text{ C}$
 $H_1: \mu < \mu_0$



Ce H_0 pri dani stopnji zapanja zavržemo za najnižjo temperaturo μ_{00} iz območja pri H_0 , potem bo avtomatično H_0 zavržena tudi pri visjih temperaturah, saj je izmerjena povprečna temperatura $\bar{x}=13.84$ nižja od mejne temperature 14.5C . Testirajmo torej ali H_0 zavzamemo pri mejni temperaturi, kar se zgodi če bo $\bar{x}$ mnogo manjši od μ_0 , oziroma če bo t_{vz} manjši od mejnega $t_{\alpha} = -t_{1-\alpha} = -1.8$. Vrednost naključne spremenljivke t_{izm} ob predpostavki da velja H_0 z mejno temperaturo je $t_{\text{vz}} = (\bar{x} - 14.5) / (sX / \sqrt{n}) = -2.2$, kar je manjše od -1.8 zato H_0 zavzamemo. To je v skladu z definicijo, da H_0 zavzamemo pri mejni temperaturi z verjetnostjo $\alpha = 0.05$. Vrednost-p, da je meritev temperature tako nizka kot $\bar{x}$ ali se nižja pri mejnem primeru H_0 je vrednost-p = 0.025 kar je manjše od α , tako da tudi iz tega stališča zavržemo H_0 . Pravimo, da smo H_0 zavržili z signifikanco 1.9 gausovih sigma ($k=1.9$).

$$\text{vrednost-p} = \int_{-\infty}^{t_{\text{vz}}} f_T(t, \nu) dt = 0.025, \quad \text{vrednost-p} = \int_{-\infty}^{-k} N(0, 1) dz \rightarrow k = 1.9$$

Bolj preprosto povedano: T visja ali enaka 14.5C dejansko ni zelo verjetna ob dani meritvi povprečne temperature 13.84C z varianco povprečja 0.3 C . Je pa temperatura 14.5 vseeno lahko prava temperatura "populacije" z verjetnostjo p-value = 0.025, a je bila izmerjena povprečna temperatura 13.84C tako nizka ker je bila meritev izpostavljena statistični fluktuaciji.

Zgled: rezultat za vrednost-p pri dejanskem eksperimentu ki poroča o pomembnem odkritju

Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC ☆

<https://doi.org/10.1016/j.physletb.2012.08.020>

ATLAS Collaboration* Spodaj so rezultati kolaboracije ATLAS (med soavtorji so tudi slovenski fiziki)

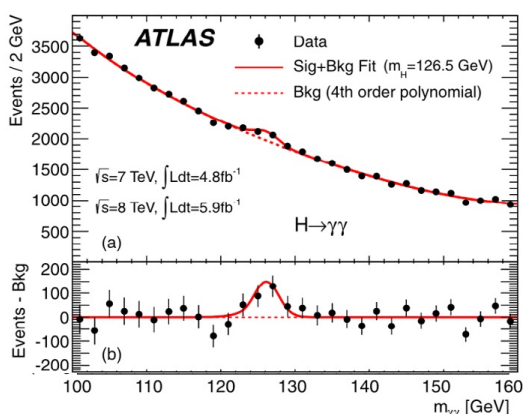
Physics Letters B 716 (2012)

Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC ☆

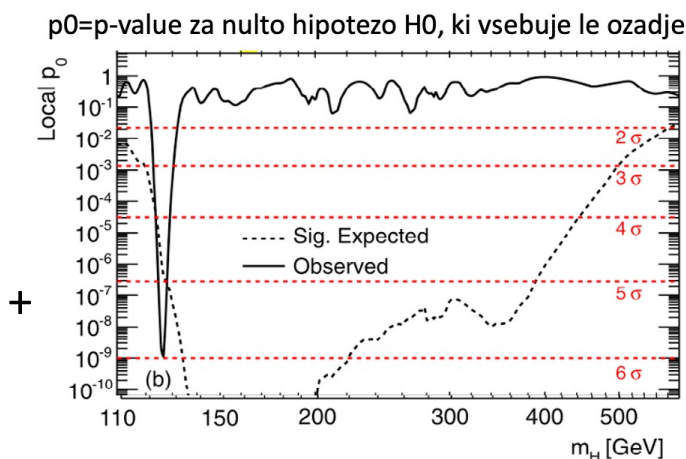
<https://doi.org/10.1016/j.physletb.2012.08.021>

CMS Collaboration*

$$\text{vrednost } -p = 1.7 \times 10^{-9} = \int_{5.9}^{\infty} N(0, 1) dz$$



$$p p \rightarrow \gamma \gamma$$



+

ABSTRACT

A search for the Standard Model Higgs boson in proton–proton collisions with the ATLAS detector at the LHC is presented. The datasets used correspond to integrated luminosities of approximately 4.8 fb^{-1} collected at $\sqrt{s} = 7 \text{ TeV}$ in 2011 and 5.8 fb^{-1} at $\sqrt{s} = 8 \text{ TeV}$ in 2012. Individual searches in the channels $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$, $H \rightarrow \gamma\gamma$ and $H \rightarrow WW^{(*)} \rightarrow e\nu\mu\nu$ in the 8 TeV data are combined with previously published results of searches for $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$ and $H \rightarrow \gamma\gamma$ channels in the 7 TeV data and results from improved analyses of the $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$ and $H \rightarrow \gamma\gamma$ channels in the 7 TeV data. Clear evidence for the production of a neutral boson with a measured mass of $126.0 \pm 0.4 \text{ (stat)} \pm 0.4 \text{ (sys)} \text{ GeV}$ is presented. This observation, which has a significance of 5.9 standard deviations, corresponding to a background fluctuation probability of 1.7×10^{-9} , is compatible with the production and decay of the Standard Model Higgs boson.

pri določeni energiji E:

H0: $N=N(b)$ stevilo dogodkov skladno z ozadjem
(SM model brez Higgsa)

H1: $N= N(b + s)$; $s+b$ =ozadje+Higgsov signal

stevilo dogodkov skladno s pričakovanji SM modela
ki vsebuje Higgsov bozon

Ob hipotezi brez Higgsovega bozona je verjetnost za opazeno stevilo dogodkov ali se večje stevilo dogodkov enaka vrednosti $-p=1.7 \times 10^{-9}$, kar je izjemno malo. Hipotezo H0 so zavrgli s signifikanco 5.9 gaussovskih sigma.

10.4 Test za vrednost variance v primeru normalno porazdeljenih spremenljivk (10.2.2 v knjigi)

Testirajmo komplementarni hipotezi za vrednost populacijske variance σ^2 , na primer

$$\begin{array}{ll} H_0: \sigma^2 = \sigma_0^2 & \text{ali} & H_0: \sigma^2 \leq \sigma_0^2 \\ H_1: \sigma^2 \neq \sigma_0^2 & & H_1: \sigma^2 > \sigma_0^2 \end{array}$$

Pri testiranju variance na podlagi vzorca $x = \{x_1, x_2, \dots, x_n\}$ iz normalno porazdeljene populacije $X \sim N(\mu, \sigma^2)$ spet ločimo dva primera: da je pravo populacijsko povprečje μ znano ali ne.

- Če pravega populacijskega povprečja μ ne poznamo, hipotezo H_0 preverjamo s statistiko χ^2 , ki je porazdeljen po porazdelitvi χ^2 z $n-1$ prostostinimi stopnjami

$$\chi^2 = \frac{n \tilde{s}_X^2}{\sigma_0^2} = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma_0^2} : f_{\chi^2}(\chi^2, \nu = n-1) \quad \bar{X} = (1/n) \sum_{i=1}^n X_i$$

- Če populacijski μ poznamo, hipotezo H_0 preverjamo s statistiko χ^2 , ki je porazdeljen po porazdelitvi χ^2 z n prostostinimi stopnjami

$$\chi^2 = \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma_0^2} : f_{\chi^2}(\chi^2, \nu = n)$$

Zgled Za detektor potrebujemo veliko število žičnih elektrod določene dolžine. Največja dovoljena varianca dolžin je $\sigma_0^2 = 100 (\mu\text{m})^2$. Natančna meritev dolžine je zelo zahtevna, zato si lahko privoščimo le majhen vzorec $n = 10$ elektrod, v katerem ugotovimo varianco $s_X^2 = 142 (\mu\text{m})^2$. S pomembnostjo $\alpha = 0.05$ presodimo, ali dolžina večine žic v neraziskani "populaciji" čezmerno fluktuiral!

Ničelna hipoteza je $H_0 : \sigma^2 \leq \sigma_0^2$, alternativna hipoteza je $H_1 : \sigma^2 > \sigma_0^2$. Hipotezo H_0 bi zavrgli, če bi vzorčna varianca presegla kritično varianco glede na α . Vrednost testne statistike je $\chi^2 = n s_X^2 / \sigma_0^2 = 10 \cdot 142 / 100 = 14.2$. Kritični χ^2 odčitamo iz tabele D.3, vrstica za $\nu = n - 1 = 9$, stolpec za $p = 1 - \alpha = 0.95$: enak je $\chi_*^2 = \chi_{0.95}^2 = 16.9$. Ker velja $\chi^2 < \chi_*^2$, ob stopnji zaupanja $1 - \alpha = 95\%$ nimamo razloga, da bi H_0 zavrgli. Sklepamo lahko, da je varianca dolžin vseh elektrod v okviru predpisane.

$$\begin{array}{l} H_0: \sigma^2 \leq \sigma_0^2 \\ H_1: \sigma^2 > \sigma_0^2 \end{array}$$



10.5 Test H_0 proti H_1 v primeru danih splošnih porazdelitev $f_x(x|H_0)$ in $f_x(x|H_1)$

(10.1 v knjigi)

Zdaj obravnavajmo primer, ko sta verjetnostni porazdelitvi relevantne naključne spremenljivke x , ki jo merimo, znani in točno določeni za ničelno hipotezo H_0 in alternativno hipotezo H_1 . Zopet imejmo v mislih primere ko H_0 in H_1 izcrpata vse možnosti. Da bomo bolj nazorni imejmo v mislih konkreten primer

H_0 : pacient je zdrav (ni okužen z določenim virusom); vrednosti opazljivke x pri krvnem testu x so porazdeljene z verjetnostno gostoto $f(x; \theta_0)$

H_1 : pacient je bolan (je okužen z določenim virusom); vrednosti opazljivke x pri krvnem testu x so porazdeljene z verjetnostno gostoto $f(x; \theta_1)$

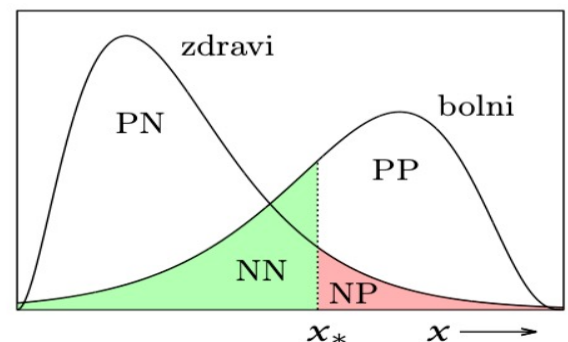
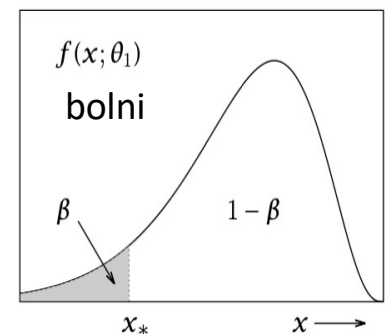
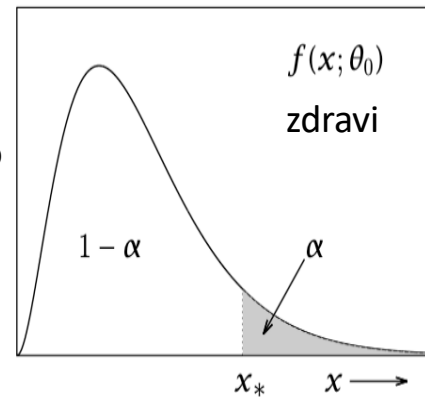
Obe verjetnostni porazdelitvi tu torej pripadata točno določenima vrednostima parametra θ , v tem primeru θ_0 in θ_1 (za razliko od nekaterih prejsnjih primerov ko sta H_0 in H_1 pripadali razponu vrednosti parametrov, npr $H_0: \mu \geq \mu_0$; $H_1: \mu < \mu_0$).

Kako pri danem α izberemo mejno vrednost x_* krvnega testa za presojo o tem ali je pacient zdrav ali bolan že vemo: Ob predpostavki, da je pa zdrav, mora biti verjetnost, da smo ga razglasili za bolnega, enaka α . Temu ustreza zgornji graf, kjer pacienta s krvnim testom x pod mejno vrednostjo x_* razglasimo za zdravega. Z verjetnostjo α smo seveda s tem zgrešili napačno presojo, zato želimo čim manjši α .

Ker tokrat alternativni hipotezi H_1 pripada točno določena verjetnostna porazdelitev za naključno spremenljivko x , določimo verjetnost, s katero opredelimo bolnega pacienca kot bolnega

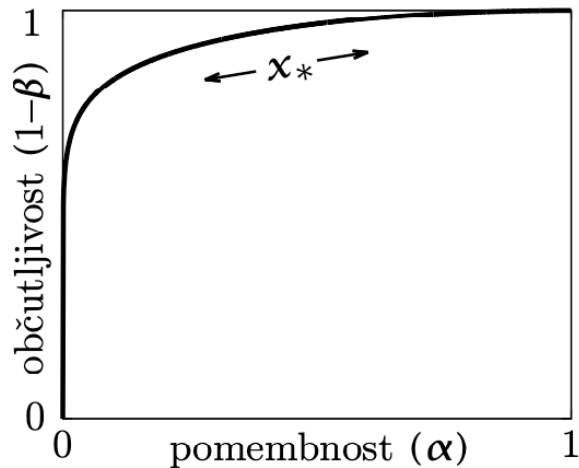
$$1 - \beta = \int_{x_*}^{\infty} f(x; \theta_1) dx, \quad \text{občutljivost testa (sensitivity)}$$

zato seveda želimo, da je $1 - \beta$ čim večji. To verjetnost moramo določiti seveda po porazdelitvi, ki ustreza H_1 .



Verjetnostna gostota, ki ustreza ničelni hipotezi s parametrom θ_0 (zgoraj) in alternativni hipotezi s parametrom θ_1 (sredina). Senčenje označuje območji zavrnitve pri stopnjah pomembnosti α oziroma β . Kritična točka je označena z x_* .
spodaj: Primer: Porazdelitev po izidih krvnega testa v zdravi in bolni populaciji. Možni so pravilni pozitivni (PP), napačni negativni (NN), pravilni negativni (PN) in napačni pozitivni (NP) izidi.

Želimo torej čim večji α in čim večji $1-\beta$. Oba odvisna od mejnega x^* , torej sta soodvisna. Tipična soodvisnost je prikazana na spodnji sliki, ki kaže tako imenovano krivuljo ROC (Receiver Operating Characteristics). V prej prikazanem primeru se α in $1-\beta$ manjšata z večanjem x^* . Čim bolj se krivulja približa zgornjemu levemu kotu, tem večjo napovedno moč ima test; v skrajnem primeru (točka $(\alpha, 1 - \beta) = (0, 1)$) sta populaciji povsem ločeni.



Dodatek k poglavju o intervalih zaupanja:

Zgornje in spodnje meje na količine, ko signala ne zaznamo

Če signala ne zaznamo, ali pa je merjena količina precej blizu napovedi za ozadje, običajno postavimo zgornjo ali spodnjo mejo na določeno opazljivko.

Zgled

Oglejmo si kar preprost zgled kjer merimo število radioaktivnih jeder, a v času meritve T ne zaznamo nobenega razpada, zato želimo določiti spodnjo mejo za razpadni čas τ pri $CL=1-\alpha=95\%$ stopnji zaupanja. Obravnavali bomo primer brez ozadja in z ozadjem. Zaradi preprostosti imejmo v mislih, da je čas meritve T mnogo manjši od razpadnega časa τ .

(a) $N_0=1000$ (zacetno število jeder)

$T=100$ s (čas opazovanja),

$n_{vz}=0$ (opazeno število razpadnih produktov v vzorcu)

predpostavimo da ni ozadja

$$P(n|\bar{N}) = \frac{1}{n!} \bar{N}^n e^{-\bar{N}}$$

Verjetnost je porazdeljena po poissonovi porazdelitvi

Ta podaja verjetnost za n razpadov ob pogoju, da je povprečno število razpadov v tem času $\bar{N}$, kjer je povprečno število razpadov edini parameter porazdelitve. Nasa prva naloga je torej da določimo mejno vrednost parametra $\bar{N}_*$, pri katerem bo verjetnost, da opazimo n_{vz} ali se manj dogodkov (torej tako ekstremno malo dogodkov ali se manj) enaka $\alpha=0.05$:

+

$$P(n \leq n_{vz} | \bar{N}_*) = \alpha$$

V našem primeru je v vzorcu n_{vz} tako ali tako 0, tako da $P(n=0 | \bar{N}_*) = \alpha=0.05 = \exp(-\bar{N}_*) \rightarrow \bar{N}_* = 3.0$. Ker nismo opazili nobenega dogodka, zaključimo da je dejansko pričakovano populacijsko povprečje $\bar{N}$ razpadov v tem času manjše od 3.0, torej smo ob 95% stopnji zaupanja dobili interval zaupanja za opazljivki $\bar{N}$ in τ

interval zaupanja $\bar{N} \leq \bar{N}_* = 3.0 \rightarrow \tau \geq \tau_* = 9.3$ h pri $CL=95\%$

Pri tem smo uporabili, da je povprečno število razpadov $\bar{N}$ za $T \ll \tau$ povezano s τ preko zpodnjih zvez, kjer $N_{pr}(t)$ označuje preostalo število nerazpadlih jeder:

$$N_{pr}(t) = N_0 e^{-t/\tau}, \quad dN_{pr} = -N_0 e^{-t/\tau} dt/\tau \simeq -N_0 dt/\tau$$

$$dt \rightarrow T, \quad dN_{pr} \rightarrow -\bar{N}$$

$$\bar{N} = \frac{N_0 T}{\tau} \rightarrow \tau_* = \frac{N_0 T}{\bar{N}_*} = \frac{1000 \cdot 100s}{3.0} = 9.3h$$

(b) $N_0=1000$ (začetno število jeder)

$T=100$ s (čas opazovanja),

$n_{vz}=0$ (opazeno število razpadnih produktov v vzorcu)

. $\bar{N}_b = 0.5$: pri tej meritvi pričakujemo v povprečju 0.5 dogodkov ozadja –
to je zaznanih produktov, ki niso posledica razpada temveč suma, kozmičnih zarkov,

Tudi za ozadje pričakujemo Poissonovo porazdelitev števila dogodkov n_b okoli povprečne vrednosti $\bar{N}_b$: $P(n_b | \bar{N}_b)$. Konvolucija dveh Poissonovih porazdelitev je Poissonova porazdelitev zato pričakujemo n dogodkov z verjetnostjo $P(n | \bar{N} + \bar{N}_b)$ kjer $\bar{N}$ se vedno označuje povprečno število dejanskih razpadov. Po vzoru prejšnjega zgleda zdaj dobimo meje

$$P(n \leq n_{vz} | \bar{N}_* + \bar{N}_b) = \alpha \rightarrow \bar{N}_* + \bar{N}_b = 3.0 \rightarrow \bar{N}_* = 3.0 - \bar{N}_b = 3.0 - 0.5$$

interval zaupanja $\bar{N} \leq \bar{N}_* = 2.5 \rightarrow \tau \geq \tau_* = 11.1$ h pri CL=95%

(c) $N_0=1000$ (začetno število jeder) +

$T=100$ s (čas opazovanja),

$n_{vz}=1$ (opazeno število razpadnih produktov v vzorcu zdaj ni nič a ni dosti večje kot pričakujemo za ozadje)

. $\bar{N}_b = 0.5$: pri tej meritvi pričakujemo v povprečju 0.5 dogodkov ozadja –
to je zaznanih produktov, ki niso posledica razpada temveč suma, kozmičnih zarkov,

$$P(n \leq n_{vz} | \bar{N}_* + \bar{N}_b) = \alpha \rightarrow P(n=0,1 | \bar{N}_* + \bar{N}_b) = \alpha \rightarrow \bar{N}_* + \bar{N}_b = 4.7 \rightarrow \bar{N}_* = 4.2$$

interval zaupanja $\bar{N} \leq \bar{N}_* = 4.2 \rightarrow \tau \geq \tau_* = 6.6$ h pri CL=95%

Zgornji intervali zaupanja povedo da je z verjetnostjo CL=0.95 dejanski razpadni čas večji od danes spodnje meje τ_* , s čimer imamo v mislih: pri mnogokratni meritvi vzorcev bo v 95% vzorcev dejanski razpadni čas nad vzorčno mejo τ_* . Pri tem je dejanski razpadni čas od vzorca do vzorca nespremenjen, spodnja meja τ_* pa se od vzorca do vzorca spreminja.

Zgled: rezultat za zgornjo mejo na relevanten parameter (moc signala μ) pri dejanskem eksperimentu, ki poroča o pomembnem odkritju

Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC [☆]

<https://doi.org/10.1016/j.physletb.2012.08.020>

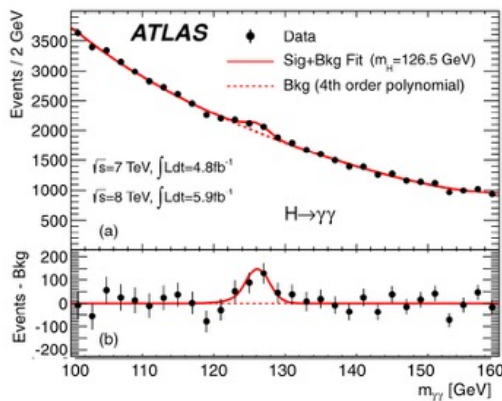
ATLAS Collaboration* Spodaj so rezultati kolaboracije ATLAS (med soavtorji so tudi slovenski fiziki)

Physics Letters B 716 (2012)

Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC [☆]

<https://doi.org/10.1016/j.physletb.2012.08.021>

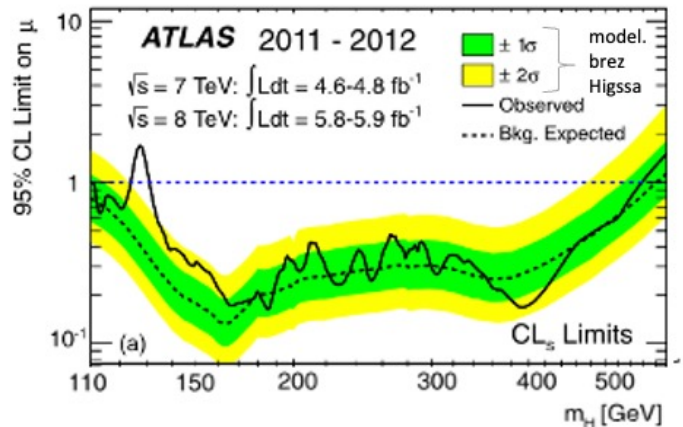
CMS Collaboration*



$$p p \rightarrow \gamma\gamma$$

ABSTRACT

A search for the Standard Model Higgs boson in proton–proton collisions with the ATLAS detector at the LHC is presented. The datasets used correspond to integrated luminosities of approximately 4.8 fb^{-1} collected at $\sqrt{s} = 7$ TeV in 2011 and 5.8 fb^{-1} at $\sqrt{s} = 8$ TeV in 2012. Individual searches in the channels $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$, $H \rightarrow \gamma\gamma$ and $H \rightarrow WW^{(*)} \rightarrow e\mu\mu$ in the 8 TeV data are combined with previously published results of searches for $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$, $WW^{(*)}$, $b\bar{b}$ and $\tau^+\tau^-$ in the 7 TeV data and results from improved analyses of the $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$ and $H \rightarrow \gamma\gamma$ channels in the 7 TeV data. Clear evidence for the production of a neutral boson with a measured mass of 126.0 ± 0.4 (stat) ± 0.4 (sys) GeV is presented. This observation, which has a significance of 5.9 standard deviations, corresponding to a background fluctuation probability of 1.7×10^{-8} , is compatible with the production and decay of the Standard Model Higgs boson.



polna crna crta:

zgornja meja na moc signala μ pri stopnji zaupnja 95%

definicija moci signala μ $N_{exp} = N_b + \mu N_s$

$\mu = 1$: SM Higgs

$\mu = 0$: samo ozadje

dogodki
ozadja

signalni dogodki
zaradi Higgsa v SM

13. Metoda Monte Carlo

Metoda oziroma simulacija Monte Carlo (MC) je generično ime za kakeršen koli postopek, pri katerem s pomočjo žrebanja naključnih števil in statističnih vzorcev približno iz vrednotimo neko matematično količino ali izraz, na primer določeni integral ali sistem enačb, z njim pa lahko rešujemo tudi bistveno splošnejše matematičnofizikalne probleme [1]. Poudarek je na besedi 'približno': kakovost rešitve je odvisna od velikosti vzorcev, ki si jih pri računu lahko privoščimo. Vendar metoda Monte Carlo z vidika obvladljivosti in natančnosti računa — zlasti pri integraciji v več dimenzijah z zapletenimi integracijskimi mejami in pri kompleksnih matematičnih modelih — ponuja v primerjavi s standardnimi numeričnimi metodami edino razumno pot.

Numerična integracija v več kot $D=10$ dimenzionalnem prostoru $\vec{x}$ z diskretizacijo in običajnim sestevanjem po vseh D dimenzijah praktično ne pride v poštev v svoji najpreprostejši obliki spodaj, pa tudi njene bolj dodelane različice v ta namen nikakor niso smotrne

$$\int g(\vec{x}) d\vec{x} \rightarrow \sum_{n_1} \dots \sum_{n_D} g(\vec{x}) h^D, \quad \vec{x} = \vec{n} h^D, \quad \vec{x} = (x_1, \dots, x_D), \quad \vec{n} = (n_1, \dots, n_D), \quad \in nN^D$$

Premnogi fizikalni in nefizikalni problemi zahtevajo integracijo v mnogo več dimenzionalnem prostoru kot je na primer $D=10$. Pri statistični termodinamiki seštevamo na primer po vseh konfiguracijah sistema in jih utežimo z Boltzmanovim faktorjem, kar predstavlja neke vrste D -dimenzionalno vsoto ali integral. Fizikalni problemi, ki jih na primer rešujem sama z namenom studija močne interakcije med kvarki, zahtevajo tipično integracijo v okoli $D \sim 10\,000\,000$ dimenzionalnem prostoru (komentar: toliko je namreč točk v diskretiziranem prostoru-času, ki ga simuliramo pri obravnavi kvantne teorije polja na mreži). Daleč najpomembnejši in najbolj uporaben pristop k numerični integraciji v $D > 10$ dimenzionalnih prostorih je Monte Carlo metoda.

13.1. Numerična integracija in pomembnostno vzorčenje (13.2, 13.3 v knjigi)

Metoda Monte Carlo je ključna na najrazličnejših področjih in pri najrazličnejših problemih, a oglejmo si jo na primeru numerične integracije. Kot je navedeno v uvodu, je skoraj neobhodno potrebna pri integraciji v mnogo-dimenzionalnem prostoru, a zaradi preprostosti si oglejmo njeno uporabo na primeru integracije v 1 dimenziji, ki jo lahko sicer opravimo tudi druge preprostejše načine, morda celo analitično.

Izračunati želimo določen integral oblike (poglavje 13.2 za $f(x)=1$)

$$\theta = \int_a^b g(x) dx \qquad \theta = \int_a^b g(x) dx = \int_a^b \frac{g(x)}{p(x)} p(x) dx$$

To z metodo Monte Carlo storimo tako, da ga najprej izrazimo z uporabo neke verjetnostne gostote $p(x)$, ki ji pravimo pomembnostna funkcija (importance function). Ta ima običajne lastnosti verjetnostne gostote $p(x) \geq 0$ ter $\int_a^b p(x) dx = 1$.

Izbira $p(x)$ je načeloma poljubna, a učinkovitost Monte Carlo metode močno zavisi od izbire $p(x)$. Pokazali pa bomo, da so za bolj učinkovito določitev integrala primerni $p(x)$ z večjo verjetnostno gostoto tam kjer je $g(x)$ večja, in $p(x)$ z manjšo verjetnostno gostoto tam kjer je $g(x)$ majhna.

Za približno določitev integrala izzrebramo N vrednosti x_i ($i=1,..,N$) ki so porazdeljene v skladu z verjetnostno gostoto $p(x)$. Kako to napravimo je seveda poglavje zase, a primerne metode so na primer povezane z verigami Markova (poglavje 13.4 in predvsem Metropolis Hastingsov algoritem 13.4.1), nam pa to na primer omogoča tudi funkcija `ProbabilityDistribution[p[x], {x, a, b}]` v Wolframovi Mathematici.

Integral θ ni nič drugega kot pričakovana vrednost funkcije $g(x)/p(x)$ glede na verjetnostno gostoto p

$$\theta = E \left[\frac{g(x)}{p(x)} \right]$$

približno pa jo ocenjuje za vzorec z N elementi x_i cenilka

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^N \frac{g(x_i)}{p(x_i)}$$

Lastnosti te cenilke θ :

(*) **Cenilka je nepristranska** saj

$$E[\hat{\theta}] = \frac{1}{N} \sum_i E[g(x_i)/p(x_i)] = \frac{1}{N} N E[g(x)/p(x)] = \theta$$

(*) predvsem nas zanima **varianca cenilke**, za katero zelimo da bo čim manjša pri dani velikosti vzorca, kar bo nase vodilo pri izbiri pomembnostne funkcije $p(x)$

$$\begin{aligned} \text{var}[\hat{\theta}] &= \frac{1}{N^2} \sum_i \text{var} \left[\frac{g(X_i)}{p(X_i)} \right] = \frac{1}{N^2} N \text{var} \left[\frac{g(X)}{p(X)} \right] \\ &= \frac{1}{N} \left(E \left[\frac{g^2(X)}{p^2(X)} \right] - \left(E \left[\frac{g(X)}{p(X)} \right] \right)^2 \right) = \frac{1}{N} \left(E \left[\frac{g^2(X)}{p^2(X)} \right] - \theta^2 \right) = \frac{1}{N} (\hat{\theta}^2 - \theta^2) \end{aligned}$$

saj $E[g(X)/p(X)] = \theta$. Zelimo najti $p(x)$ ki minimizira to varianco. Ključ v zmanjšanju variance cenilke je v členu $E[g^2/p^2]$, spodnjo mejo tega člena pa narekuje Jensenova neenakost

$$E \left[\frac{g^2(X)}{p^2(X)} \right] \geq \left(E \left[\frac{|g(X)|}{p(X)} \right] \right)^2 = \left(\int \frac{|g(x)|}{p(x)} p(x) dx \right)^2 = \left(\int |g(x)| dx \right)^2$$

Spodnja meja $E[g^2/p^2]$ oziroma enakost neenacbi je dosežena pri

$$p(x) = \frac{|g(x)|}{\int_a^b |g(x')| dx'}$$

Toda točne vrednosti integrala v imenovalcu ne poznamo, sicer je sploh ne bi računali! V praksi zato iščemo funkcijo $p(x)$, ki je čim bolj podobna funkciji $|g(x)|$, torej takšno $p(x)$, da je razmerje $|g(x)|/p(x)$ po vsem integracijskem območju približno konstantno.

Pomembnostno funkcijo $p(x)$ torej izberemo tako, da bo z večjo verjetnostjo vzorčila območja, kjer je $|g|$ velik in manj vzorčila območja z majhnim $|g|$ - temu pravimo pomembnostno vzorčenje (importance sampling)

Zgled (str 290): Kot preprost zgled za učinkovitost Monte Carlo metode izračunajmo integral z $g(x)=\cos(\pi x/2)$

$$\theta = \int_0^1 \cos\left(\frac{\pi x}{2}\right) dx \qquad \theta_0 = \frac{2}{\pi}$$

Tega tudi sicer tudi zlahka analitično določimo, in njegova analitična vrednost θ_0 nam bo služilo kot točna referenčna točka

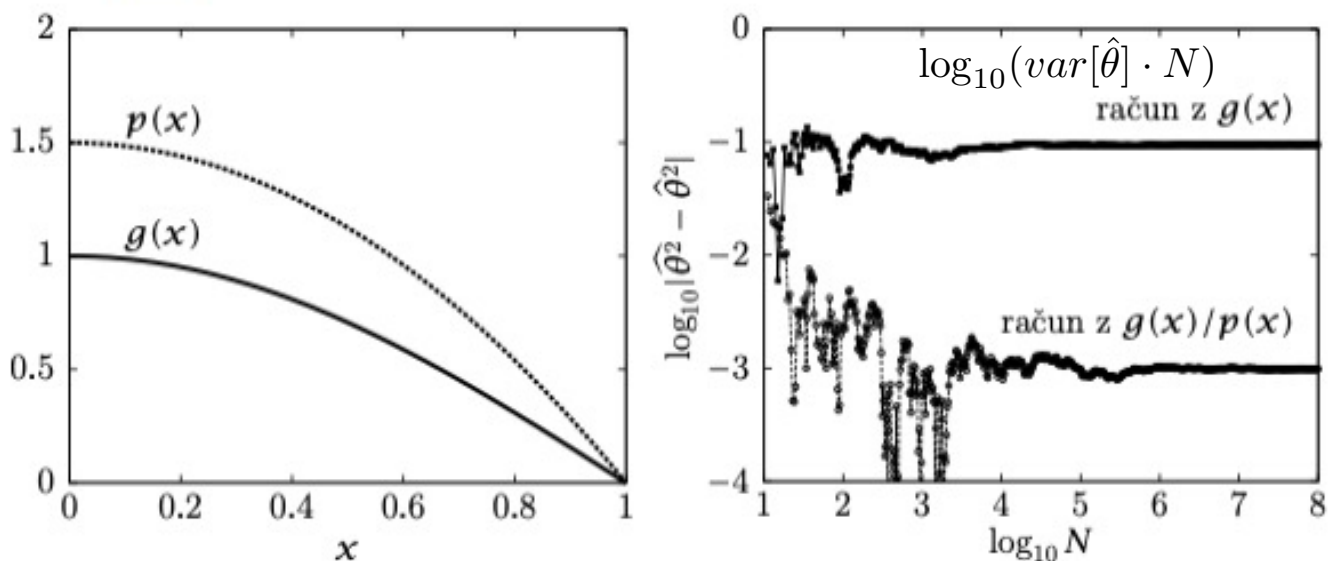
Najprej se ne menimo za pomembnostno funkcijo in delamo po starem: z vrednostmi $\{x_i\}_{i=1}^N$ enakomerno porazdeljene spremenljivke $X \sim U(0, 1)$ izračunamo

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^N g(x_i), \quad \hat{\theta}^2 = \frac{1}{N} \sum_{i=1}^N g^2(x_i). \quad (13.13)$$

Varianco $\text{var}[\hat{\theta}]$ ocene $\hat{\theta}$ v limiti velikih N lahko celo analitično izračunamo:

$$\lim_{N \rightarrow \infty} (\hat{\theta}^2 - \hat{\theta}^2) = \int_0^1 \cos^2\left(\frac{\pi x}{2}\right) dx - \left[\int_0^1 \cos\left(\frac{\pi x}{2}\right) dx \right]^2 \approx 0.09472. \quad (13.14)$$

Dobljeno vrednost $\log_{10} 0.09472 \approx -1.0236$ razberemo tudi iz zgornjega grafa na sliki 13.4 (desno), ki prikazuje približek za varianco v odvisnosti od N .



Slika 13.4 — Uteženje integranda g s pomembnostno funkcijo p . [LEVO] Grafa funkcij $g(x) = \cos(\pi x/2)$ in $p(x) = \frac{3}{2}(1 - x^2)$. [DESNO] Varianca ocene MC brez uporabe pomembnostne funkcije (zgornji graf) in s pomembnostno funkcijo p (spodnji graf).

Zdaj pa si izberimo pomembnostno funkcijo p , ki je “čim bolj podobna” funkciji g , denimo $p(x) = \frac{3}{2}(1 - x^2)$. Funkcija p je na $[0, 1]$ nenegativna in normirana na 1, torej zadošča pogojem za verjetnostno gostoto. Ker zdaj računamo integral

$$\theta = \int_0^1 \frac{g(x)}{p(x)} p(x) dx,$$

moramo vrednosti x_i v vsotah

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^N \frac{g(x_i)}{p(x_i)}, \quad \hat{\theta}^2 = \frac{1}{N} \sum_{i=1}^N \left(\frac{g(x_i)}{p(x_i)} \right)^2, \quad (13.15)$$

žrebat v skladu z gostoto p ! (Naključne vrednosti v skladu s takšno porazdelitvijo dobimo s kakšno metodo iz razdelka C.2.) Čaka nas prijetno presenečenje:

$$\lim_{N \rightarrow \infty} (\hat{\theta}^2 - \hat{\theta}^2) \approx 0.000990.$$

S posrečeno izbiro funkcije p smo torej dosegli približno stokrat manjšo varianco kot v primeru ocene (13.14); glej spodnji graf na sliki 13.4 (desno), ki se za velike N ustali pri $\log_{10} 0.000990 \approx -3.0044$. ◀

Preiskusimo odstopanje dobljenega integrala od prave vrednosti (theta-theta0) za obe pomembnostni funkciji tudi z Mathematico, pri čemer v obeh primerih to napravimo za 5 vzorcev: (gradiva: MC-str290.nb)

```
theta0 = Integrate[Cos[Pi * x / 2], {x, 0, 1}];
g[x_] = Cos[Pi * x / 2];
Print["theta0=", theta0, "=", N[theta0]]

theta0 =  $\frac{2}{\pi}$  = 0.63662

n = 10 000;
p1[x_] := 1;
Do[dist = ProbabilityDistribution[p1[x], {x, 0, 1}];
  xvec = RandomVariate[dist, n];
  theta1 = 1 / n * Sum[g[xvec[[i]]] / p1[xvec[[i]]], {i, 1, n}];
  Print["vzorec ", j, ": theta-theta0=", theta1 - theta0], {j, 1, 5}]
Print["Standardni odkon iz zgornje funkcije na sliki 13.4b:
~Sqrt[0.1] / Sqrt[n]=", Sqrt[0.1] / Sqrt[n]]

vzorec 1: theta-theta0=-0.002364
vzorec 2: theta-theta0=-0.00565783
vzorec 3: theta-theta0=-0.00350846
vzorec 4: theta-theta0=0.00210052
vzorec 5: theta-theta0=0.000517369

Standardni odkon iz zgornje funkcije na sliki 13.4b:
~Sqrt[0.1] / Sqrt[n]=0.00316228

n = 10 000;
p2[x_] := 3 / 2 * (1 - x ^ 2);
Do[dist = ProbabilityDistribution[p2[x], {x, 0, 1}];
  xvec = RandomVariate[dist, n];
  theta2 = 1 / n * Sum[g[xvec[[i]]] / p2[xvec[[i]]], {i, 1, n}];
  Print["vzorec ", j, ": theta-theta0=", theta2 - theta0], {j, 1, 5}]
Print["Standardni odkon iz spodnje funkcije na sliki 13.4b:
~Sqrt[0.001] / Sqrt[n]=", Sqrt[0.001] / Sqrt[n]]

vzorec 1: theta-theta0=0.0000408722
vzorec 2: theta-theta0=0.000339246
vzorec 3: theta-theta0=0.00052314
vzorec 4: theta-theta0=0.000362635
vzorec 5: theta-theta0=0.000187948

Standardni odkon iz spodnje funkcije na sliki 13.4b:
~Sqrt[0.001] / Sqrt[n]=0.000316228
```


13.2 Metoda Monte Carlo na osnovi verig Markova

V prejšnjem razdelku smo opredelili zakaj bi potrebovali način kako izžrebamo naključne vrednosti x , ki so porazdeljene v okviru poljubne porazdelitve $p(x)$. Spodaj bomo obravnavali primer kako izžrebati zvezne naključne vrednosti x na integracijskem intervalu $[a,b]$ v skladu s poljubno porazdelitvijo, ki je normirana na tem intervalu; primer ki bi bil relevanten za zgled iz prejšnjega razdelka je interval $[a,b]=[0,1]$ in verjetnostna gostota $p(x)=3/2 (1-x^2)$.

V ta namen se napogosteje uporablja verige Markova (poglavje 12 v knjigi). V verigi Markova obravnavamo zaporedje naključnih spremenljivk ob zaporednih časih

$$1, \quad 2, \dots, \quad t-1, \quad t, \quad t+1, \dots$$
$$f_x(x;1) f_x(x;2) f_x(x;t-1) f_x(x;t) f_x(x;t+1)$$

Pri vsakem času ima naključna spremenljivka X neko porazdelitev po intervalu $[a,b]$, in ta porazdelitev $f_x(x;t)$ je od časa odvisna. Posebnost Markovih verig je, da je verjetnost za stanje sistema ob času t odvisna le od njegovega stanja pri času $t-1$, vsa prejšnja stanja pa so za razvoj sistema nepomembna – pravimo da gre za **proces brez spomina**. Pri teh procesih je torej relevantna prehodna amplituda za določen korak **$P(x' | x)$, ki predstavlja pogojno verjetnostno gostoto da vrednost x v naslednjem koraku preide v vrednost x'** , oboja pa sta v našem primeru realni števili na intervalu $[a,b]$. **$P(x' | x)$ pravimo tudi prehodna amplituda** oz tranzicijska amplituda in se v literaturi pogosto označi z $T(x \rightarrow x')$. V naslednjem razdelku si bomo ogledali zgled metode kako izbrati $P(x' | x)$, da bo Markova veriga vodila k željeni porazdelitvi naključnih spremenljivk x .

Na osnovi verig Markova zasnujemo učinkovito metodo za žrebanje naključnih vrednosti v skladu s poljubno, lahko tudi zelo zapleteno verjetnostno porazdelitvijo (angl. Markov-Chain Monte Carlo, MCMC) [9]. Bistvo metode je v tem, da izžrebane vrednosti tvorijo stanja verige Markova, katerih ravnovesna porazdelitev je ravno zahtevana verjetnostna porazdelitev. Relevantne verige Markova namreč po določenem času dosežejo **ravnovesno porazdelitev – to je tisto pri kateri se verjetnostna gostota s časom več ne spreminja**. Ravnovesno porazdelitev se v literaturi pogosto **označi s $\pi(x)$** in le ta je ravnovesna, če zadošča **enačbam podrobnega ravnovesja (detailed balance equations)**

$$P(x' | x) \pi(x) = P(x | x') \pi(x'), \quad \text{za vsak } x \text{ in } x' \text{ na intervalu } [a,b]$$

saj se morajo biti prehodi iz x v x' enako pogosti kot prehodi x' v x .

Za primer numerične integracije iz prejšnjega razdelka nas bi, na primer, zanimalo kako doseči ravnovesno porazdelitev $\pi(x)$ z naslednjo odvisnostjo od x : $\pi(x)=3/2 (1-x^2)$.

V naslednjem razdelku si bomo ogledali zgled metode kako izbrati $P(x' | x)$, da bo Markova veriga tezila k željenemu ravnovesnemu stanju $\pi(x)$.

Metropolis Hastingsov algoritm za tvorbo ravnovesne porazdelitve $\pi(x)$

Po tem algoritmu po določenem času tvorimo poljubno ravnovesno porazdelitev $\pi(x)$, ki zadošča zgornji enačbi podrobnega ravnovesja : $P(x'|x) \pi(x) = P(x|x') \pi(x')$.

V mislih imejmo eno-dimenzionalno porazdelitev $\pi(x)$ naključne zvezne spremenljivke x na intervalu $[a,b]$.

Potrebni koraki, ki jih le v grobem navedemo:

Za celoten postopek izberemo tudi poljubno prehodno verjetnostno amplitudo $q(x'|x)$, na primer enakomerno, normalno,...

- 1) ob zacetnem casu izberi poljuben x_0 na intervalu $[a,b]$
- 2) iz x pripravi za naslednji korak x' tako da x' izzrebas v skladu z izbrano $q(x'|x)$
- 3) Izberi pripravljen x' z verjetnostjo ρ (acceptance probability)

$$\rho(x', x) = \min \left(1, \frac{q(x|x') \pi(x')}{q(x'|x) \pi(x)} \right)$$

Ce x' ni izbran, v naslednjem koraku obdrzi x .

Vrednosti x obdrzi tudi ce je x' izven intervala $[a,b]$

- 4) pojdi na korak 2

Pripadajoca prehodna amplituda **$P(x'|x) = \rho(x',x) q(x'|x)$** **zadosca enačbi podrobnega ravnovesja** $P(x'|x) \pi(x) = P(x|x') \pi(x')$ za poljuben $q(x'|x)$, (dokaz je dolg le nekaj vrstic in ga zlahka najdete v literaturi [13]) zato veriga števil po določenem času ustreza ravnovesni porazdelitvi π .

Algoritma generirata ravnovesno porazdelitev π , tudi če ne poznamo njene normalizacijske konstante, saj se ta v razmerju $\pi(x')/\pi(x)$ pokrajša. Poleg tega je fascinantno, da jo generirata ne glede na obliko funkcije q !

Zagotoviti moramo le, da imata π in q isto definicijsko območje. Kljub temu z vidika učinkovitosti in natančnosti algoritma ni vseeno, kakšno funkcijo q izberemo. Za dodatne informacije glej [13].

Poudarimo, da je ta metoda izjemno dragocena za mnoge namene. Z njo lahko recimo tvorimo konfiguracije sistema, kjer je verjetnostna porazdelitev raznih konfiguracij Boltzmanova. Primer konfiguracije pri 1D spinski verigi je $1 \uparrow \dots n \downarrow$, torej gre za $D=2^n$ dimenzionalen problem, željena Markova veriga pa bi lahko tvorila konfiguracije z verjetnostjo $e^{(-E_i/kT)}$.

11. Informacija in informacijska entropija

11.1 Mere informacije in informacijske entropije

Informacija

Ena od možnih poti k definiciji entropije vodi prek pojma *informacije*. Osvetlimo to zvezo z obravnavo diskretne naključne spremenljivke X , ki lahko zavzame končno mnogo vrednosti $\{x_1, x_2, \dots, x_n\}$ z verjetnostmi $\{p_1, p_2, \dots, p_n\}$, kjer je $p_i = P(X = x_i)$ in $\sum_{i=1}^n p_i = 1$. Predstavljamo si, da z vsakim izidom poskusa s takšno naključno spremenljivko, na primer z dogodkom $(X = x)$, ki se zgodi z verjetnostjo $p = P(X = x)$, dobimo informacijo $I(p)$. Vrednost x razumemo kot "signal" ali "sporočilo", ki nosi informacijo $I(p)$.

Kako izmerimo njeno količino? Intuitivno je jasno, da mora imeti mera informacije logaritemski značaj [1,2]. Če se namreč dogodka z verjetnostma p_1 in p_2 zgodita neodvisno drug od drugega (verjetnost $p_1 p_2$), pričakujemo, da bo informacija iz takšnega kombiniranega izida enaka vsoti informacij iz posameznih izidov: stavka "sneži" in "danes je sredo" nosita skupaj natanko toliko informacije kot stavek "sneži in danes je sredo". Primerna mera informacije naj bi bila torej aditivna,

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

Poleg tega si od funkcije $I(p)$ želimo še nenegativnost, $I(p) \geq 0$, monotonost, $p_1 < p_2 \Rightarrow I(p_1) > I(p_2)$, in zveznost: majhne spremembe v p pomenijo majhne spremembe v $I(p)$. Kar sama od sebe se ponuja funkcija

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

in pokazati je mogoče, da je v resnici tudi edina možna [3]. Tako definirana mera ima tudi smiselni lastnosti $I(1) = 0$ (gotov dogodek ne prinaša nikakršne informacije) in $\lim_{p \rightarrow 0} I(p) = \infty$ (zelo neverjeten dogodek nosi zelo veliko informacije). Poljubno realno konstanto C lahko z zvezo $\log_b x = \log x / \log b$ skrijemo v bazo logaritma in je zato nepomembna. Če se dogovorimo za $b = 2$ in $C = 1$, informacijo merimo v *bitih*, če izberemo $b = e$ in $C = 1$, pa v *natih*, ki jih od bitov loči zgolj faktor $\log 2$.

$$I(p_i) \equiv -\log p_i \quad \text{količina informacije (definicija)}$$

Informacijska entropija za diskretne naključne spremenljivke

Od informacije do informacijske entropije je od tod le še korak. Če se posamezne vrednosti x_i pojavljajo z verjetnostmi p_i , $i = 1, 2, \dots, n$, je povprečna količina sprejete ali "nastale" informacije enaka

informacijska entropija
H za diskretne xi
(definicija)

$$H(p_1, \dots, p_n) \equiv E[I] = -\overline{\log p_i} = -\sum_{i=1}^n p_i \log p_i \quad (11.1)$$

To "tehtnico" informacije po Shannonu [3] imenujemo entropija končne verjetnostne porazdelitve.

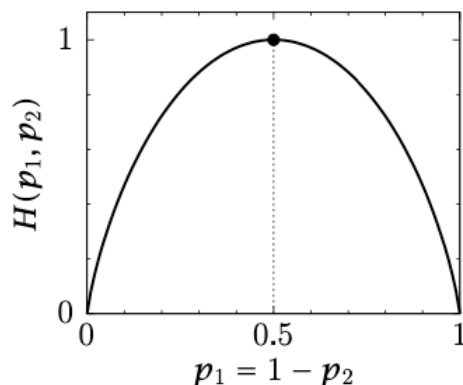
Zgled Pri metu poštenega kovanca sta grb in cifra enako verjetna: $p_1 = p_2 = \frac{1}{2}$. Entropija naključne spremenljivke s takšno porazdelitvijo je

$$H = -\left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2}\right) = 1.$$

Z metom poštenega kovanca dobimo torej natanko en bit informacije. Če kovanec ni pošten, tako da je na primer $p_1 = \frac{9}{20}$ in $p_2 = \frac{11}{20}$, pa dobimo

$$H = -\left(\frac{9}{20} \log_2 \frac{9}{20} + \frac{11}{20} \log_2 \frac{11}{20}\right) \approx 0.9928 < 1.$$

Entropija se je zmanjšala, saj je zaradi kovanca, ki "vleče" v eno stran, izid meta manj negotov. Odvisnost H od $p_1 = 1 - p_2$ prikazuje slika. ◀



Kako lahko definirano entropijo H interpretiramo?

Bistvo slehernega naključnega procesa je *negotovost*. Izidi niso predvidljivi, vendar se s sprejetjem signala (ene same vrednosti x_i od n možnih) nekoliko zmanjša negotovost, v kateri smo bili pred sprejetjem. Izraz (11.1) smemo torej razumeti kot merilo za takšno negotovost. Zavedati pa se moramo, da H meri informacijsko entropijo, ki je ne smemo mešati s termodinamsko entropijo S . V nadaljevanju izraz 'entropija' pomeni informacijsko entropijo.⁷

Mera entropije (11.1) ima vrsto ugodnih lastnosti. Negotovost gotovega dogodka je nič, $H(p = 1) = 0$. Tudi negotovost nemogočega dogodka je nič, $H(p = 0) = 0$, saj ni v dogodku, ki se nikoli ne zgodi, nič nejasnega, poleg tega je tudi formalno $\lim_{p \rightarrow 0} p \log p = 0$. Vrednost entropije je odvisna izključno od verjetnostne porazdelitve $\{p_i\}$ in od nobene druge lastnosti, ki bi jo utegnili pripisati signalu. Je neodvisna od permutacij med p_i in se ne spremeni, če n dogodkom dodamo nemogoč dogodek, $H(p_1, p_2, \dots, p_n) = H(p_1, p_2, \dots, p_n, 0)$. Entropija je maksimalna, ko smo "maksimalno negotovi", to je tedaj, ko so vsi izidi enako verjetni: $p_1 = p_2 = \dots = p_n = 1/n$ (enakomerna porazdelitev). Za katero koli drugo porazdelitev ali ob kakršnem koli pogoju za p_i se entropija zmanjša (glej zgled na str. 250 in [5]).

Informacijska entropija za zvezne naključne spremenljivke

Entropijo zvezne verjetnostne porazdelitve definiramo po analogiji z (11.1). Če je X zvezna naključna spremenljivka z verjetnostno gostoto f , je entropija njene porazdelitve enaka

$$H(X) = - \int_{-\infty}^{\infty} f(x) \log f(x) dx.$$

informacijska entropija H
za zvezne x (11.3)
(definicija)

Zgled Entropija enakomerne porazdelitve $U(a, b)$ oziroma enakomerno porazdeljene zvezne naključne spremenljivke $X \sim U(a, b)$ je

$$H(X) = -\frac{1}{b-a} \log \frac{1}{b-a} \int_a^b dx = \log(b-a). \quad (11.4)$$

Rezultat je odvisen samo od razlike $b - a$. To pomeni, da imajo vse enakomerne porazdelitve z isto širino intervala isto entropijo. ◀

11.2 Načelo maksimalne informacijske entropije

Zapisali bomo načelo maksimalne entropije, navedli nekaj primerov, in nato našli nekaj razmislekov, ki so bili vodilo pri zapisu tega načela.

Načelo maksimalne informacijske entropije pravi: Sistemu, o katerem imamo nezadostne podatke in spoštuje določene vezi, je želimo pripisati verjetnostno porazdelitev naključnih spremenljivk. Tedaj je po načelu maksimalne entropije sistemu najbolj "smiselno" pripisati verjetnostno porazdelitev, ki ji ustreza maksimalna informacijska entropija pri danih vezeh.

Gre za načelo, ki se ga v splošnem ne da matematično dokazati, a vodi do mnogih izjemno pomembnih verjetnostnih porazdelitev ki so realizirane v naravi. Navedimo nekaj primerov, kjer pomembne porazdelitve sledijo pri danih "vezeh":

- **enakomerna porazdelitev** ima največjo entropijo med vsemi porazdelitvami, kjer so vrednosti porazdeljene na koncnem intervalu. To velja za diskretne ali zveze porazdelitve (knjiga 11.2, str 250).
 - Kor primer smo ze obravnavali diskretno porazdelitev pri konancu, kjer maksimalno entropijo dobimo za enakomerno porazdelitev $p_1=p_2=1/2$.
 - Pri kocki da maksimalno entropijo $p_i=1/6$, kot bomo pokazali z zgledom spodaj.
- **normalna porazdelitev** ima največjo entropijo med vsemi zveznimi porazdelitvami z danim povprecjem in varianco, kjer so vrednosti porazdeljene na celotnem intervalu $x=[-\infty, \infty]$. (knjiga 11.4, str 257)
- **Boltzmanova porazdelitev** $P(E_i)=e^{-\beta E_i}$ ima največjo entropijo med vsemi z dano pričakovano vrednostjo energije, kot bomo pokazali spodaj (knjiga str 253).
- **Bose-Einsteinova in Fermi-Diracova porazdelitev** tudi ustrezata nacelu maksimalne entropije pri določenih vezeh (knjiga 11.3.5 in 11.3.6)

Razmisleki, ki vodijo do zgornjega načela:

Videli smo, da je v opazovanju naključnega procesa "negotovost" — torej informacijska entropija — največja, ko so vsi njegovi izidi enako verjetni. To spoznanje je v preprosto obliko strnil že Laplace v *načelu nezadostnega razloga* ali *indiferentnosti* (angl. *principle of insufficient reason / indifference*): če nimamo razloga, da bi kakor koli razlikovali med dvema ali več izidi, je najboljša strategija ta, da jih obravnavamo kot enako verjetne.

Načelo maksimalne entropije to vodilo povzema in nadgrajuje: v kakršnih koli okoliščinah, kjer imamo na voljo nepopolno informacijo — na primer vzorec izidov naključnega poskusa — iščemo za njihov kvantitativni opis verjetnostno porazdelitev, ki je konsistentna z vso *znano* informacijo, obenem pa je karseda "nezavezujoča", "negotova", "prosta" glede na *neznano* informacijo⁸. Laplaceov

11.3: Zgledi z maksimalno entropijo;

iskanje ekstrema pri danih vezeh z metodo Lagrangevih multiplikatorjev:

Marcun, matematika 2, poglavje 4.2.5

TRDITEV 4.29 (Lagrangeova metoda multiplikatorjev). Naj bo $f : \mathcal{D} \rightarrow \mathbb{R}$ totalno odvedljiva skalarna funkcija n spremenljivk, definirana na odprti podmnožici $\mathcal{D} \subset \mathbb{R}^n$, in naj bo $g : \mathcal{E} \rightarrow \mathbb{R}^m$ zvezno odvedljiva vektorska funkcija n spremenljivk, definirana na odprti podmnožici $\mathcal{E} \subset \mathbb{R}^n$. Naj bo $a \in \mathcal{E} \cap \mathcal{D}$ takšna točka, da je $g(a) = 0$ in da je $\text{rank}(Dg)(a) = m$. Če ima funkcija f v točki a vezan ekstrem pri vezi g , potem obstajajo takšne realne konstante $\lambda_1, \lambda_2, \dots, \lambda_m$, da velja

$$(\nabla f)(a) = \sum_{i=1}^m \lambda_i (\nabla g_i)(a).$$

$$\nabla \left(f - \sum_i \lambda_i g_i \right) = 0$$

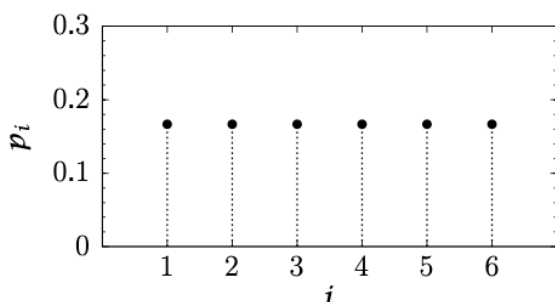
iscemo ekstrem
Lagrangeve funkcije

$$\mathcal{L} = f - \sum_i \lambda_i g_i$$

Zgled Po Laplaceu metanju poštene kocke ustreza enakomerna porazdelitev $p_i = 1/6$, $i = 1, 2, \dots, 6$. Do iste ugotovitve vodi tudi načelo maksimalne entropije. Maksimizirati moramo $-\sum_{i=1}^6 p_i \log p_i$ ob pogoju $\sum_{i=1}^6 p_i = 1$. Za to nalogo se kar ponuja klasična metoda Lagrangeevih multiplikatorjev iz variacijskega računa. Odvajamo Lagrangeovo funkcijo

$$\mathcal{L} = -\sum_{i=1}^6 p_i \log p_i - \lambda \left(\sum_{i=1}^6 p_i - 1 \right)$$

po p_i in odvod postavimo na nič. Sledi $\partial \mathcal{L} / \partial p_i = -\log p_i - 1 - \lambda = 0$ in $p_i = e^{-\lambda-1}$. Torej so p_i neodvisni od i (z drugimi besedami, vsi enaki), njihova vsota pa mora biti 1. Potemtakem $p_1 = p_2 = \dots = p_6 = 1/6$, glej sliko 11.2 (levo). V nadaljevanju ta zgled razširimo v splošnejše orodje na osnovi Lagrangeevih multiplikatorjev. $\triangleleft$



Slika 11.2 — Diskretna porazdelitev vrednosti pri metu igralne kocke in načelo maksimalne entropije. [LEVO] Normirana porazdelitev brez omejitev (vezi).

11.3.3 Maxwell-Boltzmannova porazdelitev

Denimo, da so v fizikalnem sistemu na voljo energijski nivoji z enodelčnimi energijami $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, ki jih delci zasedajo z verjetnostmi $p_1, p_2, \dots, p_n$. Naj bo pričakovana vrednost energije, ε , predpisana. Poiščimo verjetnostno porazdelitev delcev, ki je skladna s predpostavko maksimalne entropije. Maksimizirati moramo entropijo z vezema

$$\sum_{i=1}^n p_i = 1, \quad \sum_{i=1}^n p_i \varepsilon_i = \varepsilon, \quad (11.15)$$

torej Lagrangeovo funkcijo

naceloma bi to označili z λ_0 , a lahko označimo poljubno, tu označimo z $1-\lambda_0$ da dobimo lepsi izraz

$$\mathcal{L} = - \sum_{i=1}^n p_i \log p_i - (\lambda_0 - 1) \left(\sum_{i=1}^n p_i - 1 \right) - \lambda_1 \left(\sum_{i=1}^n p_i \varepsilon_i - \varepsilon \right).$$

Dobimo $\partial \mathcal{L} / \partial p_i = -\log p_i - \lambda_0 - \lambda_1 \varepsilon_i = 0$ oziroma $p_i = e^{-\lambda_0} e^{-\lambda_1 \varepsilon_i} \equiv a e^{-\lambda_1 \varepsilon_i}$, kjer je $i = 1, 2, \dots, n$. To vstavimo v enačbi (11.15), od koder sledi

$$p_i = \frac{e^{-\lambda_1 \varepsilon_i}}{\sum_{i=1}^n e^{-\lambda_1 \varepsilon_i}}, \quad \varepsilon = \frac{\sum_{i=1}^n \varepsilon_i e^{-\lambda_1 \varepsilon_i}}{\sum_{i=1}^n e^{-\lambda_1 \varepsilon_i}}. \quad (11.16)$$

Če postavimo $1/\lambda_1 = k_B T$, kjer je k_B Boltzmannova konstanta, ta izraza podajata Maxwell-Boltzmannovo porazdelitev (glej tudi podrazdelek 11.3.4).