MATEMATIČNA ANALIZA IV
izročki predavanj

Marjeta Kramar Fijavž
Fakulteta za gradbeništvo in geodezijo Univerze v Ljubljani
Katedra za matematiko in fiziko

2010/11

Vsebina predavanj

Današnja snov

Osnovni pojmi

Poskus in dogodek

Računanje z dogodki

Definicije verjetnosti

Pogojna verjetnost, neodvisnost dogodkov

Poskus, izid

Poskus je postopek, ki ga je mogoče poljubno mnogokrat pod enakimi pogoji ponoviti. Dogajanje je bodisi sproženo namenoma bodisi ga le opazujemo. *Izid je nepredvidljiv!*

Zgledi: met kocke, nakup srečke, štetje prometa, obteževanje nosilcev

Vzorčni prostor $\mathcal{S}$ je množica vseh možnih elementarnih izidov poskusov. Lahko je *diskreten* ali *zvezen*, *končen* ali *neskončen*.

Vsako ponavljanje poskusa imenujemo *proces*.

Dogodek

Poljubno podmnožico vzorčnega prostora $A \subseteq \mathcal{S}$ imenujemo *dogodek*.
Vsak dogodek sestavljajo elementarni izidi.

Označimo izid nekega poskusa z a. Če je $a \in A$ rečemo, da se je dogodek A *zgodil*.

Posebna dogodka:

G: *gotov dogodek* (zgodi se ob vsaki ponovitvi poskusa)
N: *nemogoč dogodek* (ne zgodi se ob nobeni ponovitvi poskusa)

Zgledi

- vržemo kocko

- štejemo avtomobile, ki gredo skozi predor do prve naslednje nesreče v predoru

- zanima nas mesto prve naslednje nesreče v predoru

Operacije

$a \in A$: dogodek A *vsebuje* izid a

$A \subseteq B$: dogodek A je *način* dogodka B, kadar se zgodi A se zagotovo zgodi tudi B.

$A = B$: dogodka A in B sta *enaka*, vedno se zgodita sočasno. V tem primeru velja $A \subseteq B$ in $B \subseteq A$.

$A + B$: *vsota dogodkov*, zgodi se vsaj eden od dogodkov A in B.

AB: *produkt dogodkov*, dogodka A in B se zgodita hkrati.

$\overline{A}$: *nasproten dogodek*, dogodek, da se A ne zgodi.

Nezdružljivost, popoln sistem

Dogodka A in B sta *nezdružljiva*, če se ne moreta zgoditi hkrati, tj. ko je $AB = N$.

Dogodki $A_1, \ldots, A_n$ sestavljajo *popoln sistem dogodkov*, če se v vsaki ponovitvi poskusa zgodi natanko eden izmed njih, tj.

$$A_i \neq N \quad \forall i, \qquad A_i A_j = N \quad i \neq j, \qquad \sum_{i=1}^{n} A_i = G.$$

Teorija množic

oznaka	teorija množic	verjetnostni račun	oznaka
$\mathcal{S}$	univerzalna množica	vzorčni prostor	$\mathcal{S}$
A	množica	dogodek	A
a	element	elementarni izid	a
$\emptyset$	prazna množica	nemogoč dogodek	N
$\mathcal{S}$	cela množica	gotov dogodek	G
$A \subseteq B$	podmnožica	način	$A \subseteq B$
$A \cup B$	unija	vsota	$A + B$
$A \cap B$	presek	produkt	AB
$A^{\complement}$	komplement	nasproten dogodek	$\overline{A}$
$A \cap B = \emptyset$	disjunktnost	nezdružljivost	$AB = N$

Pravila

Veljajo vsa pravila 'prepisana' v jezik dogodkov. Lahko si pomagamo z Vennovimi diagrami.

Distributivnost:

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

De Morganovi pravili:

$$\overline{A + B} = \overline{A} \cdot \overline{B} \qquad \text{in} \qquad \overline{AB} = \overline{A} + \overline{B}$$

Zgled

Mesto $\mathcal{M}$ dobavlja vodo iz dveh izvirov, pri čemer vsak izvir zase zadošča vodnim potrebam mesta. Napeljava je narejena tako, da se cevi c_1 in c_2, ki peljeta od vsakega od izvirov, pred mestom združita v skupno cev c_3. Simbolno zapiši dogodka, da mestu $\mathcal{M}$ zaradi iztekanja iz katere od cevi primanjkuje oziroma ne primanjkuje vode.

Verjetnost

Verjetnost dogodka je vrednost, ki nam pove, kolikšna je možnost, da se pri izbranem poskusu zgodi določen dogodek.

Izražamo jo z realnimi števili iz intervala $[0, 1]$, v vsakdanjem življenju pogosteje z odstotki med 0 in 100%.

Oznaka: $P(A)$.

Klasična definicija (Pascal, Fermat, 17. stoletje)

Privzamemo: $\mathcal{S} = \{a_1, \ldots, a_n\}$, kjer so a_i enako verjetni in paroma nezdružljivi izidi poskusa.

Za dogodek A je ugodnih k izmed teh izidov.V tem primeru je verjetnost dogodka A enaka

$$P(A) := \frac{k}{n}.$$

Zgleda

1. Kolikšna je verjetnost, da pri metu kocke pade sodo število?

2. Imamo tri tovornjake. Verjetnost, da se posamezen tovornjak v pol leta ne pokvari je $\frac{1}{2}$. Kolikšna je verjetnost, da bosta po pol leta vsaj dva tovornjaka brezhibna?

Geometrijska definicija (18. stoletje)

Klasično definicijo verjetnosti lahko z diskretnega posplošimo na zvezen prostor, če elementarne izide poskusa lahko predstavimo kot enakovredne točke na delu *premice*, *ravnine* ali *prostora*. Štetje nadomestimo z neko mero m (kot npr. *dolžino*, *ploščino* ali *prostornino*). Verjetnost dogodka A tako dobimo kot

$$P(A) := \frac{m(A)}{m(\mathcal{S})}.$$

Zgled

Opazujemo 10-metrski nosilec, pritrjen na obeh krajiščih. Na poljubno mesto na nosilcu položimo tovor težak 100 kg. Zanima nas reakcija v levem krajišču: R_L. Kolikšna je verjetnost, da je

(a) $10 \leq R_L \leq 20$?

(b) $R_L > 50$?

Statistična definicija (Bernoulli, 18. stoletje)

Nek poskus ponovimo n-krat, pri tem se dogodek A zgodi k-krat. Število

$$f_n(A) = \frac{k}{n}$$

imenujemo *relativna frekvenca* dogodka A.

Bernoullijev zakon stabilnosti: po dovolj veliko ponovitvah poskusa se relativne frekvence dogodkov ustalijo okoli neke fiksne vrednosti.

$$f_n(A) \rightsquigarrow P(A)$$

Zgledi

- Met kovanca (Buffon, Pearson)

- Za pravilno načrtovanje križišča nas zanima verjetnost, da več kot pet avtomobilov v danem trenutku čaka na zavijanje v levo. Rezultati opazovanj (60 štetij) so zbrani v tabeli:

št. čakajočih	0	1	2	3	4	5	6	7	8
kolikokrat opaženo	4	16	20	14	3	2	1	0	0
relativna frekvenca	$\frac{4}{60}$	$\frac{16}{60}$	$\frac{20}{60}$	$\frac{14}{60}$	$\frac{3}{60}$	$\frac{2}{60}$	$\frac{1}{60}$	0	0

Aksiomatična definicija (Kolmogorov, 20. stoletje)

Verjetnost je *preslikava* $P : \mathcal{S} \to \mathbb{R}$, ki ustreza naslednjim lastnostim:

(P1) $P(A) \geq 0$ za vsak $A \in \mathcal{S}$,
(P2) $P(G) = 1$,
(P3) za *nezdružljiva* dogodka A in B velja: $P(A + B) = P(A) + P(B)$.

Nekatere posledice lastnosti (P1)-(P3)

1. $P(\overline{A}) = 1 - P(A)$
2. $P(N) = 0$
3. $A \subseteq B \Longrightarrow P(A) \leq P(B)$
4. $P(A + B) = P(A) + P(B) - P(AB)$
5. $P(A + B + C) =$
   $P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$

Zgled

Vrnimo se k 10-metrskemu nosilcu, pritrjenem na obeh krajiščih. Predpostavimo tokrat, da je teža tovora poljubno število med 100 in 300 kg. Zanima nas reakcija tako v levem (R_L) kot v desnem (R_D) krajišču. Kolikšna je verjetnost, da je vsaj ena od vrednosti R_L oziroma R_D večja od 100?

Pogojna verjetnost

Naj bo B nek mogoč dogodek, tj. $P(B) \neq 0$. S $P(A|B)$ oz. $P_B(A)$ označimo *verjetnost dogodka A pri pogoju, da se je zgodil B*. Velja:

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

Zgled

Vržemo dve kocki in dobimo vsoto pik na obeh kockah enako 6. Kolikšna je verjetnost, da je pri tem na eni izmed kock padla 4?

Pogojna verjetnost (nad.)

Pogojna verjetnost ima prav take lastnosti kot brezpogojna (veljajo ustrezno prirejeni $(P1) - (P3)$ in vse kar iz le-teh sledi)!

Posledica

Če je $P(B) \neq 0$, za produkt dogodkov velja

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

Če je $P(A) \neq 0$, potem velja tudi

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

Neodvisnost dogodkov

Dogodka A in B sta *neodvisna*, če velja

$$P(A|B) = P(A) \quad \text{oz.} \quad P(B|A) = P(B).$$

Izrek

Dogodka A in B sta neodvisna natanko tedaj, ko velja produktna formula

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

Pozor: neodvisnost $\neq$ nezdružljivost !!!

Današnja snov

Osnovni pojmi (nad.)

Pogojna verjetnost, neodvisnost dogodkov (nad.)

Dvofazni poskusi in Bayesov obrazec

Bernoullijevo zaporedje neodvisnih poskusov

Primeri

Neodvisnost družine dogodkov

Dogodki $A_1, \ldots, A_n$ so *paroma neodvisni*, če je $P(A_i|A_j) = P(A_i)$ za vsak par $i \neq j$. To pomeni, da velja produktna formula

$$P(A_iA_j) = P(A_i)P(A_j), \quad i \neq j.$$

Dogodki $A_1, \ldots, A_n$ so *(medsebojno) neodvisni*, če je $P\left(A_{j_1}|A_{j_2}\cdots A_{j_k}\right) = P(A_{j_1})$ oziroma

$$P\left(A_{j_1}\cdots A_{j_k}\right) = P(A_{j_1})\cdots P(A_{j_k})$$

za vse možne različne indekse $j_1, \ldots, j_k$.

Zgled

Trikrat zapored vržemo kovanec. Kaj lahko povemo o neodvisnosti naslednje družine dogodkov:
$A \ldots$ v prvem metu pade cifra,
$B \ldots$ v drugem metu pade grb,
$C \ldots$ v vseh treh metih je bil izid enak,
$D \ldots$ cifra je padla natanko dvakrat zaporedoma?

Dvofazni (dvostopenjski) poskus

1. faza: zgodi se natanko eden od paroma nezdružljivih izidov $H_1, \ldots, H_n$ (= *hipoteze*)
2. faza: odvisna je od izida na prvi stopnji. Opazujemo dogodek A.

Denimo, da poznamo $P(H_i)$ in $P(A|H_i)$ za vse $i = 1, \ldots, n$. Koliko je potem $P(A)$?

Formula o (po)polni verjetnosti

Izrek

Naj bo $H_1, \ldots, H_n$ popoln sistem dogodkov. Potem je

$$P(A) = \sum_{i=1}^{n} P(H_i)P(A|H_i).$$

Zgleda

1. V neki populaciji ima 70% kadilcev in 15% nekadilcev težave z dihali. Kolikšna je verjetnost težav z dihali, če kadi 40% populacije?

2. Dvakrat vržemo kocko. Kolikšna je verjetnost, da v drugem metu vržemo strogo več pik kot v prvem?

Bayesov obrazec

Vprašanje obrnimo: denimo, da se je dogodek A zgodil. Kolikšna je verjetnost, da se je pred tem zgodila hipoteza H_k?

Izrek (Bayesov obrazec)

$$P(H_k|A) = \frac{P(H_k)P(A|H_k)}{\sum_{i=1}^{n} P(H_i)P(A|H_i)}, \quad k = 1, \dots, n.$$

Zgleda (nad.)

1. Nek pacient ima težave z dihali. Kolikšna je verjetnost, da je kadilec?

2. Pri metu dveh kock smo v drugo vrgli strogo več pik kot v prvem metu. Kolikšna je verjetnost, da smo v prvem metu vrgli 1?

Zaporedje neodvisnih poskusov

- isti poskus ponovimo *velikokrat*
- verjetnost izidov v enem poskusa je *neodvisna* od preostalih poskusov

Bernoullijevo zaporedje poskusov

- možna sta le dva izida, A ali $\overline{A}$
- v vsaki ponovitvi poskusa je $P(A) = p$
  (vrednost p je enaka za vse poskuse!)

Bernoullijev obrazec

$P_n(k) \ldots$ verjetnost, da se dogodek A v n ponovitvah poskusa zgodi natanko k-krat

$P(A) = p$

Izrek

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

Zgled

- Kovanec vržemo 10-krat. Kolikšne so verjetnosti dogodkov, da grb pade 0-krat, 1-krat, 2-krat, vsaj 3-krat?

- Kolikokrat moramo vreči kocko, da lahko z verjetnostjo 0.99 pričakujemo, da bo padla vsaj ena šestica?

Bernoullijev zakon velikih števil

Izrek (Jakob Bernoulli, 1713)

Denimo, da je relativna frekvenca dogodka A v n neodvisnih ponovitvah poskusa enaka $\frac{k}{n}$. Pri tem za vsako ponovitev velja: $P(A) = p$. Potem je za vsak $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) = 1.$$

(statistična definicija verjetnosti!)

Računanje verjetnosti

1. Kolikšna je verjetnost, da imata vsaj dva študenta v predavalnici rojstni dan na isti dan? (Pri tem 29.2. ne upoštevamo!)
2. V krogu slučajno izberemo točko. Kolikšna je verjetnost, da je točka bližje središču kot obodu?
3. Dve ladji pristajata na istem pomolu, vendar se ne moreta istočasno privezati. Njuna prihoda sta neodvisna in kadarkoli v 24h enako verjetna. Prva ladja se ob pomolu zadrži 2 uri, druga pa 4 ure. Kolikšna je verjetnost, da mora ena ladja pred pristaniščem čakati na prost privez?

Računanje z dogodki

1. Dokaži, da za poljubna dogodka A in B velja: $P(A\overline{B}) = P(A) - P(AB)$.
2. Če sta dogodka A in B neodvisna, potem so tudi pari dogodkov A in $\overline{B}$, $\overline{A}$ in B, ter $\overline{A}$ in $\overline{B}$ neodvisni. Dokaži to trditev.
3. Naj za dogodka A in B velja: $P(A) = \frac{3}{8}$, $P(B) = \frac{1}{2}$ in $P(AB) = \frac{1}{4}$. Izračunaj $P(A+B)$, $P(\overline{A})$, $P(\overline{B})$, $P(\overline{A}+\overline{B})$ $P(A\overline{B})$ in $P(\overline{A}B)$. Ali sta dogodka A in B nezdružljiva, ali sta neodvisna?

Bayesov obrazec

1. Pogostost neke bolezni v populaciji je 0.0001. Test, ki je na voljo, je pozitiven za 99% obolelih (*občutljivost*, angl. *sensibility*) in negativen v 95% zdravih (*specifičnost*, angl. *specifity*). Kolikšna je verjetnost, da smo bolni, če je bil test pozitiven?

2. Preden investitor prevzame avtocestni odsek od izvajalca z ultrazvokom preverijo debelino podlage. To storijo na vsakih 150 m in vsak 150-metrski del bo sprejet, če bo izmerjena debelina vsaj 18 cm. V nasprotnem primeru zavrnejo celoten odsek.
Izkušnje pravijo, da je 90% izmerjenih odsekov v skladu s predpisi, vendar je zanesljivost UZ meritev le 80%. Kolikšna je verjetnost, da
   - je odsek dobro zgrajen (tj. debelina podlage dejansko vsaj 18 cm) in ga bo investitor tudi prevzel?
   - je odsek slabo zgrajen, a ga bo investitor na podlagi UZ meritev vseeno sprejel?
   - da bo dobro zgrajen odsek sprejet?

Današnja snov

Slučajne spremenljivke

Definicija

Porazdelitvena funkcija

Diskretne slučajne spremenljivke

Porazdelitev

Matematično upanje in disperzija

Enakomerna diskretna porazdelitev

Porazdelitve, ki izhajajo iz Bernoullijevih poskusov

Slučajna spremenljivka

Slučajna spremenljivka X je funkcija, ki vsakemu izidu poskusa priredi neko realno število:

$$X : \mathcal{S} \to \mathbb{R}.$$

Izid poskusa je bodisi številsko podan bodisi ga lahko primerno številsko interpretiramo.
V posameznem poskusu je mogočih več vrednosti (zato *spremenljivka*), katera od njih nastopi je odvisno zgolj od slučaja (zato *slučajna*).

Zgledi:

met kocke, met kovanca, štetje prometa, merjenje obremenitve, ipd.

Porazdelitveni zakon

Slučajna spremenljivka X je določena

- z *zalogo vrednosti* $\mathcal{Z}_X$ (lahko je *diskretna* ali *zvezna*) in
- s *porazdelitvenim zakonom*, tj. predpisom, ki elementom $\mathcal{Z}_X$ določi njihove verjetnosti.

Porazdelitvena funkcija (CDF) slučajne spremenljivke X je funkcija

$$F_X : \mathbb{R} \to [0,1], \qquad F_X(x) := P(X \leq x).$$

Lastnosti porazdelitvene funkcije

1. $0 \leq F_X(x) \leq 1$ za vsak $x \in \mathbb{R}$
2. $\lim_{x \to -\infty} F_X(x) = 0$ in $\lim_{x \to +\infty} F_X(x) = 1$
3. F_X je monotono naraščajoča: $x_1 < x_2 \Longrightarrow F_X(x_1) \leq F_X(x_2)$
4. $P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$
5. F_X je z desne zvezna:
$$\lim_{h \to 0, h > 0} F_X(x+h) = F_X(x),$$
v nekaj (največ števno mnogo) točkah ima lahko skok
6. $P(X = x) = F_X(x) - \lim_{h \to 0, h>0} F_X(x-h)$

Funkcije slučajne spremenljivke

Naj bo $X : \mathcal{S} \to \mathbb{R}$ slučajna spremenljivka in $g : \mathbb{R} \to \mathbb{R}$ neka funkcija. Potem je $Y = g(X)$: $\mathcal{S} \to \mathbb{R}$ je spet slučajna spremenljivka.

V primeru, da je g strogo naraščajoča, velja

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

Zgleda

$Y = aX + b$ $(a, b \in \mathbb{R}, a > 0)$ ali $Y = X^2$

Opomba: Če sta $X, Y : \mathcal{S} \to \mathbb{R}$ slučajni spremenljivki, je tudi $X + Y : \mathcal{S} \to \mathbb{R}$ slučajna spremenljivka. Njeno porazdelitveno funkcijo bomo izračunali kasneje.

Diskretne slučajne spremenljivke

X imenujemo *diskretna slučajna spremenljivka*, kadar njeno zalogo vrednosti sestavlja števno mnogo točk:

$$\mathcal{Z}_X = \{x_1, x_2, \dots, x_k, \dots\}$$

Za poznavanje X potrebujemo še *verjetnostno funkcijo (PDF)* $p_k := P(X = x_k)$, $k = 1, 2, \dots$ Pri tem velja:

$$p_k \geq 0 \quad \text{in} \quad \sum_{k=1}^{\infty} p_k = 1$$

Verjetnostna shema pove vse o diskretni slučajni spremenljivki:

$$X : \begin{pmatrix} x_1 & x_2 & \cdots & x_k & \cdots \\ p_1 & p_2 & \cdots & p_k & \cdots \end{pmatrix}$$

Porazdelitvena funkcija

Denimo, da so x_i urejeni monotono naraščajoče. Porazdelitvena funkcija X je potem enaka

$$F_X(x_k) = P(X \leq x_k) = \sum_{i=1}^{k} p_i$$

Zgleda:

- met kocke, X: število pik, Y: sodo ali liho
- met 3 kovancev, X: število grbov

Matematično upanje

Matematično upanje diskretne slučajne spremenljivke X označimo z $E(X)$ in definiramo kot "uteženo povprečje" vrednosti

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

- Pomen: pričakovana vrednost, srednja vrednost ($\sim$ statični moment, težišče)
- slučajna spremenljivka X morda vrednosti $E(X)$ nikoli ne zavzame
- $E(X)$ ne obstaja vedno (neskončna vsota)

Zgledi

- met 2 kock, X: večje od števila pik
- met kovanca, X: število metov do prvega grba
- Na loteriji je na voljo 1000 srečk po 10€. Organizator obljublja 100 dobitkov po 5€, 20 dobitkov za 20€, 2 dobitka za 500€in glavni dobitek 5000€. Se nam splača kupiti srečko?

Osnovne lastnosti matematičnega upanja

- za poljubno $g : \mathbb{R} \to \mathbb{R}$ je
$$E\left(g(X)\right) = \sum_{i=1}^{\infty} g(x_i) p_i$$
- $E(aX + b) = aE(X) + b$ za vse $a, b \in \mathbb{R}$
posebej: $E(b) = b$ in $E\left(X - E(X)\right) = 0$
- za $X, Y : \mathbb{R} \to \mathcal{S}$ je $E(X + Y) = E(X) + E(Y)$
- $E\left((X - E(X))^2\right) = E(X^2) - E(X)^2$

Disperzija in standardni odklon

Disperzijo (ali *varianco*) diskretne slučajne spremenljivke X označimo z $D(X)$ (tudi $V(X)$, $\sigma^2(X)$). Meri razpršenost porazdelitve, dobimo jo kot

$$
\begin{aligned}
D(X) &:= E\left((X-E(X))^2\right) = \sum_{i=1}^{\infty} (x_i - E(X))^2\, p_i \\
&= E(X^2) - E(X)^2 = \sum_{i=1}^{\infty} x_i^2 p_i - \left(\sum_{i=1}^{\infty} x_i p_i\right)^2
\end{aligned}
$$

Standardni odklon $\sigma(X)$ je enak

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

Zgledi (nad.)

- met 2 kock, X: večje od števila pik, Y: vsota pik
- met kovanca do prvega grba
- loterija

Standardizacija slučajne spremenljivke

Najprej opazimo, da za disperzijo vedno velja

$$D(aX+b) = a^2 D(X) \quad \text{za vse} \quad a, b \in \mathbb{R}$$

Denimo, da za neko slučajno spremenljivko X obstajata $E(X)$ in $D(X)$. *Standardizirana slučajna spremenljivka* je definirana kot

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

Zanjo velja: $E(X^*) = 0$ in $D(X^*) = 1$.

Enakomerna diskretna porazdelitev

imamo končno mnogo enako verjetnih izidov

$$X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$$

Ob predpostavki, da so x_i urejeni monotono naraščajoče za porazdelitveno funkcijo X velja

$$F_X(x) = \begin{cases} 0, & x \leq x_1; \\ \frac{i}{n}, & x_i < x \leq x_{i+1}; \\ 1, & x > x_n. \end{cases}$$

Bernoullijeva porazdelitev

X ponazarja možna izida Bernoullijevega poskusa:
$A \rightarrowtail 1$ in $\overline{A} \rightarrowtail 0$

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

Porazdelitvena funkcija:

$$F_X(x) = \begin{cases} 0, & x < 0; \\ 1-p, & 0 \leq x < 1; \\ 1, & x \geq 1. \end{cases}$$

$$E(X) = p \quad \text{in} \quad D(X) = (1-p)p$$

Binomska porazdelitev $B(n, p)$

$X =$ število ponovitev dogodka A v n ponovitvah Bernoullijevega poskusa

$$X \sim \begin{pmatrix} 0 & 1 & \cdots & n \\ P_n(0) & P_n(1) & \cdots & P_n(n) \end{pmatrix} \qquad P_n(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$E(X) = np \quad \text{in} \quad D(X) = np(1-p)$$

Geometrijska porazdelitev

$X =$ zaporedna številka ponovitve Bernoullijevega poskusa v katerem se A prvič zgodi (oz. število poskusov med dvema ponovitvama dogodka A)

$$X \sim \begin{pmatrix} 1 & 2 & \cdots & k & \cdots \\ p_1 & p_2 & \cdots & p_k & \cdots \end{pmatrix} \qquad p_k = (1-p)^{k-1}p$$

$$E(X) = \frac{1}{p} \quad \text{in} \quad D(X) = \frac{1-p}{p^2}$$

Pomen: $E(X)$ je povprečna doba med dvema ponovitvama dogodka A

Zgled

Radijski oddajnik je projektiran tako, da zdrži 50-letni veter (t. j. močan rušilen veter, ki se v povprečju pojavi na vsakih 50 let). Kolikšna je verjetnost

- da se bo tak veter pojavil prvič v petem letu po zgraditvi oddajnika?
- da se bo prvi takšen veter pojavil v prvih petih letih po zgraditvi oddajnika?
- natanko enega takšnega vetra v petih letih?

Današnja snov

Diskretne slučajne spremenljivke (nad.)

Poissonova porazelitev

Primeri

Zvezne slučajne spremenljivke

Gostota verjetnosti

Matematično upanje in disperzija

Enakomerna zvezna porazdelitev

Eksponentna porazdelitev

Poissonov proces

Dogodki so podintervali nekega danega intervala na realni osi (zvezen proces - v času ali prostoru). Pri tem velja:

- dogodek se lahko zgodi kjerkoli (kadarkoli) na danem intervalu
- pojavitve na disjunktnih podintervalih so med seboj neodvisne
- verjetnost pojavitve dogodka na dovolj majhnem podintervalu je premosorazmerna dolžini podintervala.

(S. Poisson, ≈1830)

Izpeljava iz Bernouillijevega procesa

Opazovani interval razdelimo na n disjunktnih podintervalov. Predstavljamo si, da uspeh v i-ti ponovitvi Bernoullijevega poskusa ustreza pojavitvi dogodka na i-tem podintervalu. Za velike vrednosti n so dolžine podintervalov majhne (zato smemo privzeti, da se na enem podintervalu lahko zgodi le en dogodek).

Izrek (Poissonov limitni izrek)

Naj bodo $\lambda \in \mathbb{R}^+$, $n \in \mathbb{N}$ *in* $p = \frac{\lambda}{n}$. *Potem je*

$$\lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} = \frac{e^{-\lambda}\lambda^k}{k!}.$$

Poissonova porazdelitev: $X \sim P(\lambda)$

$X =$ število pojavitev dogodka v Poissonovem procesu na danem intervalu

$$X : \begin{pmatrix} 0 & 1 & \cdots & k & \cdots \\ p_\lambda(0) & p_\lambda(1) & \cdots & p_\lambda(k) & \cdots \end{pmatrix} \qquad p_\lambda(k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

$$E(X) = \lambda \quad \text{in} \quad D(X) = \lambda$$

Primeri

1. Denimo, da pojav nevihte v nekem mestu sledi Poissonovem procesu. Podatki kažejo, da so v tem mestu v povprečju 4 nevihte na leto. Izračunaj
   (a) verjetnost, da v naslednjem letu ne bo nevihte,
   (b) verjetnost natanko 4 neviht v naslednjem letu.
2. V knjigi s 500 stranmi je slučajno porazdeljenih 300 tiskarskih napak. Kolikšna je verjetnost, da sta na strani 13 točno dve tiskarski napaki?

Diskretne slučajne spremenljivke (nad.) Primeri

3. Na okrogli tarči s polmerom 5 *cm* sta narisana koncentrična kroga s polmeroma 1 *cm* in 3 *cm*. Vrednost zadetkov je naslednja: notranji krog prinese 10 točk, srednji kolobar 5 točk in zunanji kolobar 3 točke. Denimo, da je verjetnost, da igralec zadane tarčo 0.5, zadetki znotraj tarče pa so enako verjetni.
   Kolikšno je pričakovano število točk na met? In standardni odklon?

Zvezne slučajne spremenljivke

X je *zvezna slučajna spremenljivka*, če je njena zaloga vrednosti $\mathcal{Z}_X \subseteq \mathbb{R}$ interval ali unija intervalov.

Za zvezno slučajno spremenljivko X velja:

- $P(x_1 < X \leq x_2) = P(x_1 \leq X \leq x_2) = P(x_1 \leq X < x_2) = P(x_1 < X < x_2)$
- $P(X = x) = 0$ za katerikoli $x \in \mathbb{R}$ (čeprav to ni nujno nemogoč dogodek!)

Gostota verjetnosti (PDF)

Gostota verjetnosti zvezne slučajne spremenljivke je takšna pozitivna integrabilna funkcija p_X, da za poljubna $x_1, x_2 \in \mathbb{R}$, $x_1 < x_2$, velja

$$P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} p_X(x)\, dx.$$

Pri tem je

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

Porazdelitvena funkcija (CDF)

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

Lastnosti F_X v primeru zvezne slučajne spremenljivke:

- F_X je povsod zvezna
- če je p_X zvezna v točki x, je F_X odvedljiva v x in velja:

$$F_X'(x) = p_X(x)$$

- veljajo vse ostale lastnosti od prej

Zgled

$$f(x) = \begin{cases} \alpha x^2, & 0 \leq x \leq 10; \\ 0, & \text{sicer.} \end{cases}$$

Določi parameter α tako, da bo f gostota neke slučajne spremenljivke X. V tem primeru izračunaj $P(X > 5)$.

Matematično upanje

Matematično upanje $E(X)$ zvezne slučajne spremenljivke X definiramo kot

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

- $E(X)$ ne obstaja vedno (nepravi integral)
- veljajo enake osnovne lastnosti kot v diskretnem primeru

Disperzija in standardni odklon

Disperzija $D(X)$ zvezne slučajne spremenljivke X je definirana s prepisom

$$\begin{aligned} D(X) &:= E\left((X - E(X))^2\right) = \int_{-\infty}^{\infty} (x - E(X))^2\, p_X(x)\, dx \\ &= E(X^2) - E(X)^2 = \int_{-\infty}^{\infty} x^2 p_X(x)\, dx - \left(\int_{-\infty}^{\infty} x p_X(x)\, dx\right)^2 \end{aligned}$$

Standardni odklon $\sigma(X)$ je enak

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

Disperzija in standardni odklon

- Velja: $D(aX + b) = a^2 D(X)$ za vse $a, b \in \mathbb{R}$.
- *standardizirana slučajna spremenljivka* je kot prej

$$Z = \frac{X - E(X)}{\sigma(X)}, \quad E(Z) = 0, \quad D(Z) = 1.$$

Zgled

Izračunaj $E(X)$ in $\sigma(X)$ za porazdelitev v prejšnjem zgledu.

Enakomerna zvezna porazdelitev

$\mathcal{Z}_X = [a, b]$, vse točke so enako verjetne

$$p_X(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b; \\ 0, & \text{sicer.} \end{cases} \quad \text{in} \quad F_X(x) = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a}, & a \le x \le b; \\ 1, & x > b. \end{cases}$$

$$E(X) = \frac{a+b}{2} \quad \text{in} \quad D(X) = \frac{(b-a)^2}{12}$$

Zgled

Opazujemo 100-kilometrski odsek avtoceste, X meri razdaljo, pri kateri pride do nesreče. Predpostavimo, da je kjerkoli na odseku nesreča enako verjetna.

- Kolikšna je verjetnost, da se nesreča zgodi med 20-tim in 35-tim kilometrom?

- Kje je pričakovano mesto nesreče? Kolikšen je strandardni odklon?

Eksponentna porazdelitev

tudi: porazdelitev Poissonovega toka
X: čakalna doba do prve pojavitve dogodka v Poissonovem procesu s parametrom λ (ali čas med dvema pojavitvama), $\mathcal{Z}_X = [0, \infty)$

$$p_X(x) = \lambda e^{-\lambda x} \quad \text{in} \quad F_X(x) = 1 - e^{-\lambda x}$$

$$E(X) = \frac{1}{\lambda} \quad \text{in} \quad D(X) = \frac{1}{\lambda^2}$$

Zgled

Denimo, da je razdalja med dvema velikima razpokama na cesti porazdeljena eksponentno in je v povprečju enaka 5 *km*.

- Kolikšna je verjetnost, da na razdalji 10 *km* ni velike razpoke?
- Kolikšna je verjetnost, da se prva velika razpoka pojavi med 12 in 15-im kilometrom od začetka opazovanja?
- Kolikšna je verjetnost, da na dveh ločenih 5 *km* odsekih ni velike razpoke?
- Denimo, da v prvih 5 *km* ni razpoke. Kolikšna je verjetnost, da je tudi v naslednjih 10 *km* ni?

Današnja snov

Zvezne slučajne spremenljivke (nad.)

Normalna ali Gaussova porazdelitev

Lognormalna porazdelitev

Porazdelitev hi-kvadrat

Studentova t-porazdelitev

Neenakost Čebiševa

Primeri

Limitni izreki

Normalna ali Gaussova porazdelitev: $X \sim N(\mu, \sigma)$

- zvezna aproksimacija binomske porazdelitve
- de Moivre(1733), Gauss skoraj 100 let kasneje
- najpogosteje uporabljana, opisuje mnogo naravnih procesov (centralni limitni izrek!)
- $\mathcal{Z}_X = (-\infty, \infty)$

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

$$E(X) = \mu \quad \text{in} \quad D(X) = \sigma^2$$

Standardizirana normalna porazdelitev: $Z \sim N(0,1)$

porazdelitvena funkcija je *Gaussov integral*:

$$F_Z(z) = \Phi(z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\, dt \to \text{ tabele}$$

$$P(a < Z \leq b) = \Phi(b) - \Phi(a)$$

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

$$P(a < X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Kvantile, percentile

X slučajna spremenljivka, $0 < p < 1$
p-ta kvantila (ali *100p-ta percentila*) zvezne porazdelitve X je taka točka x_p, da velja:

$$P(X \leq x_p) = p.$$

Posebej:

- *mediana* je $x_{0.5}$
- *kvartile* so $x_{0.25}$, $x_{0.5}$ in $x_{0.75}$
- *decile* so $x_{0.1}, x_{0.2}, \ldots, x_{0.9}$
- *interkvartilni razpon (IQR)*: $x_{0.75} - x_{0.25}$
- pravilo $68 - 95 - 99.7$ (za normalno porazdelitev)

Zgled

Tlačno trdnost vzorca betona modeliramo z normalno porazdelitvijo z $\mu = 6000\, kg/cm^2$ in $\sigma = 100\, kg/cm^2$.

- Kolikšna je verjetnost, da je trdnost vzorca manj kot $6250\, kg/cm^2$?
- Kolikšna je verjetnost, da je trdnost vzorca med 5800 in $5900\, kg/cm^2$?
- Kolikšno trdnost preseže 95% vzorcev?

Lognormalna porazdelitev

Postavimo $X = e^Y$, kjer je $Y \sim N(\mu, \sigma)$. Dobimo: $\mathcal{Z}_X = (0, \infty)$,

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

$$F_X(x) = F_Y(\ln x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right),$$

$$E(X) = e^{\mu + \frac{\sigma^2}{2}} \quad \text{in} \quad D(X) = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right).$$

Porazdelitev hi-kvadrat: $X \sim \chi^2(n)$

$\mathcal{Z}_X = (0, \infty)$, $n =$ *število prostostnih stopenj,*

$$p_X(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \quad \text{kjer je} \quad \Gamma(r) = \int_0^\infty t^{r-1} e^{-t} dt.$$

Velja:

$$E(X) = n \quad \text{in} \quad D(X) = 2n.$$

- če $X \sim N(0,1)$, potem je $X^2 \sim \chi^2(1)$
- če $X_i \sim N(0,1)$ neodvisne, potem je $\sum_{i=1}^n X_i^2 \sim \chi^2(n)$
- pri velikih n se približuje normalni porazdelitvi

Studentova t-porazdelitev

$\mathcal{Z}_T = (-\infty, \infty)$, $n =$ *število prostostnih stopenj,*

$$p_T(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(\frac{n}{n+x^2}\right)^{\frac{n+1}{2}}$$

- za $n > 1$ je $E(T) = 0$, za $n > 2$ je $D(T) = \frac{n}{n-2}$
- če $Z \sim N(0,1)$ in $X \sim \chi^2(n)$ neodvisni, potem je $T = \frac{Z}{\sqrt{\frac{X}{n}}}$
- pri velikih n se približuje standardni normalni porazdelitvi

Neenakost Čebiševa

Izrek (Čebišev)

Če za slučajno spremenljivko X obstajata matematično upanje in disperzija, potem za $k > 0$ velja

$$P\big[E(X) - k\sigma(X) < X < E(X) + k\sigma(X)\big] \geq 1 - \frac{1}{k^2}$$

Zgled

1000-krat vržemo kovanec. Oceni vejetnost, da bo število grbov med 400 in 600.

Primeri

1. Poišči taka $a, b \in \mathbb{R}$ da bo $F(x) = a + b \arctan \frac{x}{2}$ porazdelitvena funkcija zvezne slučajne spremenljivke X ter skiciraj dobljeno funkcijo. Poišči gostoto verjetnosti za ta primer in izračunaj $P(|X| < 2)$. Ali lahko izračunamo $E(X)$?

2. Vzorci betona imajo povprečno tlačno trdnost 6000 kg/cm^2 s standardnim odklonom $\sigma = 100\, kg/cm^2$. Oceni pri kolikšnem deležu vzorcev je vrednost med 5800 in 6200 kg/cm^2.

3. Letna količina padavin v zbiralniku je porazdeljena lognormalno s povprečjem 60 in standardnim odklonom 15.
   - Kolikšna je verjetnost, da bo v naslednjem letu količina padavin med 40 in 70?
   - In verjetnost, da bo letna količina vsaj 30?
   - Izračunaj še prvi decil letne količine vode v zbiralniku (t. j. vrednost, pri kateri bo skupna verjetnost 10%).

4. Reši zgornjo nalogo še za primer normalne porazdelitve.

5. Zvezna slučajna spremenljivka X ima gostoto verjetnosti

$$p_X(x) = \begin{cases} \frac{x}{2}, & 0 < x < 2, \\ 0, & \text{sicer.} \end{cases}$$

   Poišči porazdelitveno funkcijo slučajne spremenljivke $Y = 1/X$.

6. Ugotovi, kako je porazdeljena slučajna spremenljivka $W = Z^2$, če je $Z \sim N(0, 1)$.

Zakon velikih števil

Naj bo $X_1, X_2, \ldots$ zaporedje neodvisnih slučajnih spremenljivk, pri čemer so vse X_i enako porazdeljene kot spremenljivka X.

$$\overline{X}_n := \frac{1}{n}\left(X_1 + X_2 + \cdots + X_n\right), n = 1, 2, \ldots$$

Potem za za vsak $\varepsilon > 0$ velja

$$\lim_{n\to\infty} P\left(\left|\overline{X}_n - E(X)\right| < \varepsilon\right) = 1.$$

(Bernoulli, Markov, Čebišev, Borel, Cantelli, Kolmogorov. . .)

Centralni limitni zakon

$$\begin{aligned} S_n &= X_1 + X_2 + \cdots + X_n \\ Z_n &= \frac{S_n - E(S_n)}{\sigma(S_n)} \end{aligned}$$

Za zaporedje $X_1, X_2, \ldots$ velja *centralni limitni zakon*, če

$$\lim_{n\to\infty} P(Z_n \leq x) = \Phi(x).$$

Izrek (Centralni limitni izrek)

Če so X_i neodvisne, enako porazdeljene z $E(X_i) < \infty$ in $D(X_i) < \infty$, potem zanje velja centralni limitni zakon.

(Laplace, de Moivre)

Današnja snov

Slučajni vektorji

Diskretni slučajni vektorji

Pogojne porazdelitve

Neodvisnost slučajnih spremenljivk

Zvezni slučajni vektorji

Pogojne porazdelitve

Neodvisnost slučajnih spremenljivk

Funkcije slučajnega vektorja

Slučajni vektorji

Slučajni vektor je n-terica slučajnih spremenljivk

$$X = (X_1, X_2, \dots, X_n) : \mathcal{S} \to \mathbb{R}^n$$

Porazdelitveno funkcijo slučajnega vektorja (CDF) X definiramo kot

$$F_X(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) : \mathbb{R}^n \to [0, 1]$$

Omejimo se na primer $n = 2$ (*dvorazsežne ali bivariatne porazdelitve*).

Lastnosti porazdelitvene funkcije slučajnega vektorja (X, Y)

1. za vsako spremenljivko je naraščajoča in z desne zvezna
2. $0 \leq F_{(X,Y)}(x,y) \leq 1$ za vse $x, y \in \mathbb{R}$
3. $F_{(X,Y)}(-\infty, -\infty) = 0$ in $F_{(X,Y)}(+\infty, +\infty) = 1$
4. $F_{(X,Y)}(-\infty, y) = F_{(X,Y)}(x, -\infty) = 0$ za vse $x, y \in \mathbb{R}$
5. $F_{(X,Y)}(+\infty, y) = F_Y(y)$ in $F_{(X,Y)}(x, +\infty) = F_X(x)$ imenujemo *robni porazdelitvi* slučajnega vektorja (X, Y)
6. $P\big((X,Y) \in (a,b] \times (c,d]\big) =$ $F_{(X,Y)}(b,d) - F_{(X,Y)}(a,d) - F_{(X,Y)}(b,c) + F_{(X,Y)}(a,c)$

Diskreten slučajni vektor

(X, Y) je *diskreten slučajni vektor*, kadar njegovo zalogo vrednosti predstavlja števno mnogo točk.

Verjetnostna funkcija slučajnega vektorja (PDF)

$$p_{ij} := P(X = x_i, Y = y_j), \quad i, j \in \mathbb{N}$$

Velja

$$p_{ij} \geq 0 \quad \text{in} \quad \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{ij} = 1$$

Verjetnostna shema

$X \setminus Y$	y_1	y_2	$\cdots$	p_X
x_1	p_{11}	p_{12}	$\cdots$	q_1
x_2	p_{21}	p_{22}	$\cdots$	q_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
p_Y	r_1	r_2	$\cdots$	1

Robni porazdelitvi spremeljivk X in Y: verjetnostni funkciji p_X in p_Y dobimo kot

$$q_i = P(X = x_i) = \sum_{j=1}^{\infty} p_{ij} \quad \text{in} \quad r_j = P(Y = y_j) = \sum_{i=1}^{\infty} p_{ij}$$

Zgled 1

Vržemo 2 kocki. Označimo slučajni spremenljivki:

X : večje od števila pik na obeh kockah,
Y : vsota pik na obeh kockah.

Zapiši verjetnostno shemo slučajnega vektorja (X, Y) in robnih porazdelitvi spremenljivk X in Y.

Pogojne porazdelitve

Za $r_k \neq 0$ definiramo *pogojno verjetnostno funkcijo $p_{i|k}$* slučajne spremenljivke X pri pogoju $Y = y_k$ kot

$$p_{i|k} := P(X = x_i | Y = y_k) = \frac{p_{ik}}{r_k}$$

Velja

$$p_{i|k} \geq 0 \quad \text{in} \quad \sum_{i=1}^{\infty} p_{i|k} = 1$$

Zgled 1 (nad.)

Zapiši verjetnostni shemi za pogojni porazdelitvi spremenljivke X pri pogojih $Y = 4$, $Y = 8$ in $Y = 12$.

Pogojno matematično upanje

Pogojno matematično upanje $E(X|y_k)$ je matematično upanje pogojne porazdelitve:

$$E(X|y_k) := \sum_{i=1}^{\infty} x_i p_{i|k} = \frac{1}{r_k} \sum_{i=1}^{\infty} x_i p_{ik}$$

Velja

$$E(X) = \sum_{k=1}^{\infty} E(X|y_k) r_k$$

Zgled 1 (nad.)

Izračunaj pogojna matematična upanja za porazdelitve iz zadnjega zgleda.

Neodvisnost slučajnih spremenljivk

Diskretni slučajni spremenljivki X in Y imenujemo *neodvisni*, če velja eden od ekvivalentnih pogojev

$$
\begin{aligned}
F_{(X,Y)}(x,y) &= F_X(x)F_Y(y) \text{ za poljubna } x,y \in \mathbb{R} \\
\iff & \quad P(X=x, Y=y) = P(X=x)P(Y=y) \;\; \forall x,y \in \mathbb{R} \\
\iff & \quad p_{ij} = q_i r_j \text{ za vse } i,j \in \mathbb{N} \\
\iff & \quad p_{i|k} = q_i \text{ za vse } i,k \in \mathbb{N}
\end{aligned}
$$

Zgled 1 (nad.)

Ali sta slučajni spremenljivki X in Y neodvisni?

Zvezen slučajni vektor

(X,Y) je *zvezen slučajni vektor*, če sta X in Y zvezni slučajni spremenljivki. Zaloga vrednosti:

$$\mathcal{Z}_{(X,Y)} = \mathcal{Z}_X \times \mathcal{Z}_Y \subseteq \mathbb{R}^2.$$

Gostota verjetnosti (PDF) slučajnega vektorja (X,Y)

je takšna pozitivna integrabilna funkcija $p_{(X,Y)}$, da je

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

Ne obstaja vedno!

Lastnosti $p_{(X,Y)}$

- $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{(X,Y)}(u,v)\,du\,dv = 1$$
- $$P\left((X,Y) \in (a,b] \times (c,d]\right) = \int_a^b \int_c^d p_{(X,Y)}(u,v)\,du\,dv$$
- $$p_{(X,Y)}(x,y) = \frac{\partial^2 F_{(X,Y)}(x,y)}{\partial x \partial y}$$
- *robni gostoti vejetnosti*:

$$p_X(x) = \int_{-\infty}^{\infty} p_{(X,Y)}(x,y)\,dy \quad \text{in} \quad p_Y(y) = \int_{-\infty}^{\infty} p_{(X,Y)}(x,y)\,dx$$

Zgled 2

$$f(x,y) = \begin{cases} xe^{-xy}, & x \geq 0, y \geq 1; \\ 0, & \text{sicer.} \end{cases}$$

Preveri, da je $f(x,y)$ gostota verjetnosti nekega zveznega slučajnega vektorja in poišči robni gostoti.

Pogojne porazdelitve

Za $p_Y(y) \neq 0$ definiramo *pogojno gostoto verjetnosti $p_{X|y}(x)$* slučajne spremenljivke X pri pogoju $Y = y$ kot

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

- $\int_{-\infty}^{\infty} p_{X|y}(x)\,dx = 1$
- $P\left(a \leq X \leq b \mid Y = y\right) = \int_a^b p_{X|y}(x)\,dx$

Zgled 2 (nad.)

Zapiši pogojno gostoto verjetnosti slučajne spremenljivke X pri pogoju $Y = 2$. Koliko je $P(0 \leq X \leq 1 \mid Y = 2)$?

Pogojno matematično upanje

Pogojno matematično upanje $E(X \mid Y = y)$ je matematično upanje pogojne porazdelitve:

$$E(X \mid Y = y) := \int_{-\infty}^{\infty} x\, p_{X|y}(x)\,dx = \frac{1}{p_Y(y)} \int_{-\infty}^{\infty} x\, p_{(X,Y)}(x,y)\,dx$$

Velja

$$E(X) = \int_{-\infty}^{\infty} E(X \mid Y = y) p_Y(y)\,dy$$

Pogojno matematično upanje $E(X \mid Y = y)$ zapisano kot funkcija y se imenuje tudi *regresijska enačba*.

Neodvisnost slučajnih spremenljivk

Zvezni slučajni spremenljivki X in Y imenujemo *neodvisni*, če velja eden od ekvivalentnih pogojev

$$\begin{aligned} F_{(X,Y)}(x,y) &= F_X(x)F_Y(y) \text{ za poljubna } x,y\in\mathbb{R} \\ \iff \quad & p_{(X,Y)}(x,y) = p_X(x)p_Y(y) \text{ za vse } x,y\in\mathbb{R} \\ \iff \quad & p_{X|y}(x) = p_X(x) \text{ za vse } x,y\in\mathbb{R} \end{aligned}$$

Zgled 2 (nad.)

Ali sta slučajni spremenljivki X in Y neodvisni?

Funkcije slučajnega vektorja

Naj bo $(X,Y):\mathcal{S}\to\mathbb{R}^2$ slučajni vektor in $g:\mathbb{R}^2\to\mathbb{R}$ preslikava. Potem je

$$Z(s) := g\left(X(s),Y(s)\right),\ s\in\mathcal{S}$$

funkcija slučajnega vektorja (X,Y). Pišemo: $Z=g(X,Y)$.

Za diskreten slučajni vektor (X,Y) velja:

$$\begin{aligned} p(Z=z_k) &= \sum_{i,j:g(x_i,y_j)=z_k} p_{ij} \\ E(Z) &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} g(x_i,y_j)p_{ij} \end{aligned}$$

Funkcije slučajnega vektorja (nad.)

Za zvezen slučajni vektor (X, Y) velja:

$$\begin{aligned} F_Z(z) &= \iint_{g(x,y)\leq z} p_{(X,Y)}(x,y)\, dx\, dy \\ E(Z) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)p_{(X,Y)}(x,y)\, dx\, dy \end{aligned}$$

Pomembni primeri: $X+Y$, XY, $\frac{X}{Y}$

Zgled: Dvorazsežna enakomerna porazdelitev

Naj bo G neko območje v ravnini s končno ploščino $S(G)$. Zvezni slučajni vektor (X, Y) je *dvorazsežno enakomerno porazdeljen*, če

$$p_{(X,Y)}(x,y) = \begin{cases} \frac{1}{S(G)}, & (x,y) \in G; \\ 0, & \text{sicer.} \end{cases}$$

- Denimo, da je G trikotnik z oglišči $(0,0)$, $(1,0)$ in $(1,1)$. Ali sta X in Y porazdeljeni enakomerno zvezno? Ali sta neodvisni?
- Naj bosta X in Y neodvisni enakomerno zvezno porazdeljeni slučajni spremenljivki, X na intervalu $[0,2]$ in Y na intervalu $[0,1]$. Pokaži, da je slučajni vektor (X, Y) dvorazsežno enakomerno porazdeljen. Poišči gostoto verjetnosti in matematično upanje slučajne spremenljivke $W = XY$. Izračunaj še $P(X < Y)$.

Današnja snov

Slučajni vektorji (nad.)

Kovarianca in korelacija

Dvorazsežna normalna porazdelitev

Primeri

Momenti

Kovarianca

Trditev

Če sta X in Y neodvisni slučajni spremenljivki, velja:

$$E(XY) = E(X)E(Y) \quad \text{in} \quad D(X+Y) = D(X) + D(Y).$$

Definicija

Naj imata slučajni spremenljivki X in Y končni matematični upanji in končni disperziji. Kovarianco slučajnih spremenljivk X in Y definiramo kot

$$Cov(X,Y) = \sigma_{XY} = K(X,Y) := E\left((X - E(X))(Y - E(Y))\right).$$

Lastnosti kovariance

- $Cov(X, X) = D(X)$
- $Cov(X, Y) = Cov(Y, X)$
- $Cov(a + bX, c + dY) = bdCov(X, Y)$, $a, b, c, d \in \mathbb{R}$
- $Cov(X, Y) = E(XY) - E(X)E(Y)$
- $D(X + Y) = D(X) + D(Y) + 2Cov(X, Y)$
- $|Cov(X, Y)| \leq \sigma(X)\sigma(Y)$

Korelacija

$$\rho(X, Y) = Corr(X, Y) = r(X, Y) := \frac{Cov(X, Y)}{\sigma(X)\sigma(Y)}$$

imenujemo *korelacijski koeficient* slučajnih spremenljivk X in Y.

- $-1 \leq \rho(X, Y) \leq 1$
- $\rho(X, Y) = 0$: X in Y sta *nekorelirani* slučajni spremenljivki; v nasprotnem primeru sta X in Y *korelirani*
- $|\rho(X, Y)| = 1$: X in Y sta *linearno odvisni*

Če sta X in Y neodvisni, sta tudi nekorelirani. Obrat ne velja!

Matematično upanje in kovariančna matrika

Matematično upanje slučajnega vektorja $X = (X_1, \ldots, X_n)$ definiramo kot

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

Razpršenost (disperzijo) opisuje *kovariančna ali disperzijska matrika* $K = (K_{ij})_{i,j=1}^n$ z elementi

$$K_{ij} = Cov(X_i, X_j),\ i, j = 1, \ldots, n.$$

- $K_{ii} = D(X_i)$ za vse $i = 1, \ldots, n$
- $K_{ji} = K_{ij}$ za vse $i, j = 1, \ldots, n$
- vse lastne vrednosti K so nenegativne

Dvorazsežna normalna porazdelitev

Slučajni vektor (X, Y) je porazdeljen *dvorazsežno normalno* s parametri $\mu_X, \mu_Y \in \mathbb{R}$, $\sigma_X > 0$, $\sigma_Y > 0$ in $-1 < \rho < 1$, če ima gostoto verjetnosti enako

$$p_{(X,Y)}(x, y) =$$
$$= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)}$$

$$\mathcal{Z}_{(X,Y)} = \mathbb{R}^2$$

Lastnosti

- robni porazdelitvi: $X \sim N(\mu_X, \sigma_X)$ in $Y \sim N(\mu_Y, \sigma_Y)$
- pogojni porazdelitvi sta normalni
- regresijski krivulji sta premici - *linearna regresija*
- korelacijski koeficient $\rho(X, Y) = \rho$
- $p_{(X,Y)}(x, y) = p_X(x)p_Y(y) \iff \rho = 0$

Za normalno porazdeljeni slučajni spremenljivki X in Y torej velja:
neodvisni $\iff$ nekorelirani!

Standardna dvorazsežna normalna porazdelitev

Slučajni vektor (X, Y) je porazdeljen *standardno dvorazsežno normalno* s parametrom $-1 < \rho < 1$, če je porazeljen dvorazsežno normalno s parametri

$$\mu_X = \mu_Y = 0 \quad \text{n} \quad \sigma_X = \sigma_Y = 1.$$

- $X, Y \sim N(0, 1)$
- pogojna porazdelitev $Y\,|_{X=x} \sim N(\rho x, \sqrt{1-\rho^2})$
- pogojna porazdelitev $X\,|_{Y=y} \sim N(\rho y, \sqrt{1-\rho^2})$
- regresijski krivulji sta premici: $E(Y\,|_{X=x}) = \rho x$ in $E(X\,|_{Y=y}) = \rho y$

1.

Slučajna vektorja (X, Y) in (X', Y') imata verjetnostni shemi

$X \setminus Y$	4	10	p_X
1	1/4	1/4	1/2
3	1/4	1/4	1/2
p_Y	1/2	1/2	1

$X' \setminus Y'$	4	10	$p_{X'}$
1	0	1/2	1/2
3	1/2	0	1/2
$p_{Y'}$	1/2	1/2	1

Ugotovi, ali sta X in Y oziroma X' in Y' korelirani slučajni spremenljivki. Zapiši tudi kovariančni matriki.

2.

Naj bo slučajni vektor (X, Y) porazdeljen enkomerno zvezno na štirikotnem območju, ki ga določajo oglišča $(-1, 0)$, $(0, 1)$, $(1, 0)$ in $(0, -1)$. Pokaži, da sta slučajni spremenljivki nekorelirani a odvisni.

3.

Konzolni nosilec je na razdaljah a in $2a$ od stene obtežen s tovoroma, katerih teži sta neodvisni slučajni spremenlivki S_1 in S_2. Strižna sila Q in navor M sta potem tudi slučajni spremenljivki oblike

$$Q = S_1 + S_2 \quad M = aS_1 + 2aS_2.$$

Denimo, da poznamo $E(S_1) = \mu_1$, $\sigma(S_1) = \sigma_1$, $E(S_2) = \mu_2$ in $\sigma(S_2) = \sigma_2$.

(a) Poišči $E(Q)$, $E(M)$, $\sigma(Q)$ in $\sigma(M)$.
(b) Ali sta Q in M korelirani? Zapiši njuno kovariančno matriko.
(c) Izračunaj njun korelacijski koeficient v primeru, ko je $\sigma_1 = \sigma_2$.

Momenti višjega reda

Moment reda k glede na točko $a \in \mathbb{R}$ definiramo kot

$$m_k(a) := E\left((X-a)^k\right)$$

- $a = 0$: *začetni moment* $z_k = m_k(0)$ (primer: $z_1 = E(X)$)
- $a = E(X)$: *centralni moment* $m_k = m_k(E(X))$ (primer: $m_2 = D(X)$)
- *koeficient asimetrije* γ_1:

$$\gamma_1(X) = A(X) := \frac{m_3}{\sigma^3(X)}$$

- *koeficient sploščenosti ali kurtosis* γ_2:

$$\gamma_2(X) := \frac{m_4}{\sigma^4(X)}$$

Vsebina

Statistika

realen svet	←--	odločanje, načrtovanje
↓ *vzorčenje*		↑ *preskušanje domnev*
zbiranje podatkov		ocene verjetnosti dogodkov
↓ *ocenjevanje parametrov*		↑ *verjetnostni račun*
izbira verjetnostnega modela	⟶	teoretični model
opisna statistika		**inferenčna statistika**

Pomembni izrazi

- *(statistična) enota*= predmet proučevanja
- *populacija* $\mathcal{P}$ = množica vseh enot
- *vzorec*: $a_1, \ldots, a_n \in \mathcal{P}$
- *(statistična) spremenljivka* X = lastnost enote, ki nas zanima $X(a_i) = X_i, i = 1, \ldots n$ (opisne ali številske spremenljivke)
- *parameter* θ: značilnost populacije (ustreza porazdelitvi X)
- *(vzorčna) statistika*: $Y = f(X_1, \ldots, X_n)$ meri značilnosti vzorca
- *porazdelitev vzorčnih ocen* = verjetnostna porazdelitev statistike
- *cenilka* $\hat{\theta}$ parametra θ je vzorčna statistika, katere porazdelitev je odvisna le od θ

Vzorčenje

Vzorcu ustreza slučajni vektor $(X_1, \ldots, X_n)$.
Vzorec je

- *enostaven slučajni*, če so X_i med seboj neodvisne in enako porazdeljene slučajne spremenljivke
- *reprezentativen* (t. j. dobro predstavlja populacijo), če je
  - dovolj velik in
  - nepristransko izbran (problem: priročni vzorci, prostovoljni vzorci)

Pomembno: dobro načrtovanje eksperimentov.
Utemeljitelj sodobne statistike: Sir Ronald A. Fisher (1890 - 1962)

Urejanje in prikazovanje podatkov

- tabela
- frekvenčna tabela
- histogram
- krožni prikaz
- poligon
- prikaz s škatlami
- ...

Vzorčno povprečje $\overline{X}$

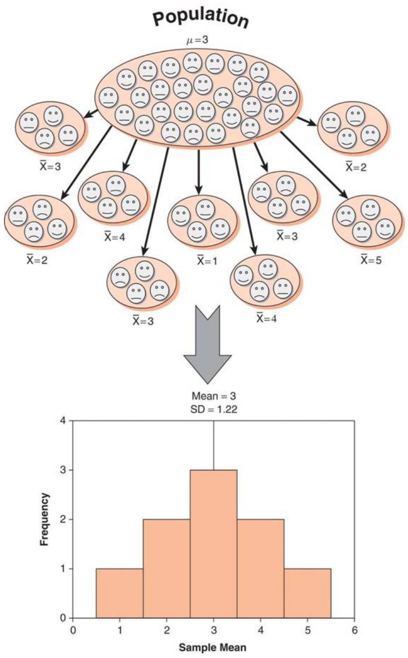
$$\overline{X} := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Izrek

Naj $(X_1, \ldots, X_n)$ *prestavlja enostaven slučajni vzorec z* $E(X_i) = \mu$ *in* $D(X_i) = \sigma^2$, $i = 1, \ldots, n$. *Potem velja*

$$E(\overline{X}) = \mu \quad \textit{in} \quad D(\overline{X}) = \frac{\sigma^2}{n}.$$

Zgled



Slika: A.Field, Discovering statistics using SPSS

Porazdelitev vzorčnega povprečja

- Če je $X \sim N(\mu, \sigma)$, je tudi vzorčno povprečje normalno porazdeljeno:

$$\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

- **ne glede na osnovno porazdelitev** X je po Centralnem limitnem izreku $\overline{X}$ za velike n (dovolj že $n > 30$) blizu normalni porazdelitvi $N(\mu, \frac{\sigma}{\sqrt{n}})$ oziroma:

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

Vzorčna disperzija S^2

$$S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Izrek

Naj $(X_1, \dots, X_n)$ prestavlja enostaven slučajni vzorec z $E(X_i) = \mu$ in $D(X_i) = \sigma^2$, $i = 1, \dots, n$. Potem je $E(S^2) = \sigma^2$.

Vzorčna standardna deviacija $S := \sqrt{S^2}$

Porazdelitev vzorčne disperzije

- Velja: če je $X \sim N(\mu, \sigma)$, za porazdelitev vzorčne disperzije velja
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$
- Po Centralnem limitnem izreku je za velike n porazdelitev χ^2 blizu normalni porazdelitvi $N(n-1, \sqrt{2(n-1)})$ in za vzorčno disperzijo velja
$$S^2 \sim N\left(\sigma^2, \sqrt{\frac{2}{n-1}}\sigma^2\right)$$

Standardizirano vzorčno povprečje

σ^2 ne poznamo, nadomestimo s S^2

Trditev

Naj bosta $\overline{X}$ vzorčno povrečje in S^2 vzorčna disperzija slučajnega vzorca, ki ustreza $X \sim N(\mu, \sigma)$. Potem je

$$T = \frac{\overline{X} - \mu}{\sqrt{\frac{S^2}{n}}}$$

Studentova t-porazdelitev z $(n-1)$ prostostnimi stopnjami.

(pomembno le za $n < 30$)

Vzorčna korelacija R

$(X_1, Y_1), \ldots, (X_n, Y_n)$ enostaven slučajni vzorec slučajnega vektorja (X, Y)

$$R := \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$$

($\overline{X}$ in $\overline{Y}$ sta vzorčni povprečju vzorcev za X in Y)

Ocenjevanje parametrov

- *intervalsko ocenjevanje*: poiščemo slučajni interval (t. i. interval zaupanja), ki s predpisano verjetnostjo vsebuje parameter θ
- *točkasto ocenjevanje*: parameter θ ocenimo s številsko vrednostjo cenilke (statistike) $\hat{\theta}$.

Lastnosti cenilk

- Statistika $\hat{\theta}$ je *nepristranska cenilka* za θ, če je $E(\hat{\theta}) = \theta$ in je *asimptotično nepristranska*, če je $\lim_{n\to\infty} E(\hat{\theta}) = \theta$.
- V primeru dveh nepristranskih cenilk θ_1 in θ_2 za θ je θ_1 *učinkovitejša* od θ_2, če velja: $D(\theta_1) < D(\theta_2)$.
- Standardno deviacijo $\sigma(\hat{\theta}) = SE(\hat{\theta})$ imenujemo *standardna napaka* cenilke $\hat{\theta}$.
- Statistika $\hat{\theta}$ je *dosledna cenilka*, če je za vsak $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(|\hat{\theta} - \theta| \geq \varepsilon\right) = 0.$$

Primeri cenilk

$(X_1, \ldots, X_n)$ enostaven slučajni vzorec z $E(X_i) = \mu$ in $D(X_i) = \sigma^2$

- Vzorčno povprečje $\overline{X}$ je nepristranska in dosledna cenilka parametra μ s standardno napako $SE(\overline{X}) = \frac{\sigma}{\sqrt{n}}$.
- Vzorčna disperzija S^2 je nepristranska in dosledna cenilka za σ^2 s standardno napako $SE(S^2) = \frac{\sqrt{2}\sigma^2}{\sqrt{n-1}}$.
- Vzorčna standardna deviacija S je pristranska a asimptotično nepristranska cenilka za σ.

Momenti višjega reda

Moment reda k glede na točko $a \in \mathbb{R}$ definiramo kot

$$m_k(a) := E\left((X - a)^k\right)$$

- $a = 0$: *začetni moment* $z_k = m_k(0)$ (primer: $z_1 = E(X)$)
- $a = E(X)$: *centralni moment* $m_k = m_k(E(X))$ (primer: $m_2 = D(X)$)
- *koeficient asimetrije* γ_1:

$$\gamma_1(X) = A(X) := \frac{m_3}{\sigma^3(X)}$$

- *koeficient sploščenosti ali kurtosis* γ_2:

$$\gamma_2(X) := \frac{m_4}{\sigma^4(X)}$$

Metoda momentov

- $z_k := E(X^k)$ k-ti začetni moment slučajne spremenljivke X
- *k-ti vzorčni moment* slučajnega vzorca $(X_1, \dots, X_n)$:

$$\widehat{z_k} := \frac{1}{n} \sum_{i=1}^{n} X_i^k, \quad k = 1, 2, \dots$$

- $\widehat{z_k}$ je cenilka za z_k (npr. $z_1 = E(X)$ in $\widehat{z_1} = \overline{X}$)
- če se parametri $\theta_1, \theta_2, \dots, \theta_m$ izražajo z momenti

$$\theta_i = f_i(z_1, z_2, \dots, z_m), \quad i = 1, 2, \dots, m,$$

dobimo *cenilke po metodi momentov* kot

$$\widehat{\theta_i} = f_i(\widehat{z_1}, \widehat{z_2}, \dots, \widehat{z_m}), \quad i = 1, 2, \dots, m.$$

Zgled

Naj bo $(X_1, \dots, X_n)$ enostaven slučajni vzorec. Poišči cenilke po metodi momentov za parametre navedenih porazdelitev in ugotovi ali so nepristranske.

1. normalna: $X_i : N(\mu, \sigma)$
2. Poissonova: $X_i : P(\lambda)$

Metoda največjega verjetja (R.A. Fisher)

- $(X_1, \ldots, X_n)$ enostaven slučajni vzorec, porazdelitev je podana z verjetnostno funkcijo $p(x, \theta)$, ki je odvisna od parametrov $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$
- *funkcija verjetja* je

$$L(\theta) := p(X_1, \theta) \cdot p(X_2, \theta) \cdots p(X_n, \theta)$$

- *cenilka po metodi največjega verjetja* je vrednost $\widehat{\theta}$ pri kateri doseže $L(\theta)$ maksimum
- izračunamo ga s pomočjo parcialnih odvodov, pogosto si pomagamo z logaritmiranjem: iščemo $\max \ln L(\theta)$

Zgledi

1. Poišči cenilke po metodi največjega verjetja za
   - parameter q Bernoullijeve porazdelitve
   - parametra μ in σ^2 normalne porazdelitve
2. Primerjaj obe metodi za pridobivanje cenilk na primeru enakomerne zvezne porazdelitve na intervalu $(0, a)$
3. Časovne intervale med prihodi vozil v križišče modeliramo z eksponentno porazdelitvijo. Poišči cenilko za parameter λ na podlagi izmerjenih intervalov:

$$1.2 \quad 3.0 \quad 6.3 \quad 10.1 \quad 5.2 \quad 2.4 \quad 7.1 \quad [s]$$

Vsebina

Intervalsko ocenjevanje parametrov

Zanima nas kvalitativna ocena cenilke nekega parametra. Poiščimo slučajen interval $[A, B]$ (= *interval zaupanja*), ki s predpisano verjetnostjo $1 - \alpha$ (= *stopnja zaupanja*) vsebuje parameter θ:

$$P(\theta \in [A, B]) = 1 - \alpha.$$

Število α imenujemo *stopnja tveganja*.

Interval zaupanja za μ

- $(X_1, \ldots, X_n)$ enostaven slučajni vzorec z $E(X_i) = \mu$ in $D(X_i) = \sigma^2$, $i = 1, \ldots, n$
- $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ je nepristranska cenilka za μ
- Po centralnem limitnem izreku je za velike n

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

- simetrična porazdelitev zato simetričen interval:

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$

Iskanje intervala zaupanja za μ, σ^2 poznamo

1. izberemo stopnjo tveganja α
2. določimo $z_{\alpha/2}$ (tabele, računalnik): $\phi(z_{\alpha/2}) = 1 - \alpha/2$
3. izračunamo $\overline{X}$

Interval zaupanja za μ s stopnjo zaupanja $1 - \alpha$ je potem

$$\left[\overline{X} - \frac{z_{\alpha/2}\sigma}{\sqrt{n}}, \overline{X} + \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right]$$

Posledica

Če želimo, da je z verjetnostjo $1 - \alpha$ napaka manjša od ε, mora za velikost vzorca veljati

$$n > \left(\frac{z_{\alpha/2}\sigma}{\varepsilon}\right)^2.$$

Zgled

Opazujemo naselje z 8000 stanovanji. Na vzorcu 100 stanovanj ugotovimo, da je povprečno število avtomobilov na stanovanje 1.6 s standardnim odklonom 0.8.

- Kolikšna je ocenjena standardna napaka za povprečno vrednost $\overline{X}$?
- Poišči 95% interval zaupanja za $\overline{X}$.
- Oceni skupno število avtomobilov N v naselju in standardno napako te vrednosti.
- Poišči 95% interval zaupanja za N.
- Kako velik vzorec bi morali vzeti, da bi širino gornjega intervala zmanjšali na ± 500?

Iskanje intervala zaupanja za μ, σ^2 ne poznamo

Nepoznan σ^2 nadomestimo s cenilko $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

- za velike vzorce ($n \geq 30$) postopek enak kot prej
- pri majhnem vzorcu ($n < 30$) namesto normalne vzamemo Studentovo porazdelitev z $(n-1)$ prostostnimi stopnjami

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

- določimo $t_{\alpha/2}$ da velja $P(T > t_{\alpha/2}) = \alpha/2$

Interval zaupanja za μ s stopnjo zaupanja $1 - \alpha$ je potem

$$\left[\overline{X} - \frac{t_{\alpha/2}S}{\sqrt{n}}, \overline{X} + \frac{t_{\alpha/2}S}{\sqrt{n}}\right]$$

Zgled

Na vzorcu 12 meritev smo dobili $\overline{X} = 81.282$ in $S^2 = 11.833$. Določi 90% interval zaupanja za μ.

Enostranski interval zaupanja

Včasih je smiselna le ocena v eno smer. Postopamo kot prej, le da ustrezno mejo intervala nadomestimo s $+\infty$ ali $-\infty$ ter $z_{\alpha/2}$ z z_α.

Spodnja meja zaupanja za μ s stopnjo zaupanja $1-\alpha$ je tako

$$\overline{X} - \frac{z_\alpha \sigma}{\sqrt{n}} \leq \mu$$

in *zgornja meja zaupanja* za μ s stopnjo zaupanja $1-\alpha$

$$\overline{X} + \frac{z_\alpha \sigma}{\sqrt{n}} \geq \mu.$$

Zgled

Meritve tlačne trdnosti 100 jeklenih profilov nam povedo:

$$\overline{X} = 220\ kPa \quad S = 22\ kPa$$

Kolikšna je 95% spodnja meja zaupanja za povprečno tlačno trdnost takšnih profilov? Primerjaj dobljeno vrednost s spodnjo mejo dvostranskega intervala zaupanja pri istih podatkih.

Interval zaupanja za σ^2, μ poznamo

- $(X_1, \dots, X_n)$ enostaven slučajni vzorec z $E(X_i) = \mu$ in $D(X_i) = \sigma^2$
- vzamemo cenilko $\widehat{D} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$
- če $X_i \sim N(\mu, \sigma)$ potem

$$X^2 := \frac{n\widehat{D}}{\sigma^2} \sim \chi^2(n)$$

- porazdelitev ni simetrična, poiskati moramo taka a in b, da bo $P(X^2 > b) = \alpha/2$ in $P(X^2 > a) = 1 - \alpha/2$

Interval zaupanja za σ^2 s stopnjo zaupanja $1 - \alpha$ je potem

$$\left[\frac{n\widehat{D}}{b}, \frac{n\widehat{D}}{a} \right]$$

Interval zaupanja za σ^2, μ ne poznamo

Cenilki za μ in σ^2 sta $\overline{X}$ in S^2

- če $X_i \sim N(\mu, \sigma)$ potem

$$X^2 := \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

- poiščemo taka a in b, da bo $P(X^2 > b) = \alpha/2$ in $P(X^2 > a) = 1 - \alpha/2$

Interval zaupanja za σ^2 s stopnjo zaupanja $1 - \alpha$ je potem

$$\left[\frac{(n-1)S^2}{b}, \frac{(n-1)S^2}{a}\right]$$

Interval zaupanja za σ

Izrek

Če je $[A, B]$ interval zaupanja za σ^2 s stopnjo zaupanja $1 - \alpha$, potem je $[\sqrt{A}, \sqrt{B}]$ interval zaupanja za σ z enako stopnjo zaupanja.

Zgled

Rezultati meritev:

22.2 24.7 20.9 26.0 27.0 24.8 26.5 23.8 25.6 23.9

- Poišči 99% interval zaupanja za σ^2.
- Poišči 99% spodnjo mejo zaupanja za σ^2.
- Poišči 90% spodnjo mejo zaupanja za σ.
- Denimo, da poznamo $\mu = 25$. Poišči 99% interval zaupanja za σ^2 v tem primeru.

Normalna aproksimacija binomske porazdelitve

- Zanima nas verjetnost p, da slučajen element populacije ustreza danemu kriteriju
- vzorec velikosti n, $X=$ število elementov vzorca, ki ustrezajo željenemu kriteriju
- $X \sim B(n,p)$, $E(X) = np$ in $D(X) = np(1-p)$
- če n velik v primerjavi s p (in p ne preblizu 0 ali 1) velja:

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \sim N(0,1)$$

- v praksi: aproksimacija bo dobra za $np > 5$ in $n(1-p) > 5$

Interval zaupanja za delež populacije (velik vzorec)

- Cenilka za p je *vzorčni delež $\widehat{p}$*:

$$\widehat{p} = \frac{X}{n}$$

- Velja:

$$Z = \frac{\widehat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

- določimo $z_{\alpha/2}$, da velja: $\phi(z_{\alpha/2}) = 1 - \alpha/2$
- neznani p v standardni napaki nadomestimo s $\widehat{p}$
- interval zaupanja za p s stopnjo zaupanja $1 - \alpha$ je

$$\left[\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right]$$

Zgled

V proizvodnji so preizkusili 85 izdelkov in ugotovili, da 10 izmed njih ne ustreza specifikacijam. Določi 95% interval zaupanja za delež neustreznih izdelkov.

Vsebina

Preskušanje domnev

Domneva, testna statistika, napake

Preskušanje domneve o matematičnem upanju

Preskušanje skladnosti

Preskušanje neodvisnosti

Primeri

Statistična domneva

Statistična domneva ali *hipoteza* je izjava o porazdelitvi neke slučajne spremenljivke. Ponavadi vsebuje parametre. Govori o neki lastnosti populacije.
Na podlagi vzorca želimo ugotoviti, ali je domneva pravilna.

Oznake (Neyman-Pearson):

H_0: *ničelna domneva* (želimo preskusiti)
H_A: *alternativne domneve* (vse ostale, ničelni komplemetarne)

Pri domnevah o parametrih porazdelitev ločimo *enostranske* in *dvostranske* domneve.

Testna statistika

- $(X_1, \ldots, X_n)$ slučajni vzorec
- $T : (X_1, \ldots, X_n) \to \mathbb{R}$ *testna statistika*
- $Z(T) \subseteq \mathbb{R}$: zaloga vrednosti testne statistke na danem vzorcu
- $W_0 \subseteq Z(T)$: *kritično območje testa* (H_0 zavrnemo)
- $Z(T) \setminus W_0$: *sprejemljivo območje testa* (H_0 ne zavrnemo)

Lastnosti testov

- *Napaka I. vrste*: zavrnemo pravilno domnevo H_0; maksimalno verjetnost te napake α imenujemo *stopnja značilnosti ali tveganje testa* (ponavadi jo določimo vnaprej - "100α% test").
- *Napaka II. vrste*: ne zavrnemo H_0, čeprav je napačna; verjetnost te napake β je odvisna od dejanske vrednosti parametrov, velikosti vzorca, ipd.
- *Moč testa* je verjetnost zavrnitve napačne domneve $= 1 - \beta$.
- *p-vrednost* je najmanjša stopnja značilnosti, pri kateri na podlagi vzorca H_0 še zavrnemo.

Zgled

Opazujemo vzorčno vsoto 25 elementov, porazdeljenih Bernoullijevo z nekim parametrom p. Testna statistika: $Y \sim B(25, p)$. Postavimo domnevi:

$$H_0 : p = 0.45 \qquad H_A : p > 0.45$$

1. Določi stopnjo značilnosti testa α, če je kritično območje testa: $W_0 = [16, \infty)$.
2. Denimo, da je prava vrednost parametra $p = 0.55$. Kolikšna je moč zgornjega testa?
3. Na vzorcu dobimo vrednost $y = 14$. Določi p-vrednost.

Splošni postopek

1. postavimo domnevi H_0 in H_A
2. izberemo stopnjo značilnosti α
3. določimo ustrezno testno statistiko T
4. poiščemo kritično območje W_0, da bo

$$P(T \in W_0 | H_0) = \alpha$$

5. izračunamo vrednost testne statistike T na vzorcu
6. za $T \in W_0$ domnevo H_0 zavrnemo s stopnjo značilnosti α, sicer domneve ne zavrnemo (a tudi ne sprejmemo!)

Preskušanje μ normalne porazdelitve, σ^2 poznamo

- $(X_1, \ldots, X_n)$, $X \sim N(\mu, \sigma)$
- $H_0 : \mu = \mu_0$, $H_A : \mu \neq \mu_0$
- testna statistika:
$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$
- $W_0 = (-\infty, -z_{\alpha/2}) \cup (z_{\alpha/2}, \infty)$, kjer je $\phi(z_{\alpha/2}) = 1 - \alpha/2$
- podobno za enostransko domnevo
- povezava z intervali zaupanja?

Zgled

Po specifikacijah mora trdno pogonsko gorivo za vesoljska vozila izgorevati v povprečju 50 cm/s. Standardni odklon je 2 cm/s. Testiranje 25 vzorcev je vrnilo $\overline{X} = 51.3$ cm/s. Kaj lahko s stopnjo značilnosti 0.05 rečemo o dejanski povprečni vrednosti?

Preskušanje skladnosti (χ^2-test)

H_0 je domneva o modelu porazdelitve za dano populacijo

- slučajni vzorec velikosti n razdelimo na k razredov $S_1, \ldots, S_k$
- o_i= frekvenca razreda S_i
- e_i= $nP(X \in S_i | H_0)$ = domnevna frekvenca razreda S_i
- testna statistika:
$$X_0^2 := \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$
- če je H_0 pravilna in m nepoznano število parametrov domnevane porazdelitve, za velike n (dovolj: $e_i \geq 5$) velja
$$X_0^2 \sim \chi^2(k - m - 1)$$
- $W_0 = (c_\alpha, \infty) : P(\chi^2(k - m - 1) > c_\alpha) = \alpha$
- domnevo zavrnemo, če $X_0^2 > c_\alpha$

Zgled

V tabeli so zbrani podatki o močnih deževjih na nekem območju v obdobju 66 let.

št. deževij/leto	0	1	2	3	4
št. let	20	23	15	6	2

Na podlagi χ^2-testa s 5% stopnjo značilnosti ugotovi, ali je Poissonova porazdelitev ustrezen model za opis letnega števila močnih deževij.

Preskušanje neodvisnosti (χ^2-test)

- H_0: izbrani lastnosti populacije sta (statistično) neodvisni
- vzorec velikosti n na podlagi obeh lastnosti razdelimo v r oziroma s razredov
- o_{ij}= frekvenca i-tega razreda prve in j-tega razreda druge lastnosti ($r \times s$ *kontingenčna tabela*)
- e_{ij}= domnevne frekvence razredov
- testna statistika:

$$X_0^2 := \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(o_i - e_i)^2}{e_i} \sim \chi^2((r-1)(s-1))$$

Zgled

Pivovarna pred načrtovano oglaševalsko akcijo preveri ciljno skupino svojih proizvodov. Ali je izbira vrste piva odvisnja od starosti kupcev? Preverimo to na spodnjih podatkih s stopnjo značilnosti 0.05.

vrsta piva \ starost	< 25	$25 - 45$	$45 - 65$	> 65
svetlo	25	42	33	12
temno	17	25	25	10
brezalkoholno	18	23	22	8

1. Testiranje skladnosti normalne porazdelitve

Proizvajalec želi preveriti domnevo o normalni porazdelitvi neke lastnosti proizvoda. Na vzorcu velikosti 100 dobi $\overline{X} = 5.04$ in $S = 0.08$.

Po splošni praksi določi frekvenčne razrede tako, da bodo pričakovane frekvence e_i enake za vse razrede. Izbere 8 razredov in dobi $e_i = 12.5$. Izmerjene frekvence so v tabeli na naslednji strani.

Utemelji intervale frekvenčnih razredov in ugotovi, ali s 5% stopnjo značilnosti lahko sprejmemo domnevo o normalni porazdelitvi.

Tabela za Primer 1

razredni interval	o_i
$X < 4.948$	12
$4.948 \leq X < 4.986$	14
$4.986 \leq X < 5.014$	12
$5.014 \leq X < 5.040$	13
$5.040 \leq X < 5.066$	12
$5.066 \leq X < 5.094$	11
$5.094 \leq X < 5.132$	12
$5.132 \leq X$	14

2. Preverjanje domneve o deležu populacije

Zveza potrošnikov objavi, da je vsaj 10% otroških avtosedežev na tržišču potencialno nevarnih. Testiranje 200 sedežev pokaže 16 neustreznih.

- So rezultati testiranja v skladu s trditvijo Zveze potrošnikov?
- Kako bi na isto vprašanje odgovorili s pomočjo intervalov zaupanja?

3. Olimpijske igre 2008

V tabeli na naslednji strani je prvih deset držav po skupnih medaljah na poletnih Olimpijskih igrah v Pekingu 2008. Na podlagi teh podatkov s 5% stopnjo značilnosti preveri, ali je število medalj odvisno od bruto družbenega proizvoda (BDP) države.
Ali je vzorec dovolj velik za smiselen rezultat?

Tabela: Olimpijske igre 2008

Tabela: Povzeto iz: `http://www.symworld.com/medals/`

rang	država	št. medalj	BDP (v mil. $)
1	ZDA	110	13,790,000
2	Kitajska	100	3,299,000
3	Rusija	72	1,286,000
4	Velika Britanija	47	2,773,000
5	Avstralija	46	889,700
6	Nemčija	41	3,259,000
7	Francija	40	2,515,000
8	Južna Koreja	31	981,900
9	Italija	28	2,068,000
10	Ukrajina	27	140,500