

PODROBNEJŠA IZPITNA VPRAŠANJA

0. “Ponovitev snovi iz srednje šole”

1. Opisna statistika

- Vrste slučajnih spremenljivk, primeri
- Frekvenčna porazdelitev
- Grafični prikaz podatkov
- Vzorčne statistike (povprečje, standardni odklon, mediana, kvartili), formule
- Grafi in opisna statistika glede na vrsto slučajne spremenljivke

2. Kombinatorika in preštevanje

- Pojasni pravili za vsoto in produkt, ki ju uporabljamo pri štetju
- Definiraj permutacije (in omeni kakšen primer)
- Definiraj variacije (in omeni kakšen primer)
- Definiraj kombinacije (in omeni kakšen primer)
- Definiraj permutacije s ponavljanjem
- Iz zgornje definicije izpelji posebna primera: permutacije, kombinacije
- Podaj binomski obrazec (in definiraj Pascalov trikotnik)
- Z limito vpelji število $e = 2.71 \dots$, ki predstavlja osnovo za naravni logaritem
- Podaj primer, kjer pridemo do tega števila v kombinatoriki

3. Računanje z dogodki

- Definiraj prostor dogodkov (tudi nemogoč, gotov in nasproten dogodek)
- Definiraj vsoto dveh dogodkov
- Definiraj produkt dveh dogodkov
- Naštej lastnosti za vsoto dveh dogodkov
- Naštej lastnosti za produkt dveh dogodkov
- Podaj pravilo o vključitvi/izključitvi

4. Popoln sistem dogodkov in definicije verjetnosti

- Podaj najmanjši popoln sistem dogodkov
- Kdaj je množica dogodkov popoln sistem
- Statistična definicija verjetnosti
- Klasična definicija verjetnosti (izpeljava)
- Geometrična definicija verjetnosti
- Podaj zvezo med verjetnostmi dveh dogodkov ter njunima vsoto in produktom (in jo utemelji bodisi s statistično ali geometrijsko definicijo verjetnosti)

5. Pogojna verjetnost

- Primer
- Definiraj pogojno verjetnost (s kompleksom pogojev)
- Podaj formulo, ki poveže običajno in pogojno verjetnost
- Zakaj je pomenbna grafična predstavitev z drevesom?
- Izpelji formulo za računanje pogojne verjetnosti (z uporabo statistične definicije verjetnosti in/ali geometrijske definicije verjetnosti)

6. Dvofazni poskusi

- definicija popolnega sistema dogodkov
- formula za pogojno verjetnost
- primer
- drevesna struktura
- formula za popolno verjetnost
- Bayesov obrazec
- dokaz za Bayesov obrazec in formulo za popolno verjetnost

7. Bernullijevo zaporedje neodvisnih poskusov

- primeri
- večfazni poskusi
- kombinacije
- drevesna predstavitev
- formula za računanje verjetnosti (binomska porazdelitev)
- računanje verjetnosti in/ali Laplaceov točkovni obrazec
- zveza med binomsko in normalno porazdelitvijo in/ali Poissonovo porazdelitvijo
- Dokazovanje pričakovane vrednosti z indikatorji

8. Sredine

- aritmetična sredina ($k = 1$)
- kvadratna sredina ($k = 2$)
- geometrijska sredina ($k = 0$)
- harmonična sredina ($k = -1$)
- potenčna sredina stopnje k
- neenakosti med njimi
- dokaz $A_2 \geq G_2$ (in karakterizacija enakosti)
- dokaz $K_2 \geq A_2$ (in karakterizacija enakosti)
- dokaz brez besed

9. Uporaba statistike v kriptografiji

- porazdelitev slovenskih črk
- porazdelitev angleških črk
- definicija zamenjalne/substitucijske šifre
- velikost prostora ključev
- postopek za razbijanje zamenjalne šifre
- avtomatizacija postopka razbijanja

10. Paradoks rojstnih dnevo*

- primer
- definicija trčenja
- splošna formula za velikost vzorca, da pride do trčenja z verjetnostjo $1/2$
- uporaba v kriptografiji
- izpeljava zgoraj omenjene splošne formule

1. del

1. Slučajne spremenljivke

- primeri
- definicija (zaloga vrednosti in porazdelitveni zakon)
- porazdelitvena funkcija
- lastnosti porazdelitvene funkcije
- diskretna, zvezna
- primer mešane slučajne spremenljivke

2. Diskretne slučajne spremenljivke

- primer tabele
- primeri: enakomerna, binomska, Poissonova, Pascalova, hipergeometrijska
- definicija (tabela, zaloga vrednosti je števna množica)
- porazdelitvena funkcija je stopničasta funkcija
- izpeljava netrivialne pričakovane vrednosti ali odklona

3. Poissonova porazdelitev

- primer, npr. štetje prometa
- zaloga vrednosti
- predpostavke
- definicija - verjetnostna funkcija
- povezava z binomsko porazdelitvijo
- razlaga parametra λ
- pričakovana vrednost in disperzija
- Poissonov obrazec

4. Pascalova porazdelitev

- primeri (npr. geometrijska)
- zaloga vrednosti
- definicija
- verjetnostna funkcija
- pričakovana vrednost in disperzija
- izpeljava pričakovane vrednosti za geometrijsko porazdelitev
- uporaba pri problemu "zbiranje kuponov"

5. Hipergeometrijska porazdelitev

- zaloga vrednosti
- definicija (verjetnostna funkcija)
- omejitve parametrov
- primer uporabe
- je več parametrov boljše ali slabše
- povezava z binomsko

6. Zvezne slučajne spremenljivke

- primer - slika gostote verjetnosti
- opisi primerov: enakomerna, normalna, eksponentna (in Gama), Cauchyjeva
- definicija zvezne s.s. z gostoto verjetnosti
- zveza med gostoto verjetnosti in porazdelitveno funkcijo
- računanje verjetnosti na podintervalu
- izpeljava netrivialne pričakovane vrednosti ali odklona

7. Normalna porazdelitev

- slika: unimodalna, zvonasta krivulja
- povezava z vsoto slučajnih spremenljivk in/ali CLI
- zaloga vrednosti, gostota verjetnosti, funkcija napake
- standardizacija in tabela za $N(0, 1)$
- kje sta μ in σ na grafu $y = p(x)$ in pravilo 68-95-99,7
- večrazsežna gostota porazdelitve
- vektorska oblika s kovariančno matriko

8. Bernulijev zakon velikih števil

- statistična in klasična definicija verjetnosti
- definicija funkcija napake
- podroben zapis izreka
- uporaba/primer za izračun verjetnosti ali velikost n
- dokaz z aproksimacijo prek zveze med binomsko in normalno porazdelitvijo
- Zakaj to smatramo za prvo verzijo CLI?

9. Eksponentna porazdelitev

- primer
- slika, zaloga vrednosti, gostota verjetnosti
- porazdelitvena funkcija
- definicija - povezava s Poissonovim procesom
- razlaga parametra λ
- pričakovana vrednost in disperzija
- karakterizacija: zvezna slučajna spremenljivka brez spomina

10. Gama porazdelitev

- posebni primeri (npr. eksponentna, hi-kvadrat),
kako lahko iz eksponentne pridemo do Gama porazdelitve
- slika, zaloga vrednosti, gostota verjetnosti
in kako pridemo do zgornjih primerov iz Gama porazdelitve
- Gama funkcija (definicija, rekurzija in nekatere vrednosti, tudi za $1/2$)
- Pričakovana vrednost in disperzija
- Uporaba hi-kvadrat porazdelitve
- Dedna lastnost gama porazdelitve, skica dokaza za aditivnost,
tj. $\Gamma(k, \lambda) + \Gamma(h, \lambda) = \Gamma(k + h, \lambda)$

2. del

1. Pričakovana vrednost slučajne spremenljivke (matematično upanje)

- primer: npr. igralna kocka (3.5) ali binomska (np)
- motivacija za definicijo (uteženo povprečje), tj. težišče
- definicija za diskretne slučajne spremenljivke (kdaj obstaja)
- definicija za zvezne slučajne spremenljivke (kdaj obstaja)
- primer slučajne spremenljivke za katero ne obstaja $E(X)$
- lastnosti: linearnost, z dokazom za homogenost
- Skica dokaz aditivnosti $E(X + Y) = E(X) + E(Y)$
ali Trditev 7.1: $E(|XY|) \leq \sqrt{E(X^2)E(Y^2)}$

2. Disperzija (razpršenost oz. varianca) slučajne spremenljivke, odklon in standardizacija

- primeri: npr. binomska ($np(1 - p)$), enakomerna $(b - a)^2/12$
- definicija s pričakovano vrednostjo in obstoj
- $D(X) = 0 \iff X$ je konstanta
- lastnosti (aditivnost za neodvisni slučajni spremenljivki)
- Standardizacija, slučajne spremenljivke in njena pričakovana vrednost oz. odklon
- Skica dokaza zveze “kosinusnu izrek”

3. Korelacija in kovarianca

- meri algebraično povezanost dveh številskih slučajnih spremenljivk
- definicij korelacije in njen obstoj (CS $E(|XY|) \leq \sqrt{E(X^2)E(Y^2)}$)
- lastnosti korelacije
- definicij korelacijskega koeficienta, vedno na $[-1, 1]$
- kdaj lahko zaključimo linearno odvisnost
- Ali lahko izračunamo korelacijo iz disperzij (tj. $D(X)$, $D(Y)$ in $D(X + Y)$)
- kovariančna matrika
- povezava z regresijsko premico

4. Slučajni vektorji $2D$, $3D$, nD

- definicija slučajnega vektorja (primer)
- definicija porazdelitvene funkcije (primer)
- verjetnostna in kontingenčna tabela, verjetnostna funkcija (primer)
- gostota verjetnosti (primer)
- robne porazdelitvene funkcije
- Ali se da iz verjetnostne funkcije slučajnega vektorja ugotoviti neodv. komponent?
- zveza med gostoto verjetnosti in porazdelitveno funkcijo
- definicija kvadranta in izpeljava formule pravokotnik
- neodvisnost

5. Polinomska porazdelitev

- primeri
- zaloga vrednosti
- definicija
- verjetnostna funkcija (zapiši $p_{i,j,\dots,k}$)
- povezava z binomsko
- pričakovana vrednost in disperzija

6. Funkcije slučajnih spremenljivk

- primeri enostavnih funkcij
- definicija in povezava med ustreznima porazdelitvenima funkcijama
- zveza med gostatami verjetnosti
- formula za pričakovano vrednost
- neodvisnost
- izpeljava zveze med $N(0, 1)$ in $\chi^2(1)$

7. Funkcije slučajnih vektorjev

- primer
- definicija
- definicija konvolucije
- uporaba za vsoto dveh neodvisnih normalnih porazdelitev
- uporaba za vsoto dveh neodvisnih Gama porazdelitev
- formula za pričakovano vrednost produkta
- Jacobijeva determinanta in prehod na nove spremenljivke

8. Pogojna porazdelitev

- primer za diskretni slučajni vektor
- definicija v diskretnem primeru in v zveznem primeru
- pogojna porazdelitvena funkcija v obeh primerih
- izpeljava pogojne verjetnostne funkcije v diskretnem primeru
- izpeljava pogojne gostote v zveznem primeru
- primer za zvezni slučajni vektor

9. Momenti in kvantili

- katere momente poznaš
- definicija momenta reda k glede na točko a
- začetni moment, centralni moment
- lastnosti
- definicija kvantila
- povezava z inverzom porazdelitvene funkcije

10. Studentova porazdelitev

- kako lahko pridemo do te porazdelitve, primer
- zaloga vrednosti
- kaj predstavlja njen parameter
- predpostavke
- gostota verjetnosti
- definicija beta funkcije
- posebne vrednosti beta funkcije
- katero porazdelitev dobimo za eno prostostno stopnjo in katero, če je število prostostnih stopenj dovolj veliko

11. Fisherjeva porazdelitev

- kako lahko pridemo do te porazdelitve, primer
- zaloga vrednosti
- kaj predstavljajo njeni parametri
- predpostavke
- definicija - verjetnostna funkcija
- definicija beta funkcije
- posebne vrednosti beta funkcije
- primer statistike, ki se porazdeljuje po Fisherjevo
- še kakšne lastnosti Fisherjeve porazdelitve

3. del

1. Vzorčenje in cenilke

- Definicija enostavnega slučajnega vzorca
- Vzorčne sredinske mere (modus, mediana, povprečje)
- Vzorčne mere razpršenosti (razmik, varianca, standardni odklon)
- Vzorčne mere oblike porazdelitve (koeficienta asimetrije in sploščenosti)
- Definicija cenilke
- Vpelji nepristranskost in doslednost cenilke
- CLI

2. CLI za \bar{X}

3. CLI za delež

4. CLI za S^2

5. CLI za razliko vzorčnih povprečij

6. CLI za razliko deležev

- Kako začnemo (kaj nas pravzaprav zanima)?
- predpostavke
- izbira vzorca
- slučajni vektor, ki spremlja naključno izbran vzorec

- izbira cenilke
- izrek, ki nam zagotovi normalno porazdelitev
- parameter μ te normalne porazdelitve
- parameter σ te normalne porazdelitve
- kaj je pomembno pri drugem parametru
- uporaba

7. “CLI” za kvocient S_1^2/S_2^2

- Kako začnemo (kaj nas pravzaprav zanima)?
- predpostavke
- izbira vzorca
- slučajni vektor, ki spremlja naključno izbran vzorec
- izbira cenilke
- kaj nam zagotovi Fisherjevo porazdelitev
- parameteri Fisherjeve porazdelitve
- uporaba

8. Intervali zaupanja

- definicija točkovne ocene za parameter, primeri (vsaj 3)
- pomen stopnje zaupanja
- postopek intervalskega ocenjevanja
- kdaj uporabimo Studentovo porazdelitev
- kaj je drugače, ko imamo majhen vzorec
- izbira velikosti vzorca
- izpeljava formule za interval zaupanja

9. Preverjanje domnev

- definicije (H_0 in H_1 , testna statistika, α in β , kritično območje)
- opis formalnega postopka
- definiraj moč testa
- pregled enostavnih statističnih testov
- vpelji P -vrednost in podaj primer računanja te vrednosti
- Opiši postopek preverjanje domneve o porazdelitvi slučajne spremenljivke

10. Bivariatna analiza in regresija

- preverjanje povezanosti slučajnih spremenljivk
(glede na tip slučajnih spremenljivk)
- preverjanje neodvisnosti imenskih slučajnih spremenljivk
- koeficient korelacije (Pearsonov, Spearmanov),
preverjanje domneve o parametru ρ
- linearna regresija (definicije, predpostavke, metoda najmanjših kvadratov)
- časovne vrste in definicija trenda
- Statični test linearnosti modela

SEZNAM KLJUČNIH TOČK PREDMETA OVS/VIS

- Kombinatorika (**permutacije s ponavljanjem**, variacije/kombinacije)
- **Popoln sistem dogodkov in definicije verjetnosti** (statistična, klasična in geometrijska)
- Verjetnost in računanje z dogodki (**vsota in produkt**, nasproten dogodek), vključ/izključ.
- **Pogojna verjetnost** (grafična predstavitev z drevesom), definicija, formula z izpeljavo
- Dvofazni poskusi, formula za **popolno verjetnost** (marmorni kol.) in **Bayesov obrazec**
- Bernullijevo zaporedje neodvisnih poskusov in Laplaceov točkovni obrazec
- **Slučajne spremenljivke**, porazdelitveni zakon, diskretne in zvezne, lahko pa tudi mešane
- Diskretna slučajna spremenljivka (tabela, **enakomerna** in **binomska** porazdelitev)
- Zvezne slučajne spremenljivke, **porazdelitvena funkcija** $F(x) = P(X \leq x)$ in njene lastnosti (npr. ali je lahko nepadajoča), gostota verjetnosti $p(x) = F'(x)$, ploščina
- **Normalna porazdelitev** (funkcija napake in $N(0, 1)$), pravilo 68-95-99,7, kje sta μ in σ na grafu $y = p(x)$, tudi večrazsežna gostota porazdelitve in opis kovariančne matrike

-
- Sredine (potenčna stopnje k , npr. harm./ $k = -1$, aritm./ $k = 1$, geomet./ $k = 0$, kvadrat./ $k = 2$)
 - **Pričakovana vrednost** (matično upanje) slučajne spremenljivke (kdaj obstaja!)
 - Disperzija (razpršenost oz. varianca) slučajne spremenljivke in **odklon**
 - **Standardizacija** slučajne spremenljivke (in njena pričakovana vrednost oz. odklon)
 - Povezanost dveh številskih slučajnih spremenljivk, **kovarianca**, korelacijski koeficient
 - Funkcije slučajnih spremenljivk in slučajni vektorji **2D**, **3D**, **nD** (**kontingenčna tabela**, robna porazdelitvena funkcija, verjetnostna funkcija, **neodvisnost**, pogojne porazdelitve), **polinomska porazdelitev** (definicija: npr. kaj je zaloga vrednosti, zapiši $p_{i,j,\dots,k}$)
 - **Centralni limitni izrek** (ni dovolj napisati, da gre za CLI, pač pa je potrebno razumeti, kaj je centralni limitni zakon) in neenakost Čebiševa

-
- Opisna statistika (koraki statistične analize ter urejanje in prikazovanje podatkov: npr. škatla z brki in **histogram**; mediana, **kvantil**, kvartil in kvartilni razmik, vrste spremenljivk)
 - Mere razpršenosti (odklon, varianca, standardni odklon) in asimetrije ter sploščenosti, **momenti** (centralni, začetni)
 - Porazdelitve vzorčnih statistik (povprečja, deležev, razlike povprečij in razlike deležev)
 - Sklepanje iz vzorca na populacijo, **vzorčne statistike**, **cenilke** (nepristranske in dosledne)
 - **Intervali zaupanja**, ocenjevanje parametrov (povprečje, delež, odklon, tudi za razliko oz. kvocijent), majhen in velik vzorec (porazdelitve **Studentova**, **Fisherjeva** in **hi-kvadrat**)
 - **Preverjanje domnev** (tj. testiranje hipotez): alternativna domneva, stopnja zaupanja/tveganja, napake (1., 2.), stopnja značilnosti testa, **P-vrednost**, kritično območje, **formalen postopek**
 - Regresija (**regresijska premica** z metodo najmanjših kvadratov) in časovne vrste, **trend**

Za vsak slučaj ponovite še enkrat CLI in njegovo zgodbo pri vzorčenju

CLI IN NJEGOVA ZGODBA PRI VZORČENJU

1. Opazujemo določeno lastnost (neke populacije velikosti N), ki jo spremlja spremenljivka X .
2. Radi bi ocenili pričakovano vrednost $E(X)$,
pri čemer predpostavimo $D(X) < \infty$ (oznaki $\mu = E(X)$ in $\sigma^2 = D(X)$).
3. V ta namen si izberemo vzorec velikosti $n \ll N$ (nimamo namreč možnosti, da bi opravili meritve na celotni populaciji ali pa je to enostavno predrago):
 - (a) vzorec je vektor (x_1, \dots, x_n) , sestavljen iz n meritev,
 - (b) izberemo ga naključno, meritve in izbira pa so med seboj neodvisne,
 - (c) vzorec mora biti dovolj velik (npr. vsaj $n \geq 30$).
4. Naj bo X_i ($1 \leq i \leq n$) spremenljivka, ki spremlja i -to meritev.
Lahko predpostavimo, da ima enako porazdelitev kot X ter zato tudi

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

5. Iščemo dobro formulo oz. funkcijo $f(x_1, \dots, x_n)$ (cenilko), za katero velja, da je simetrična in da bo njena vrednost na danem vzorcu z dovolj veliko verjetnostjo blizu $E(X)$.
V primeru $E(X)$ je to vzorčno povprečje (ki je tudi slučajna spremenljivka):

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

6. CLI nam zagotavlja, da se \bar{X} porazdeljuje normalno, tj. $\bar{X} \sim N(\mu_{\bar{X}}, \sigma_{\bar{X}})$ (ob predpostavkah: $n \geq 30$ in $\sigma < \infty$!). V praksi nas zanima, koliko sta parametra $\mu_{\bar{X}}, \sigma_{\bar{X}}$!
7. Iz lastnosti pričakovane vrednosti (linearnost) izračunamo $\mu_{\bar{X}} = E(\bar{X}) = \dots = \mu$ (kar pomeni, da je cenilka \bar{X} nepristranska).
8. Iz lastnosti odklona (linearnost za nekorelirane slučajne spremenljivke) izračunamo $\sigma_{\bar{X}}^2 = D(\bar{X}) = \dots = \sigma^2/n$ ("..." v točkah 7. in 8. je potrebno znati dopolniti).

Povzemimo (točke 6-8): za vzorčno povprečje vemo naslednje

$$\boxed{\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$

in ne pozabimo, da za računanje verjetnosti uporabljamo tabelirano funkcijo napake za $N(0, 1)$, zato pa vpeljemo še standardizirano slučajno spremenljivko $Z = (\bar{X} - \mu_{\bar{X}})/\sigma_{\bar{X}}$.

Mimogrede: kje smo zgoraj uporabili $D(X + Y) = D(X) + D(Y) + 2 \text{Cov}(X, Y)$

(asociacija na $(a + b)^2 = a^2 + 2ab + b^2$ oz. Pitagorjev/kosinusni izrek, glej skripto, razdelki 7.1-4)?

To je bil Centralni limitni izrek (CLI) za μ (glej razdelek 10.3, konkretno Trditev 10.1 in za tem še CLI prirejen za \bar{X} , opisan zgoraj; tudi Izrek 8.1 in razdelek 8.3).

Kaj pa **CLI za delež π** (glej 11.7.2), **CLI za odklon σ** (glej 11.4.2) ali **CLI za celo razliko pričakovanih vrednosti $\mu_1 - \mu_2$** (glej 11.7.3) ali **CLI za deležev $\pi_1 - \pi_2$** (glej 11.7.4), da ne govorimo o **CLI za σ_1/σ_2** ? Odgovore na slednje vprašanje najdete v (11.) poglavju o intervalih zaupanja (kot smo že nakazali/precizirali v oklepajih).