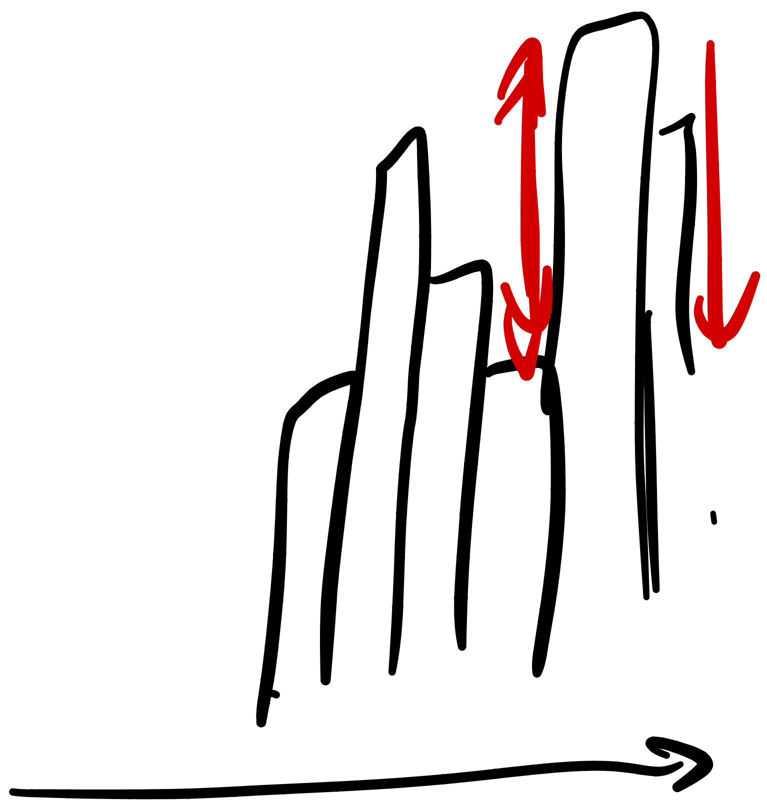
Literatura : spletna učilnica : skripta

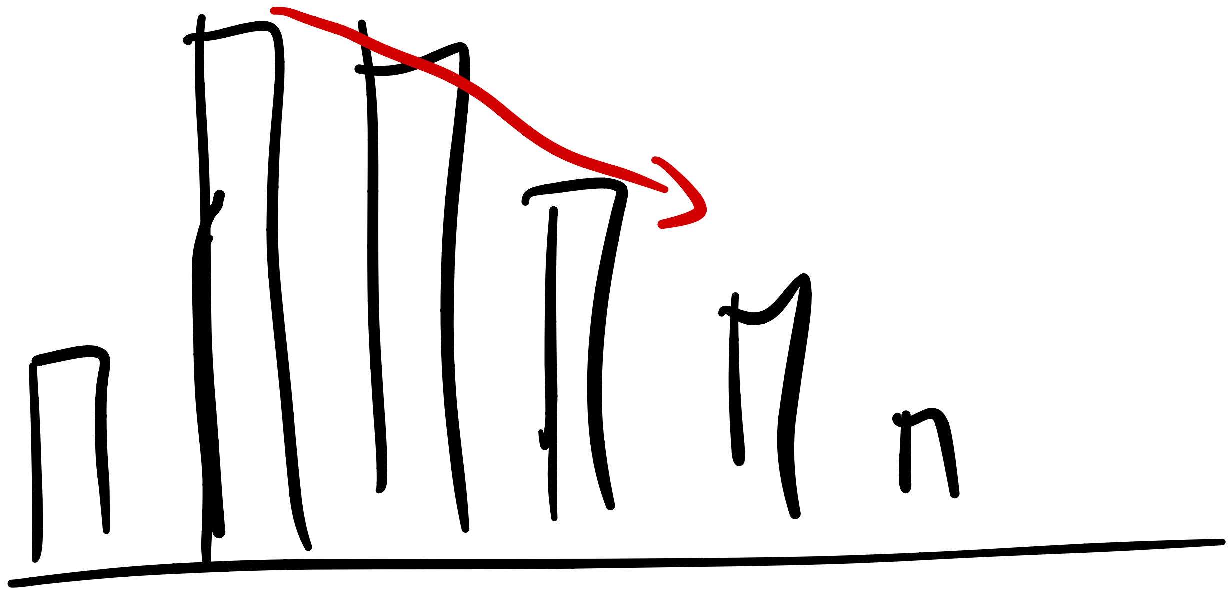
Rice

Mathematical statistics and data analysis

Inferenčna statistika = statistično sklepanje



čas



Poglavje 1 Vzorčenje

Rice 7+8

1.1. Populacijske vrednosti

1.1.1. Povprečje in varianca končne populacije

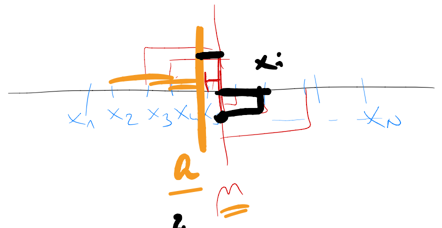
N ... št. vseh elementov populacije

$x_1, \dots, x_N$

$$\bar{x} = \mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$ ← povprečje v populaciji

Na kakšen način je povprečje na sredini:

- vsota odmikov je 0
- vsota kvadriranih odmikov je najmanjša možna



$$\sum_{i=1}^{N}(x_i-\mu) = \sum_{i=1}^{N} x_i - \sum_{i=1}^{N}\mu = N\mu - N\mu = 0$$

$$\sum_{i=1}^{N}(x_i-\mu)^2$$ ← to je najmanjše možno.

$(a+b)^2 = a^2+b^2+2ab$

$$\sum_{i=1}^{N}(x_i-a)^2 = \sum_{i=1}^{N}\big((x_i-\mu)+(\mu-a)\big)^2$$
$$= \sum_{i=1}^{N}\left[(x_i-\mu)^2 + (\mu-a)^2 + 2(x_i-\mu)(\mu-a)\right]$$
$$= \sum_{i=1}^{N}(x_i-\mu)^2 + \sum_{i=1}^{N}(\mu-a)^2 + 2\sum_{i=1}^{N}(x_i-\mu)(\mu-a)$$
$$= \sum_{i=1}^{N}(x_i-\mu)^2 + N(\mu-a)^2 + 2(\mu-a)\underbrace{\sum_{i=1}^{N}(x_i-\mu)}_{=0}$$
$$= \sum_{i=1}^{N}(x_i-\mu)^2 + \boxed{\geq 0} + 0$$

Ali obstaja a, da je enako
$\mu - a = 0$
$\Leftrightarrow a = \mu$

NE

$$\sum_{i=1}^{N}(x_i-a)^2 \geq \sum_{i=1}^{N}(x_i-\mu)^2 \quad \text{za } \forall a$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$$ Povprečje kvadriranih odmikov

Pravila za računanje s pričakovano vrednostjo in varianco

Ponovi.: vse možne vrednosti x in njihove verjetnosti

$$E(X) = \sum_x x_i \cdot p(x_i)$$

diskretna porazd. $= \sum_x x_i \, P(X = x_i)$

zvezna porazd. $= \int_{-\infty}^{\infty} x f(x)\,dx$

$$E(aX) = a\,E(X)$$
$$E(X+Y) = E(X) + E(Y)$$
$$var(X) = E\left([X-E(X)]^2\right) = \sum [X_i - E(X)]^2 \cdot p(x_i)$$
$$var(aX) = a^2\,var(X) \qquad sd(aX) = a\,sd(X)$$
$$var(X+Y) = var(X) + var(Y) + 2cov(X,Y) \qquad \rho = \frac{cov(X,Y)}{\sqrt{var(X)\,var(Y)}}$$

$cov(X,Y) = E\left[(X-E(X))(Y-E(Y))\right]$

$cov(X,X) = E\left[(X-E(X))(X-E(X))\right] = E\left([X-E(X)]^2\right) = var(X)$

$cov(X,Y) = cov(Y,X)$

$cov(X_1+X_2, Y) = cov(X_1,Y) + cov(X_2,Y)$

$var(X+Y) = cov(X+Y, X+Y) = cov(X, X+Y) + cov(Y, X+Y)$
$= cov(X+Y, X) + cov(X+Y, Y) = cov(X,X) + cov(Y,X) + cov(X,Y) + cov(Y,Y)$
$= var(X) + var(Y) + 2cov(X,Y)$

$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$ Pokažimo, da velja $E(X) = \mu$

X slučajna spr. iz naše populacije

$P(X = x_i) = \frac{1}{N}$ $P(X=x) = \frac{1}{N}$

$E(X) = \sum x_i P(X = x_i) = \sum_{i=1}^{N} x_i \cdot \frac{1}{N} = \mu$

$var(X) = E[(X - E(X))^2] = \sum_{i=1}^{N}(x_i - \mu)^2 P(X = x_i) = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2 = \sigma^2$

$\sum_{i=1}^{N}(x_i - a)^2 = \sum_{i=1}^{N}(x_i-\mu)^2 + N(\mu - a)^2$, $a = 0$

$\sum x_i^2 = \sum (x_i - \mu)^2 + N\mu^2$

$\sum_{i=1}^{N} x_i^2 \frac{1}{N} = \sum x_i^2 P(X = x_i) = E(X^2)$

$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \mu^2$

$var(X) = E(X^2) - E(X)^2$

Primer: N enot, x_i ima samo 2 različni vrednosti 0, 1

Delež enic označimo z π: $\pi = \frac{\sum I(x_i = 1)}{N}$

$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{N}\left[\sum_{i; x_i = 0} x_i + \sum_{i; x_i = 1} x_i\right]$

$= \frac{1}{N}\sum_{i; x_i=1} x_i = \frac{1}{N}\sum_i I(x_i = 1) = \pi$

$0^2 = 0$, $1^2 = 1$, $x_i^2 = x_i$

$\sigma^2 = \frac{1}{N}\sum x_i^2 - \mu^2 = \frac{1}{N}\sum x_i - \mu^2$

$= \mu - \mu^2 = \pi - \pi^2 = \pi(1-\pi)$

1.1.2. Neskončna populacija - zvezna spr.

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

$\Rightarrow \int_{-\infty}^{\infty} x f(x)\,dx$

$var(X) = \sigma^2 = E([X - \mu]^2)$

$= \int_{-\infty}^{\infty}(x - \mu)^2 f(x)\,dx$

$= \int_{-\infty}^{\infty}(x^2 - 2x\mu + \mu^2) f(x)\,dx$

$= \int_{-\infty}^{\infty} x^2 f(x)\,dx - \int 2x\mu f(x)\,dx + \int_{-\infty}^{\infty}\mu^2 f(x)\,dx$

$= E(X^2) - 2\mu\int_{-\infty}^{\infty} x f(x)\,dx + \mu^2\int_{-\infty}^{\infty} f(x)\,dx$

$= E(X^2) - 2\mu E(X) + \mu^2 \cdot 1$

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

$-2E(X)^2 + E(X)^2$

1.2. Vzorec

X slučajna spr. $f(x)$ $P(X)$

$n > 1$ $X_1, \dots, X_n$

naključni vzorec: $X_1, \dots, X_n$ parno neodvisne in enako porazdeljene i.i.d. (independent, identically distributed)

neskončna populacija X_1, X_2

končna populacija N. Vzorčenje brez vračanja → odvisnost; z vračanjem

1.2.1. Vzorčenje iz neskončne populacije

Imamo vzorec $X_1, \dots, X_n$ i.i.d., $E(X_i) = \mu$, $var(X_i) = \sigma^2$

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

Ali $\bar{X}$ ocenjuje μ?

- $\bar{X} \neq \mu$ (X zvezna spr. $P(\bar{X} = \mu) = 0$)
- $\bar{X}$ „skače" okoli μ: - $E(\bar{X}) \overset{?}{=} \mu$ - $var(\bar{X}) = ?$ - porazdelitev $\bar{X}$

Definicija:

Cenilka: Predpis - kako s slč. spr. ocenimo količino $f(X_1, \dots, X_n)$ primer: $\bar{X} = \frac{1}{n}\sum X_i$

Ocena: Vrednost, ki jo izračunamo na vzorcu, naj bi ocenjevala neko populacijsko količino $f(x_1, \dots, x_n)$ primer $\bar{x} = \frac{1}{n}\sum x_i$

Nepristranskost vzorčnega povprečja

$X_1, \dots, X_n$ i.i.d. $E(X_i) = \mu$ $var(X_i) = \sigma^2$, $cov(X_i, X_j) = 0$ ($i \neq j$)

$\bar{X} = \frac{1}{n}\sum X_i$ nepristranska cenilka za μ? $E(\bar{X}) = \mu$

$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E(X_1 + X_2 + \dots + X_n)$

$= \frac{1}{n}[E(X_1) + E(X_2) + \dots + E(X_n)]$

$= \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}\cdot n\mu = \mu$

Varianca vzorčnega povprečja

$cov(X_1 + \dots + X_n, \sum X_i)$

$var(\bar{X}) = var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}var\left(\sum X_i\right) = \frac{1}{n^2}cov\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i\right)$

$= \frac{1}{n^2}\left[cov(X_1, \sum_{i=1}^{n} X_i) + cov(X_2, \sum_{i=1}^{n} X_i) + \dots + cov(X_n, \sum X_i)\right]$

$= \frac{1}{n^2}\left[[cov(X_1, X_1) + cov(X_1, X_2) + \dots + cov(X_1, X_n)] + \sum_{i=1}^{n} cov(X_2, X_i) + \dots + \sum_{i=1}^{n} cov(X_n, X_i)\right]$

$\sum_{i=1}^{n} cov(X_1, X_i)$

i.i.d. neodv. $\Rightarrow$ $= \frac{1}{n^2}[cov(X_1, X_1) + cov(X_2, X_2) + \dots + cov(X_n, X_n)]$

$= \frac{1}{n^2}[var(X_1) + \dots + var(X_n)] = \frac{1}{n^2}\sum var(X_i)$

i.i.d. $var(X_i) = \sigma^2$ $\Rightarrow$ $= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$

$\sqrt{var(\bar{X})} = SE$

Preden

$\sigma^2 =$

$\tilde{\sigma}^2 = \frac{1}{n}$

Recimo

$E(\tilde{\sigma}^2)$

$E(X) = \sum_x x P(X = x)$

Primer:

$X \sim Ber(0{,}5)$ $n = 3$ $\mu = 0{,}5$

000
1 00
1 1 0
1 1 1

porazdelitev vzorčnega povp.

$\bar{X}$	P	$\bar{X} - \mu$	$(\bar{X} - \mu)^2$	$\frac{1}{3}\sum(x_i - \bar{x})^2$
0	$\frac{1}{8}$	$-\frac{1}{2}$	$\frac{1}{4}$	0
$\frac{1}{3}$	$\frac{3}{8}$	$-\frac{1}{6}$	$\frac{1}{36}$	$\frac{2}{9}$
$\frac{2}{3}$	$\frac{3}{8}$	$\frac{1}{6}$	$\frac{1}{36}$	$\frac{2}{9}$
1	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{4}$	0

možne vrednosti

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

$\frac{1}{3} - \frac{1}{2} =$

$\sigma^2 = \frac{1}{4} = \frac{2}{8}$

$E(\bar{X}) = 0 \cdot \frac{1}{8} + \frac{1}{3}\cdot\frac{3}{8} + \frac{2}{3}\cdot\frac{3}{8} + 1 \cdot \frac{1}{8}$

$= \frac{1}{8} + \frac{2}{8} + \frac{1}{8} = \frac{4}{8} = \frac{1}{2} = 0{,}5 = \mu$

Po formuli $var(\bar{X}) = \frac{\sigma^2}{n} = \frac{var(X)}{n} = \frac{\pi(1-\pi)}{n} = \frac{0{,}5\cdot(1-0{,}5)}{n} = \frac{0{,}25}{3} = \frac{1}{12}$

$var(\bar{X}) = \sum_{\bar{x}}(\bar{x} - \mu)^2 \cdot P(\bar{X} = \bar{x}) = \frac{1}{4}\cdot\frac{1}{8} + \frac{1}{36}\cdot\frac{3}{8} + \frac{1}{36}\cdot\frac{3}{8} + \frac{1}{4}\cdot\frac{1}{8} =$

$= 2\cdot\frac{1}{4}\cdot\frac{1}{8} + 2\cdot\frac{1}{36}\cdot\frac{3}{8} = \frac{1}{16} + \frac{1}{48} = \frac{3+1}{48} = \frac{4}{48} = \frac{1}{12}$

Varianca:

X_1, X_2, X_3

gremo na podatke. Kaj pa moramo dobiti?

$E((X - E(X))^2) = \pi(1 - \pi) = \frac{1}{4}$

$\sum_{i=1}^{3}(X_i - \bar{X})^2$

$X_1 = 0$, $X_1 = 1$, $X_2 = 0$ $\bar{X} = \frac{1}{3}$

$\frac{1}{3}\left[\left(0 - \frac{1}{3}\right)^2 + \left(1 - \frac{1}{3}\right)^2 + \left(0 - \frac{1}{3}\right)^2\right] = \frac{1}{3}\left[\frac{1}{9} + \frac{4}{9} + \frac{1}{9}\right]$

$= \frac{2}{9}$ na tem vzorcu podcenimo σ^2

$= 0\cdot\frac{1}{8} + \frac{2}{9}\cdot\frac{3}{8} + \frac{2}{9}\cdot\frac{3}{8} + 0\cdot\frac{1}{8} = \frac{1}{6}$

0,0,0 1,0,0 1,1,0 1,1,1

Primer: 2 meseca gonega zdravila $var(D) = 25^2$

$\sigma = 25$, $m = 100$. Recimo, da v populaciji ni razlik $E(D) = 0$

Ali lahko na vzorcu dobimo $\bar{D} = 10$

$SE = \frac{\sigma}{\sqrt{m}} = \frac{25}{\sqrt{100}} = \frac{25}{10} = 2{,}5$

Neenačba Čebiševa: $P(|X-\mu| \geq b) \leq \frac{\sigma^2}{b^2}$

$P(|\bar{X}-\mu| \geq 4SE) \leq \frac{var(\bar{X})}{16 SE^2} = \frac{SE^2}{16 SE^2} = \frac{1}{16}$

Imamo vzorec, $m = 100$ $\quad \bar{X} = \hat{\mu} = \frac{1}{m}\sum_{i=1} X_i$

$var(\bar{X}) = \frac{\sigma^2}{100}$

$\tilde{\mu} = \frac{1}{5}\sum_{i=1}^{5} X_i$

- Ali je $\tilde{\mu}$ cenilka ✓
- Ali je $\tilde{\mu}$ nepristranska?
  $E(\tilde{\mu}) = E\left(\frac{1}{5}\sum_{i=1}^{5} X_i\right) = \frac{1}{5}E\left(\sum_{i=1}^{5} X_i\right) = \frac{1}{5}\sum_{i=1}^{5} E(X_i) = \frac{1}{5}\sum_{i=1}^{5}\mu = \frac{1}{\cancel{5}}\cancel{5}\mu = \mu$
- $var(\tilde{\mu}) = var\left(\frac{1}{5}\sum_{i=1}^{5} X_i\right) = \frac{\sigma^2}{5}$

Ocenjevanje variance σ^2

Primer

$\tilde{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m}(X_i - \bar{X})^2 \Rightarrow$ ta cenilka podcenjuje σ^2!

$X_1, \dots, X_m$ i.i.d.

$E(X_i) = \mu$, $var(X_i) = \sigma^2$

Kaj želimo: $\hat{\sigma}^2 = c\sum_{i=1}^{m}(X_i - \bar{X})^2$

Želimo najti tak c, da bo $E(\hat{\sigma}^2) = \sigma^2$

$var(X - Y)$

$= var(X + (-1)\cdot Y) = \dots$

$= var(X) + var(Y) - 2cov$.

$E\left(\sum_{i=1}^{m}(X_i - \bar{X})^2\right) = c\sum_{i=1}^{m} E[(X_i - \bar{X})^2] = *$

$\dots X_i - \bar{X})]^2 + var(X_i - \bar{X})$

$\dots) - E(\bar{X})]^2 + var(X_i) + var(\bar{X}) - 2cov(X_i, \bar{X})$

$\dots - \mu]^2 + \sigma^2 + \frac{\sigma^2}{m} - 2cov\left(X_i, \frac{1}{m}\sum_{i=1}^{m} X_i\right)$

od prej

$-\frac{2}{m}cov(X_i, X_1 + X_2 + \dots + X_m)$

$\dots + \sigma^2 + \frac{\sigma^2}{m} - \frac{2}{m}\left(cov(X_i, X_1) + cov(X_i, X_2) + \dots + cov(X_i, X_m)\right)$

$\dots^2 + \frac{\sigma^2}{m} - \frac{2}{m}cov(X_i, X_i) \rightarrow var(X_i)$

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

$\dots) = c \cdot \cancel{m}\,\sigma^2\left(\frac{m-1}{\cancel{m}}\right) = c\sigma^2(m-1) \Rightarrow$

Zakaj $(m-1)$

Kaj pa cenilke:

$\breve{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m}(X_i - \mu)^2$

$var(X + a) = var(X) + var(a) + 2cov(X, a)$

$E(\breve{\sigma}^2) = E\left(\frac{1}{m}\sum_{i=1}^{m}(X_i - \mu)^2\right) = \frac{1}{m}\sum_{i=1}^{m} E[(X_i - \mu)^2]$

$= \frac{1}{m}\sum_{i=1}^{m}\left[[E(X_i - \mu)]^2 + var(X_i - \mu)\right]$

$= \frac{1}{m}\sum_{i=1}^{m}\left[[E(X_i) - E(\mu)]^2 + var(X_i)\right]$

$= \frac{1}{m}\sum_{i=1}^{m}\left[(\mu - \mu)^2 + \sigma^2\right]$

$= \frac{1}{m}\sum_{i=1}^{m}[\sigma^2] = \frac{1}{m}\cdot m \cdot \sigma^2 = \sigma^2$

$$\hat{\sigma}^2 = \frac{1}{m-1}\sum(X_i - \bar{X})^2$$

$$\Rightarrow E(\hat{\sigma}^2) = \sigma^2$$

$$c = \frac{1}{m-1}$$

Ocenjevanje variance, povzetek

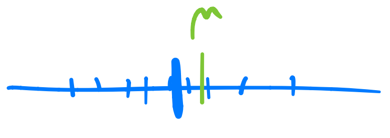
Ocenjujemo $\sigma^2 = E\left((X_i - E(X))^2\right)$ nepr.

Več cenilk:

$\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar X)^2$ ✗

$\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar X)^2$ ✓

$\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ ✓



$\sum(X_i-\bar X)^2 \le \sum(X_i-a)^2$ (μ)

Za konec

$$\hat\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar X)^2}$$

ali je ta cenilka nepr. za σ?

$E(\hat\sigma) \neq \sigma$

$E(Y) = a$

$E(Y^2) \neq a^2$

$E(Y^2) = var\,Y + E(Y)^2$

$$var(\hat\sigma) = E(\hat\sigma^2) - E(\hat\sigma)^2 = \sigma^2 - \sigma^2 = 0$$

$\hat\sigma$ ni nepristranska cenilka !

D.N. Preizkusi na primeru Bernoulli.

Doslednost

Do sedaj 2 lastnosti: - nepristr.
- varianca cenilke

Naj bo $\hat\theta_n$ cenilka na vzorcu velikosti n. Pravimo, da je cenilka dosledna, če za vsak $\varepsilon > 0$ velja

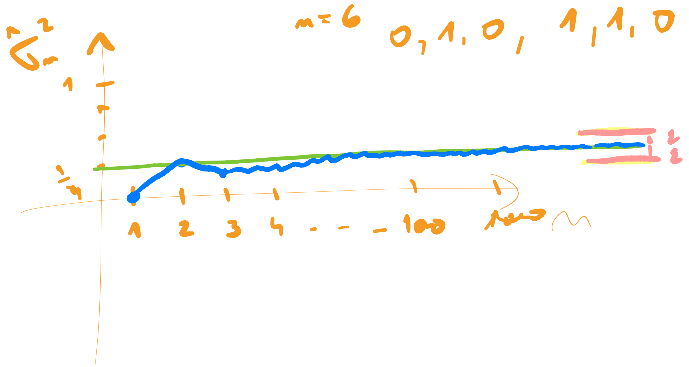
$$\lim_{n\to\infty} P(|\hat\theta_n - \theta| > \varepsilon) = 0$$

$\tilde\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar X)^2$

n=1: 0

n=3: 0, 1, 0 $\hat\sigma_3 = \frac{2}{9}$

n=6: 0, 1, 0, 1, 1, 0 $\hat\sigma_6 = \dots$



Pomožni izrek:

Naj bo $\hat\theta_n$ zaporedje cenilk, za katerega velja (ko gre $n\to\infty$)

• Cenilke so asimptotsko nepristranske

$$E(\hat\theta_n) \xrightarrow[n\to\infty]{} \theta \qquad \lim_{n\to\infty} E(\hat\theta_n) = \theta$$

Primer $\bar X$: $E(\bar X) = \mu$

• Varianca cenilk gre proti 0

$$\lim_{n\to\infty} var(\hat\theta_n) = 0$$

Primer $\bar X$: $var(\bar X) = \frac{\sigma^2}{n} \xrightarrow[n\to\infty]{} 0$

Potem je zaporedje $\hat\theta_n$ dosledno za θ.

Primer:
Vzorčno povprečje je dosledna cenilka.

Šibki zakon velikih števil:
Naj bodo X_i i.i.d., $E(X_i) = \mu$, $var(X_i) = \sigma^2 < \infty$.
Označimo z $\bar X_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Potem za vsak $\varepsilon > 0$ velja

$$\lim_{n\to\infty} P(|\bar X_n - \mu| < \varepsilon) = 1$$

Denimo, da zaporedje $X_1, X_2, \dots$ konvergira v verjetnost proti slučajni spremenljivki X in da je h zvezna funkcija. Potem zaporedje $h(X_1), h(X_2)\dots$ konvergira v verjetnost proti $h(X)$.

Posledica: če je $\hat\theta_n$ ($\bar X$) dosledna cenilka za θ (μ) in h zvezna funkcija, ($h(x) = x^2$) je tudi $h(\hat\theta_n)$ ($(\bar X)^2$) dosledna cenilka za $h(\theta)$. ($(\mu)^2$)

Primer: Cenilka $\tilde\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar X)^2$ Doslednost za σ^2?

$\tilde\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar X^2$ $\bar X$ dosledna za μ ✓ zakon velikih števil

$Y_i = X_i^2$ $X_1, X_2, \dots$ i.i.d.   1, 6, 3, 2, ...

$\frac{1}{n}\sum_{i=1}^{n} Y_i$ dosl. za $E(Y_i)$ $X_1^2, X_2^2, \dots$ i.i.d.   1, 36, 9, 4, ...

$X_1, \dots, X_n, \dots$ iid $\quad \frac{1}{n}\sum X_i$ dosl. za $E(X_i) = \mu$ $\qquad var(Y) = E(Y^2) - E(Y)^2$

$Y_1, \dots, Y_n, \dots$ iid $\quad \frac{1}{n}\sum Y_i$ dosl. za $E(Y_i)$

$X_1^2, \dots, X_n^2, \dots$ iid $\quad \frac{1}{n}\sum X_i^2$ dosl. za $E(X_i^2) = \mu^2 + var(X_i)$

$= \mu^2 + \sigma^2$

$$\tilde{\sigma}^2 = \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i^2}_{\text{dosledna za } \mu^2+\sigma^2} - \underbrace{\bar{X}^2}_{\text{dosledna za } \mu^2}$$

$$\underbrace{\mu^2 + \sigma^2 - \mu^2}_{\sigma^2}$$

$\bar{X}$ dosledna za μ

$\downarrow$ $h(x) = x^2$ zvezna funkcija

$\bar{X}^2$ dosledna za μ^2

Povzetek:

Asimptotske lastnosti ko gre $n \to \infty$ (za velike vzorce)

Lažje doslednost !

| Lastnosti nedvne od n za vsak n

nepristranskost

varianca

porazdelitev

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

$var(\bar{X}) = \frac{\sigma^2}{n}$

$E(\bar{X}) = \mu$

$$\hat{\sigma}^2 = \frac{n}{n-1}\tilde{\sigma}^2 \to \frac{n}{n-1}\sigma^2 \to \sigma^2$$

Kaj pa $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$?

$$\hat{\sigma}^2 = \frac{n}{n-1}\cdot\tilde{\sigma}^2 \qquad \hat{\sigma}^2 = h(\tilde{\sigma}^2)$$

$\Rightarrow h$ je zvezna $\Rightarrow \hat{\sigma}^2$ dosledna za $h(\sigma^2)$

$\frac{n}{n-1} \to 1$

$h(\sigma^2) \to \sigma^2$

$$\hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2} \quad \text{ali je dosledna za } \sigma \text{ ?} \checkmark$$

$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ ✓

dosledna za σ

$h(x) = \sqrt{x}$

$\frac{1}{n-1}\sum(X_i - \bar{X})^2$ dosledna za σ^2

$\sqrt{\frac{1}{n-1}\sum(X_i-\bar{X})^2}$ dosledna za σ

Primer: Ali so vse nepristranske cenilke tudi dosledne?

$X_1, X_2, \dots$

$\tilde{X} = \frac{1}{5}\sum_{i=1}^{5} X_i$ je nepristranska za μ $\quad E(\tilde{X}) = \mu$ za vsak n

$\Rightarrow$ je tudi asimptotsko nepristranska

$var(\tilde{X}) = \frac{\sigma^2}{5} \xrightarrow{n\to\infty} \frac{\sigma^2}{5} \not\to 0$

Ni dosledna !

1.2.2. Vzorčenje iz končne populacije

Do vedaj: $X_1, X_2, \ldots$ i.i.d. independent, identically distributed

Zdaj: vzorčili bomo brez vračanja → težave z neodvisnostjo

Primer: $N=5$ $x_1=1$ $x_2=3$, 5, 7, 9

X_2 in X_1 odvisna: če X_1 nekaj pove o X_2

Enake porazdelitve?

X_1	1	3	5	7	9	← vrednost
	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	← verjetnost

$P(X_2=x_i) = P(X_1=x_i)$

$P(X_2=x_i \mid X_1=x_j)$

X_2

Kako si predstavljamo vzorčenje iz končne populacije?

X_1	X_2	X_3		
5	3	7	1	9
1	9	3	7	5
5	9	1	7	3

Formal. dokaz: ← zakon o popolni verjetnosti

$$P(X_2=x_j) = \sum_{i=1}^{N} P(X_2=x_j \mid X_1=x_i)\,P(X_1=x_i)$$

$$= \sum_{\substack{i=1\\ i\neq j}}^{N} \underbrace{P(X_2=x_j \mid X_1=x_i)}_{\frac{1}{N-1}}\,\underbrace{P(X_1=x_i)}_{\frac{1}{N}} + \underbrace{P(X_2=x_j \mid X_1=x_j)}_{0}\,P(X_1=x_j)$$

$$= \sum_{\substack{i=1\\ i\neq j}}^{N} \left(\frac{1}{N-1}\right)\cdot\frac{1}{N} = \underbrace{\cancel{(N-1)}}_{\text{št. členov v vsoti}}\cdot\frac{1}{\cancel{N-1}}\cdot\frac{1}{N} = \frac{1}{N}$$

⇒ Pri vzorčenju z vračanjem:

$X_1, \ldots, X_n$ so enako porazdeljene, a odvisne sluč. spr.

Za kakšno odvisnost gre? cov ?

$\mathrm{cov}(X_1, X_2)$ $\mathrm{cov}(X_1, X_3)$

$\mathrm{cov}(X_2, X_3) = \mathrm{cov}(X_1, X_3)$

$\mathrm{cov}(X_i, X_j)$ enaka za vse pare $i \neq j$

$\mathrm{cov}(X_i, X_j) = ?$

$1, 3, 5, 7, 9$

$\mathrm{cov}(X_3, \sum_{j=1} X_j) =$

$$\mathrm{cov}\left(X_i, \sum_{j=1}^{N} X_j\right) = 0$$

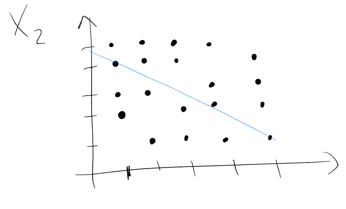
$$\mathrm{cov}(X_i, X_1) + \ldots + \mathrm{cov}(X_i, X_i) + \ldots + \mathrm{cov}(X_i, X_N) =$$

$$= (N-1)\,\underset{i\neq j}{\mathrm{cov}(X_i, X_j)} + \mathrm{var}(X_i)$$

$$\underset{i\neq j}{\mathrm{cov}(X_i, X_j)} = -\frac{\mathrm{var}(X_i)}{N-1}$$

$$\underset{i\neq j}{\mathrm{cor}(X_i, X_j)} = \frac{\mathrm{cov}(X_i, X_j)}{\sqrt{\mathrm{var}X_i}\,\sqrt{\mathrm{var}X_j}} \overset{\text{enako porazd.}}{=} -\frac{\cancel{\mathrm{var}(X_i)}}{(N-1)\cdot\cancel{\mathrm{var}(X_i)}} = -\frac{1}{N-1}$$

interpretacija:
- odvisna izključno od velikosti populacije
- predznak: negativen
- povezanost je šibka



$X_1, \dots, X_n$ enako porazd.: $E(X_i) = \mu$, $var(X_i) = \sigma^2$, odvisne: $cov(X_i, X_j) = -\frac{\sigma^2}{N-1}$

- Nepristranskost $\overline{X}$?

$E(\overline{X}) = \mu$ za dokaz ne potrebujem neodvisnosti

- $var(\overline{X}) = \frac{1}{n^2} cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right)$

$$= \frac{1}{n^2} \sum_{i=1}^{n} cov\left(X_i, \sum_{j=1}^{n} X_j\right) = \frac{1}{n^2} \sum_{i=1}^{n} \left(cov(X_i, X_1) + cov(X_i, X_2) + \dots + cov(X_i, X_i) + \dots + cov(X_i, X_n)\right)$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \left(cov(X_i, X_i) + (n-1) \cdot \underset{i \neq j}{cov} (X_i, X_j)\right)$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \left(\sigma^2 - \frac{(n-1)\sigma^2}{N-1} \right) = \frac{1}{n^{\cancel{2}}} \cancel{n} \cdot \sigma^2 \left(\frac{N-1}{N-1} - \frac{n-1}{N-1} \right)$$

$$= \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

Interpretacija: varianca še vedno odvisna od σ^2 in n

Pomnožena s faktorjem: $\frac{N-n}{N-1} \leq 1$

Popravek je pri veliki populaciji N majhen

$N = 11$ $n = 6$

$$var(\overline{X}) = \frac{\sigma^2}{n} \frac{N-n}{N-1} = \frac{\sigma^2}{6} \frac{(11-6)}{(11-1)} = \frac{\sigma^2 \cdot 5}{6 \cdot 10} = \frac{\sigma^2}{12}$$

$N = 100$, recimo σ^2 enak

$n = 10$ $var(\overline{X}) = \frac{\sigma^2}{10} \frac{(100-10)}{(100-1)} = \frac{\sigma^2 \cdot \cancel{90}\,9}{\cancel{10} \cdot 99} = \frac{\sigma^2}{11}$

$n = 11 \Rightarrow var(\overline{X}) = \frac{\sigma^2}{12{,}2}$

$N = 1000$
$n = 11$ $\Rightarrow var(\overline{X}) = \frac{\sigma^2}{11} \cdot \frac{1000 - 11}{1000 - 1} = \frac{\sigma^2}{11} \frac{989}{999}$

Ocenjevanje populacijske variance σ^2. $X_1, \dots, X_n$ $E(X_i) = \mu$, $var(X_i) = \sigma^2$, odvisni

$$E(Y^2) = E(Y)^2 + var(Y)$$

$$E\left(c \sum_{i=1}^{n} (X_i - \overline{X})^2\right) = c \sum_{i=1}^{n} E(X_i - \overline{X})^2$$

$$= c \sum_{i=1}^{n} \left[\left(\underbrace{E(X_i - \overline{X})}_{(E(X_i) - E(\overline{X}))^2 \;=\; \mu - \mu \;=\; 0}\right)^2 + var(X_i - \overline{X}) \right]$$

$$= c \sum_{i=1}^{n} \left[var(X_i) + var(\overline{X}) - 2 cov(X_i, \overline{X}) \right]$$

$$= c \sum_{i=1}^{n} \left[\sigma^2 + \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} - 2 cov\left(X_i, \frac{1}{n} \sum_{j=1}^{n} X_j\right) \right] = *$$

$$\frac{1}{n}\left[cov(X_i, X_1) + \dots + cov(X_i, X_i) + \dots + cov(X_i, X_n) \right]$$

$$= \frac{1}{n}\left[var(X_i) + (n-1) \underset{i \neq j}{cov}(X_i, X_j) \right]$$

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

$$* = c \sum_{i=1}^{n} \left[\sigma^2 + \frac{\sigma^2}{n} \frac{N-n}{N-1} - 2 \frac{\sigma^2}{n} \frac{N-n}{N-1} \right]$$

$$= c \sum_{i=1}^{n} \left[\sigma^2 - \frac{\sigma^2}{n} \frac{N-n}{N-1} \right]$$

$$= c \cdot n \sigma^2 \frac{(N-1)n - N + n}{(N-1)n}$$

$$= c n \sigma^2 \frac{Nn - N}{n(N-1)} = \frac{c \cancel{n} N (n-1) \sigma^2}{\cancel{n}(N-1)} = c \cdot \frac{N}{N-1} \cdot (n-1) \cdot \sigma^2 = \sigma^2$$

$$\Downarrow$$

$$c = \frac{1}{(n-1)} \cdot \frac{N-1}{N}$$

Neodv: $var(X+Y) = var(X) + var(Y)$

odv. $var(X+Y) = var(X) + var(Y) + 2cov(X,Y)$

neodv: $var(X+Y) =$

$$= cov(X+Y, X+Y) =$$

$$= cov(X, X+Y) + cov(Y, X+Y) =$$

$$= cov(X,X) + \underbrace{cov(X,Y)}_{=0} + \underbrace{cov(Y,X)}_{=0} + cov(Y,Y)$$

$$= cov(X,X) + cov(Y,Y)$$

$$= var(X) + var(Y)$$

Cenilka $\frac{N-1}{N} \cdot \frac{1}{n-1} \sum (X_i - \overline{X})^2$ je nepristranska cenilka za σ^2.