TMA4240
F16

Punktestimering [9.1-9.3]

MÅL: finne anslag for ukjent parameter i valgt populasjon ut i fra et tilfeldig utvalg.

a) Spiringsgrad

Populasjon: alle frø fra arten, med fokus på spiringsgrad.

Trekker ett frø fra populasjonen, plant og se om frøet spirer.

Ukjent spiringsgrad: p

Utvalg: trekker tilfeldig n frø og X av de spirer. X = antall spirede frø

$$\underbrace{\text{spirede frø}}_{\text{suksess}}$$

$X \sim \text{bin}(n, p)$

b) VO_2max

Populasjon: menn i Nord-Trøndelag - se på deres oksygenopptak. Antar normalfordelt populasjon med ukjent forventningsverdi μ og ukjent standardavvik σ.

Utvalg: $X_1, \dots, X_n$ uavhengige

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

ANSLAG FOR PARAMETERE

a) Anslag for p: $\frac{X}{n}$, skriver $\hat{p} = \frac{X}{n}$

(n, X)

SV

SV

tall

b) Anslag for μ: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, skriver $\hat{\mu} = \bar{X}$

$(X_1, \dots, X_n)$

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

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

Fokus på $\bar{X}$: fordelingen til gjennomsnittet og sentralgrenseteoremet [8.4]

$X_1, X_2, \ldots, X_n$ u.i.f. $f(x)$

uavhengige identisk fordelt

$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$, $E(X_i) = \mu$, $Var(X_i) = \sigma^2$

For generell $f(x)$ kan vi bruke mgf til å finne fordeling til $\bar{X}$, men det kan være en ukjent fordeling.

Når n er stor ($n \geq 30$) bruker vi <u>sentralgrenseteoremet</u>

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

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

$SD(\bar{X}) = \sqrt{Var(\bar{X})} = \frac{\sigma}{\sqrt{n}}$

$$Z = \frac{\bar{X} - E(\bar{X})}{SD(\bar{X})} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{n \to \infty} N(0,1)$$

Så, selv om X ikke er normalfordelt vil $\bar{X}$ bli normalfordelt for store n.
For symmetriske fordelinger skjer dette raskere enn for asymmetriske (skjeve) fordelinger.

Merk: Hvis $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{n \to \infty} N(0,1)$

Betyr det også at

varians

$$\bar{X} \text{ er tilnærmet } N\left(\mu, \frac{\sigma^2}{n}\right)$$

og siden $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$, betyr det også at

lukse

$$\sum_{i=1}^{n} X_i \text{ er tilnærmet } N(n\mu, n\sigma^2)$$

FORMELT:

$$\text{Estimator} - \text{for en parameter } \overbrace{(p, \mu, \sigma^2)}^{\Theta}$$

er en SV

$$\underbrace{(\hat{p}, \hat{\mu}, \hat{\sigma}^2, S^2)}_{\hat{\Theta}}$$

Estimat: verdien av estimatoren når vi har satt inn observerte data.

$(\hat{p}, \hat{\mu}, s^2, \hat{\sigma}^2)$

a) Spiring av frø, $n = 1200$, $x = 987$

Estimator: $\underline{\hat{p} = \frac{X}{n}}$, Estimat: $\hat{p} = \frac{987}{1200} = \underline{0.82}$

b) VO_2 max

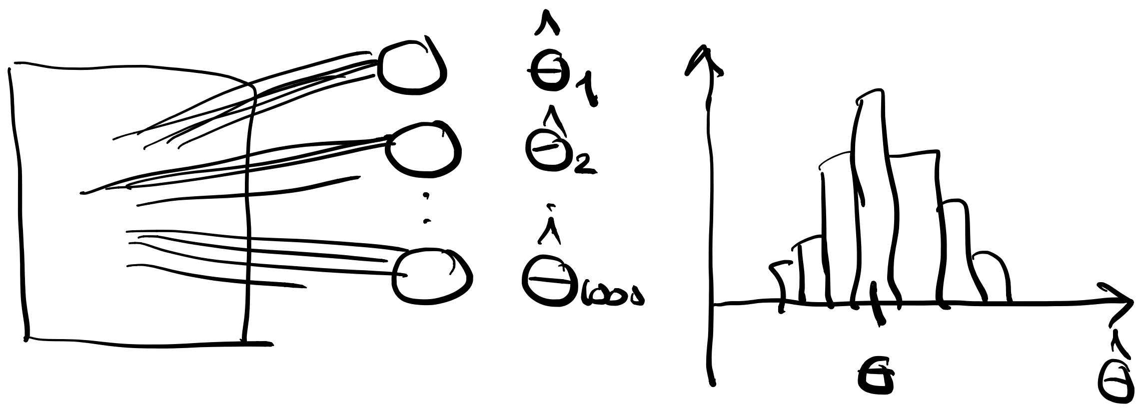
Estimator: $\hat{\mu} = \overline{X}$, estimat $\hat{\mu} = 136$

$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$, estimat $s^2 = 665$

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

Hvilke egenskaper bør en god estimator, $\hat{\theta}$, ha?

1) Hvis vi bruker estimatoren gjentatte ganger (på ulike utvalg fra samme populasjon) vil vi at gjennomsnittet angir rett verdi av parameteren.



$E(\hat{\theta}) = \theta$ ← estimatoren $\hat{\theta}$ er forventningsrett. (unbiased)

$E(\hat{\theta}) > \theta \Rightarrow$ gjennomgående for høye anslag av θ

$E(\hat{\theta}) < \theta \Rightarrow$ ——"—— lave

a) $\hat{p} = \frac{X}{n}$, $X \sim bin(n, p)$

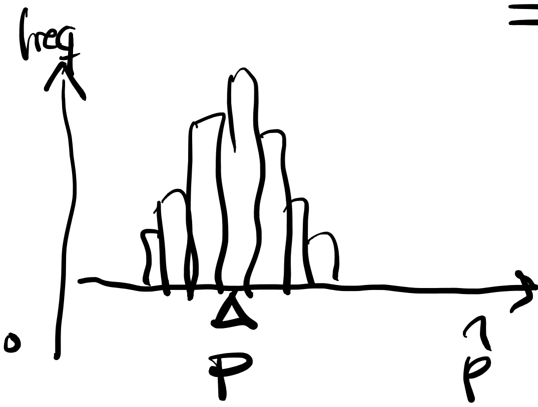
$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n}\overbrace{E(X)}^{n \cdot p} = \frac{1}{n} \cdot n \cdot p = \underline{\underline{p}}$$

utvalg 1: $X \rightarrow \frac{X}{n} = \hat{p}_1$

2 $\hat{p}_2$

$\vdots$

1000 $\hat{p}_{1000}$



b) $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$

$$E(\hat{\mu}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\overbrace{E(X_i)}^{\mu}$$

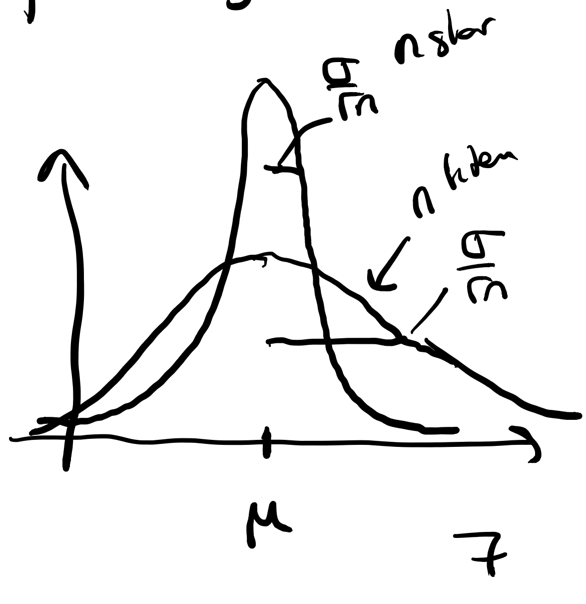
$$= \frac{1}{n} \cdot n \cdot \mu = \underline{\underline{\mu}}$$ forventningsrett

utvalg 1: $X_1, \ldots, X_n$: $\overline{X}_1$

2 $\overline{X}_2$

$\vdots$

1000 $\overline{X}_{1000}$



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

Er S^2 en forventningsrett estimator?

$$E(S^2) = E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right)$$

$$= \cdots = \sigma^2$$ forventningsrett

S^2 er forventningsrett → som er grunnen til at den brukes som estimator for σ^2.

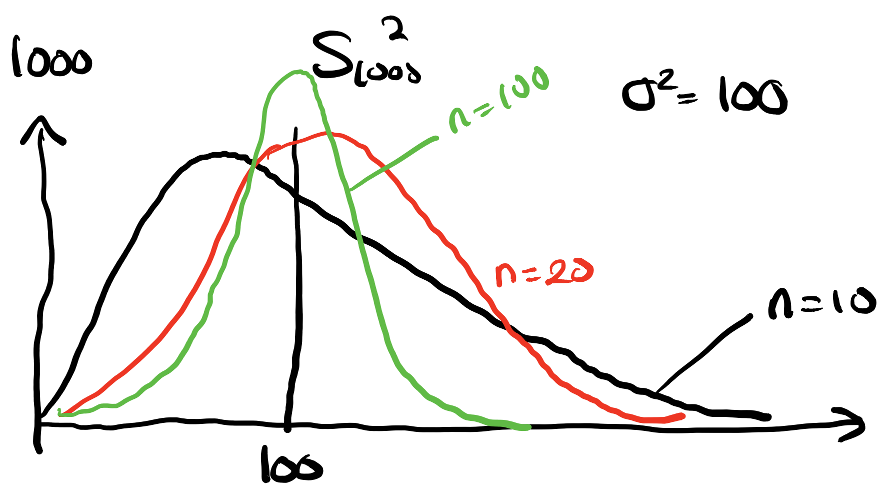
Dette gjelder uansett hvilken fordeling X har.

\# individ i utvalget

Utvalg 1: $X_1, .., X_n$: S_1^2

2 X_n : S_2^2

⋮ ⋮



Hvilken fordeling har S^2? [6.7, 8.5]

La $X_1, X_2, \ldots, X_n$ u.i.f $N(\mu, \sigma^2)$

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

Da er
$$V = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)\cdot S^2}{\sigma^2}$$
$$\sim \chi^2_{\nu = n-1}$$

khikvadratfordeling med $\nu = n-1$

Hvis $V \sim \chi^2_\nu$ så er $E(V) = \nu$
$$Var(V) = 2\nu$$

$$V = \frac{(n-1)S^2}{\sigma^2} \iff S^2 = \frac{\sigma^2 V}{n-1}$$

Fordelingen til S^2:
hvis vi vil finne f_{S^2} kan vi gjøre det ved transformasjonsformelen fra kap 7.

Hva er da $E(S^2)$?

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

$$E(V) = \nu = n-1$$

$$E(S^2) = \frac{\sigma^2}{n-1} \cdot n-1 = \underline{\underline{\sigma^2}}$$

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

$$\hat{\sigma}^2 = \frac{(n-1)S^2}{n}$$

$$E(\hat{\sigma}^2) = \frac{n-1}{n} \cdot \underbrace{E(S^2)}_{\sigma^2} = \underbrace{\frac{n-1}{n}}_{\text{alltid} < 1} \cdot \sigma^2$$

$$E(\hat{\sigma}^2) < \sigma^2$$

$\hat{\sigma}^2$ underestimerer σ^2

$\Longrightarrow$ Derfor bruker vi S^2 som estimator!