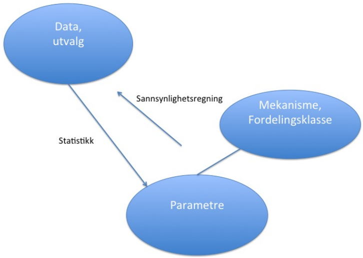
Statistikk

Jo Eidsvik

Matematiske fag, NTNU

Utvalg og Estimering

- Data er et utvalg: $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt med tetthet eller punktsannsynlighet $f(x; \theta)$.
- Vi vil estimere modellparameter $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$.
- Vi antar typisk at fordelingstype (Normal, eksponensial, Poisson, eller lignende) er kjent fra erfaring omkring mekanismen.



Gjennomsnittet i et utvalg

$X_1, \ldots, X_n$ er uavhengige normalfordelte med $E(X_i) = \mu$, $Var(X_i) = \sigma^2$.

$$\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$$

1. $\bar{X} = \hat{\mu}(X_1, \ldots, X_n)$ er en estimator for μ.
2. $Z = \sqrt{n}\frac{\bar{X}-\mu}{\sigma}$, Z er normalfordelt med $E(Z) = 0$, $Var(Z) = 1$. Dette gjør at man kan se på de statistiske egenskapene til estimatoren $\hat{\mu} = \hat{\mu}(X_1, \ldots, X_n)$.
3. $\bar{x} = \hat{\mu}(x_1, \ldots, x_n)$ er et estimat for μ.

Varians i et utvalg

$X_1, \ldots, X_n$ er uavhengige normalfordelte med $E(X_i) = \mu$, $Var(X_i) = \sigma^2$.

$$\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

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

1. S^2 estimerer σ^2.
2. $Y = \frac{(n-1)S^2}{\sigma^2}$. Y er kji-kvadratfordelt med $n-1$ frihetsgrader.

Eksempel - akslinger

$X_1, \ldots, X_{35}$, $S^2 = 0.0029$. Er dette betydelig mindre enn tidligere usikkerhet $\sigma^2 = 0.005$?
Dersom $\sigma^2 = 0.005$ fortsatt er gyldig teoretisk varians: $\frac{34S^2}{0.005}$ er kji-kvadratfordelt med 34 frihetsgrader.
Realisasjon er $\frac{34 \cdot 0.0029}{0.005} = 20.3$. $P(Y < 20.3) = 0.022$.
Det ser ut til at den gamle verdien gir lavere verdi (trolig at vi burde dele på en mindre σ^2, men det ikke ekstremt lav verdi i forhold til kji-kvadrat fordeling.