



# Forelesing 16

## Estimatorar

Torstein Fjeldstad

Institutt for matematiske fag, NTNU

28.02.2018

# I dag



- Sentralgrenseteoremet
- Estimatorar
- Kahoot

# Repetisjon I



## Populasjon

Ein populasjon består av alle moglege observasjonar me kan gjere.

## Utval

Eit utval er ei delmengde av ein populasjon.

## Tilfeldig utval

Tilfeldig utval (eng: random sample): einingane som trekkes frå populasjonen velges tilfeldig og uavhengig av kvarandre.

# Repetisjon II



## Observator

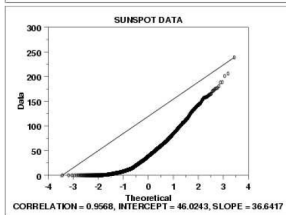
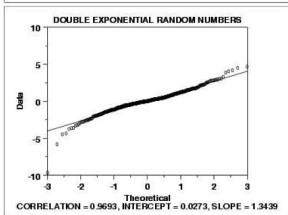
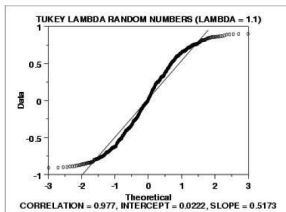
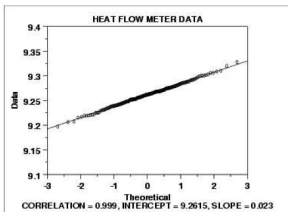
Ein **observator** (eng: statistic) er ein funksjon av stokastiske variablar i eit tilfeldig utval.

## Utvalsfordeling

Fordelinga til ein observator kallast utvalsfordelinga (eng: sampling distribution).

# Repetisjon III

**Normalplott:** Plott  $x_{(1)}, \dots, x_{(n)}$  mot  $\Phi^{-1}(\frac{i}{n})$  og avgjør om det er en lineær sammenheng.



# Sentralgrenseteoremet



## Teorem

La  $X_i$  vere uavhengig identisk fordelte (u.i.f.) stokastiske variabler med  $E(X_i) = \mu_X$  og  $Var(X_i) = \sigma_X^2 < \infty$  for  $i = 1, \dots, n$ . La  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Då vil fordelinga til

$$Z = \frac{\bar{X}_n - E(\bar{X}_n)}{SD(\bar{X}_n)} = \frac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}}$$

gå mot ei standard normalfordeling når  $n \rightarrow \infty$ .

# Sentralgrenseteoremet

## Teorem

La  $X_i$  vere uavhengig identisk fordelte (u.i.f.) stokastiske variablar med  $E(X_i) = \mu_X$  og  $Var(X_i) = \sigma_X^2 < \infty$  for  $i = 1, \dots, n$ . La  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Då vil fordelinga til

$$Z = \frac{\bar{X}_n - E(\bar{X}_n)}{SD(\bar{X}_n)} = \frac{\bar{X}_n - \mu_X}{\sigma_X / \sqrt{n}}$$

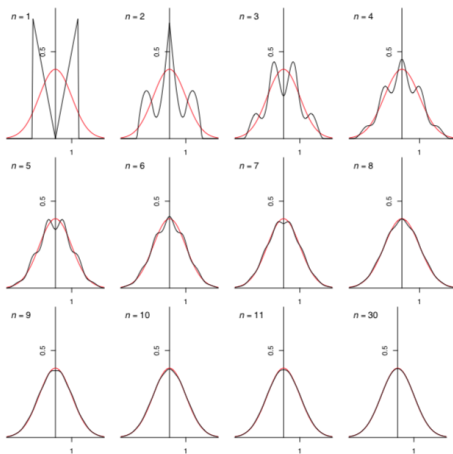
gå mot ei standard normalfordeling når  $n \rightarrow \infty$ .

Dette tilsvarar

$$\bar{X}_n \approx n \left( x; \mu_X, \frac{\sigma_X}{\sqrt{n}} \right).$$

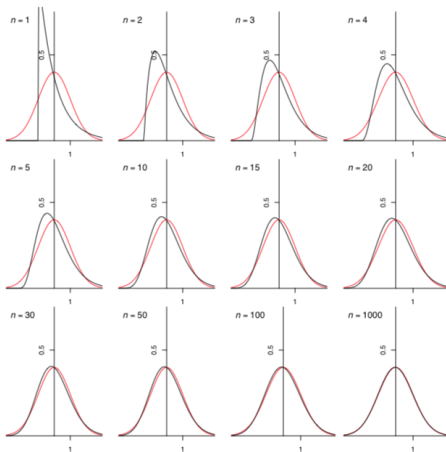
Merk: Sjølv om  $X_i$  ikkje er normalfordelt vil  $\bar{X}_n$  vere tilnærma normalfordelt når  $n$  er tilstrekkeleg stor (ofte nok med  $n \geq 30$ ).

# Sentralgrenseteoremet for ei symmetrisk fordeling





# Sentralgrenseteoremet for eksponentialfordelinga



## Forventningsrett estimator

Ein observator  $\hat{\theta}$  seies å verre ein forventningsrett (eng: unbiased) estimator for parameteren  $\theta$  dersom

$$E(\hat{\theta}) = \theta.$$

## Effisient estimator

Av fleire forventningsrette estimatorar for  $\theta$  seier me at den med minst varians er den mest effisiente.

Me føretrekk den mest effisiente estimatoren.

## Illustrasjon ved simulering

— Situasjon:

- $X_1, X_2, \dots, X_n$  tilfeldig utvalg fra  $n(x; \mu, \sigma)$ -populasjonen.
- Verdien til  $\mu$  er ukjent.
- Skal estimere (anslå) verdien til  $\mu$ .

— Naturlige estimatorer:

$$\hat{\mu}_1 = \bar{X} \text{ (gjennomsnitt)} \quad \text{og} \quad \hat{\mu}_2 = \tilde{X} \text{ (empirisk median)}$$

## Illustrasjon ved simulering

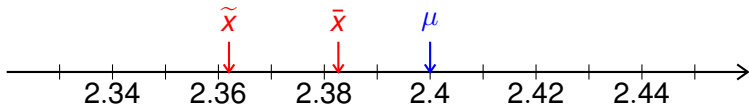
— Situasjon:

- $X_1, X_2, \dots, X_9$  tilfeldig utvalg fra  $n(x; \mu, \sigma)$ -populasjonen.
- Verdien til  $\mu$  er ukjent.
- Skal estimere (anslå) verdien til  $\mu$ .

— Generer observasjoner  $X_i \sim n(x_i; 2.4, 0.1)$ .

2.480, 2.262, 2.560, 2.281, 2.503, 2.362, 2.258, 2.282, 2.456

$$\bar{x} = 2.3827 \quad \text{og} \quad \tilde{x} = 2.362$$



## Illustrasjon ved simulering

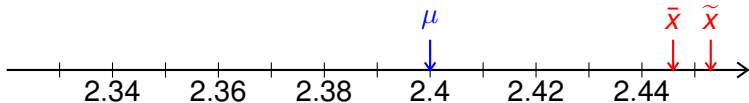
— Situasjon:

- $X_1, X_2, \dots, X_9$  tilfeldig utvalg fra  $n(x; \mu, \sigma)$ -populasjonen.
- Verdien til  $\mu$  er ukjent.
- Skal estimere (anslå) verdien til  $\mu$ .

— Generer observasjoner  $X_i \sim n(x_i; 2.4, 0.1)$ .

2.482, 2.407, 2.415, 2.406, 2.480, 2.453, 2.512, 2.331, 2.527

$$\bar{x} = 2.4459 \quad \text{og} \quad \tilde{x} = 2.453$$



## Illustrasjon ved simulering

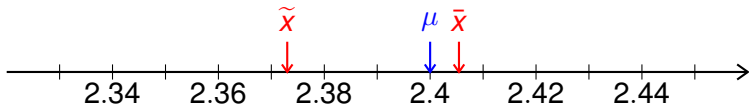
— Situasjon:

- $X_1, X_2, \dots, X_9$  tilfeldig utvalg fra  $n(x; \mu, \sigma)$ -populasjonen.
- Verdien til  $\mu$  er ukjent.
- Skal estimere (anslå) verdien til  $\mu$ .

— Generer observasjoner  $X_i \sim n(x_i; 2.4, 0.1)$ .

2.464, 2.277, 2.547, 2.393, 2.373, 2.369, 2.321, 2.369, 2.536

$$\bar{x} = 2.4054 \quad \text{og} \quad \tilde{x} = 2.373$$



## Illustrasjon ved simulering

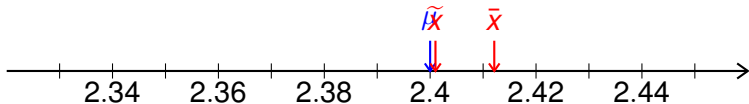
— Situasjon:

- $X_1, X_2, \dots, X_9$  tilfeldig utvalg fra  $n(x; \mu, \sigma)$ -populasjonen.
- Verdien til  $\mu$  er ukjent.
- Skal estimere (anslå) verdien til  $\mu$ .

— Generer observasjoner  $X_i \sim n(x_i; 2.4, 0.1)$ .

2.544, 2.401, 2.465, 2.430, 2.389, 2.461, 2.330, 2.325, 2.364

$$\bar{x} = 2.4121 \quad \text{og} \quad \tilde{x} = 2.401$$

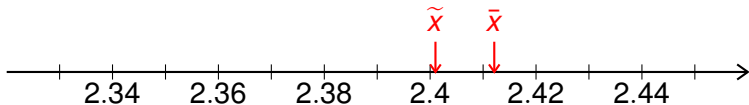


## Illustrasjon ved simulering

- Situasjon:
  - $X_1, X_2, \dots, X_9$  tilfeldig utvalg fra  $n(x; \mu, \sigma)$ -populasjonen.
  - Verdien til  $\mu$  er ukjent.
  - Skal estimere (anslå) verdien til  $\mu$ .
- Observerer verdier for  $X_1, X_2, \dots, X_9$ .

2.544, 2.401, 2.465, 2.430, 2.389, 2.461, 2.330, 2.325, 2.364

$$\bar{x} = 2.4121 \quad \text{og} \quad \tilde{x} = 2.401$$





## Illustrasjon ved simulering

- Situasjon:
  - $X_1, X_2, \dots, X_9$  tilfeldig utvalg fra  $n(x; \mu, \sigma)$ -populasjonen.
  - Verdien til  $\mu$  er ukjent.
  - Skal estimere (anslå) verdien til  $\mu$ .
- Gjentar forsøket 10 000 ganger.
  - Lager histogram over de 10 000 verdiene av  $\bar{x}$  og  $\tilde{x}$ .



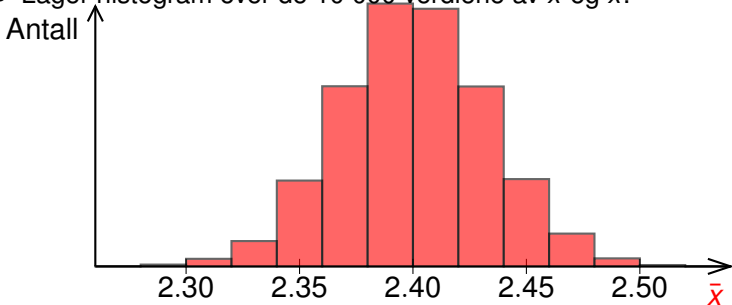
# Illustrasjon ved simulering

— Situasjon:

- $X_1, X_2, \dots, X_9$  tilfeldig utvalg fra  $n(x; \mu, \sigma)$ -populasjonen.
- Verdien til  $\mu$  er ukjent.
- Skal estimere (anslå) verdien til  $\mu$ .

— Gjentar forsøket 10 000 ganger.

- Lager histogram over de 10 000 verdiene av  $\bar{x}$  og  $\tilde{x}$ .



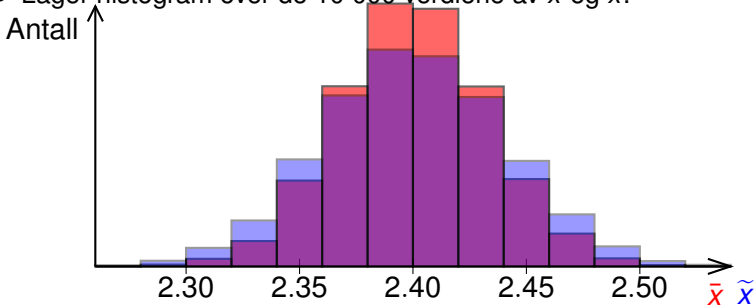
# Illustrasjon ved simulering

— Situasjon:

- $X_1, X_2, \dots, X_9$  tilfeldig utvalg fra  $n(x; \mu, \sigma)$ -populasjonen.
- Verdien til  $\mu$  er ukjent.
- Skal estimere (anslå) verdien til  $\mu$ .

— Gjentar forsøket 10 000 ganger.

- Lager histogram over de 10 000 verdiene av  $\bar{x}$  og  $\tilde{x}$ .



## Neste veke



- Sannsynsmaksimeringsestimator
- Konfidensintervall