Estimatorer

- Def: En observerbar funksjon av stokastiske variabler kalles en observator (eng: statistic)
- Evaluering av estimatorer:
  - situasjon: To estimatorer for θ: $\widehat{\theta}_1$ og $\widehat{\theta}_2$
  - spøsrmål: Hvilken estimator er best?
  - ide: tenker oss at vi gjentar forsøket/målingene (uendelig) mange ganger
    - foretrekker den estimatoren som vanligvis treffer nærmest den sanne verdien θ
  - MEN, vi gjør forsøket/målingene bare en gang!

Estimatorer

- Def: En observerbar funksjon av stokastiske variabler kalles en observator (eng: statistic)
- Evaluering av estimatorer:
  - situasjon: To estimatorer for θ: $\widehat{\theta}_1$ og $\widehat{\theta}_2$
  - spøsrmål: Hvilken estimator er best?
  - ide: tenker oss at vi gjentar forsøket/målingene (uendelig) mange ganger
    - foretrekker den estimatoren som vanligvis treffer nærmest den sanne verdien θ
  - MEN, vi gjør forsøket/målingene bare en gang!
- Def: En observator $\widehat{\theta}$ er en forventningsrett (unbiased) estimator for θ dersom $\mathrm{E}\left[\widehat{\theta}\right] = \theta$. Hvis ikke er $\widehat{\theta}$ forventningsskjev (biased)
- Def: Av flere forventingsrette estimatorer for θ sier vi at den med minst varians er mest effisient

Illustrasjon ved simulering

- Situasjon:
  - $X_1, X_2, \ldots, X_n$ tilfeldig utvalg fra $\mathrm{n}(x; \mu, \sigma)$-populasjonen.
  - Verdien til μ er ukjent.
  - Skal estimere (anslå) verdien til μ.
- Naturlige estimatorer:

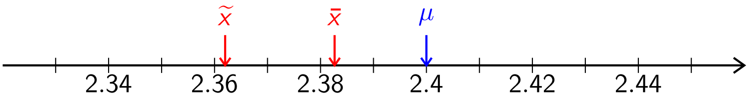
$$\hat{\mu}_1 = \bar{X} \text{ (gjennomsnitt)} \qquad \text{og} \qquad \hat{\mu}_2 = \widetilde{X} \text{ (empirisk median)}$$

Illustrasjon ved simulering

- Situasjon:
  - $X_1, X_2, \ldots, X_9$ tilfeldig utvalg fra n$(x; \mu, \sigma)$-populasjonen.
  - Verdien til μ er ukjent.
  - Skal estimere (anslå) verdien til μ.
- Generer observasjoner $X_i \sim \text{n}(x_i; 2.4, 0.1)$.

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

$$\bar{x} = 2.3827 \qquad \text{og} \qquad \widetilde{x} = 2.362$$

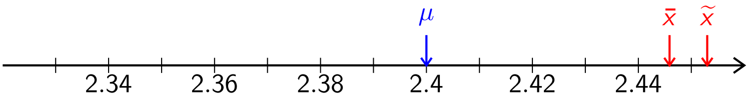


Illustrasjon ved simulering

- Situasjon:
  - $X_1, X_2, \ldots, X_9$ tilfeldig utvalg fra $n(x; \mu, \sigma)$-populasjonen.
  - Verdien til μ er ukjent.
  - Skal estimere (anslå) verdien til μ.
- Generer observasjoner $X_i \sim n(x_i; 2.4, 0.1)$.

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

$$\bar{x} = 2.4459 \qquad \text{og} \qquad \widetilde{x} = 2.453$$

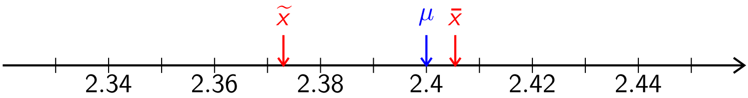


Illustrasjon ved simulering

- Situasjon:
  - $X_1, X_2, \ldots, X_9$ tilfeldig utvalg fra $\mathrm{n}(x; \mu, \sigma)$-populasjonen.
  - Verdien til μ er ukjent.
  - Skal estimere (anslå) verdien til μ.
- Generer observasjoner $X_i \sim \mathrm{n}(x_i; 2.4, 0.1)$.

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

$$\bar{x} = 2.4054 \qquad \text{og} \qquad \widetilde{x} = 2.373$$

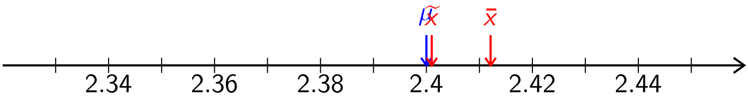


Illustrasjon ved simulering

- Situasjon:
  - $X_1, X_2, \ldots, X_9$ tilfeldig utvalg fra $n(x; \mu, \sigma)$-populasjonen.
  - Verdien til μ er ukjent.
  - Skal estimere (anslå) verdien til μ.
- Generer observasjoner $X_i \sim n(x_i; 2.4, 0.1)$.

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

$$\bar{x} = 2.4121 \qquad \text{og} \qquad \widetilde{x} = 2.401$$

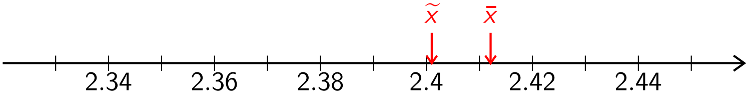


Illustrasjon ved simulering

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

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

$$\bar{x} = 2.4121 \qquad \text{og} \qquad \widetilde{x} = 2.401$$

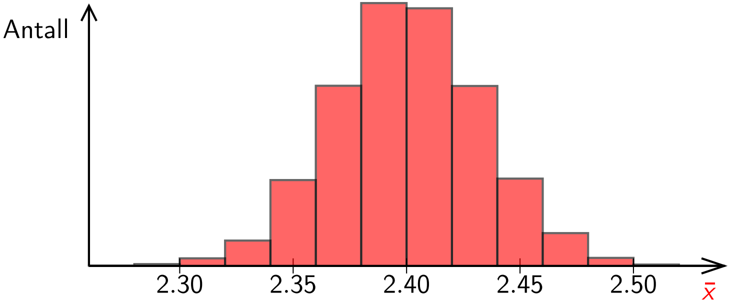


Illustrasjon ved simulering

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

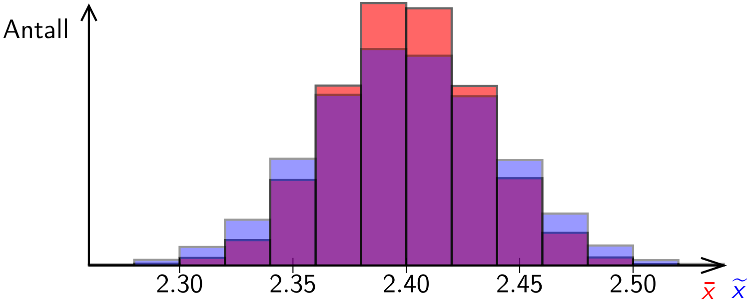
Illustrasjon ved simulering

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



Illustrasjon ved simulering

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



Oppsummering

- Krav til en god estimator:
  - Estimatoren skal være forventningsrett,
    $$\mathsf{E}\left(\hat{\theta}\right) = \theta.$$
  - Variansen til estimatoren skal være minst mulig,
    $$\mathsf{Var}\left(\hat{\theta}\right) \text{ minst mulig.}$$

Oppsummering

- Krav til en god estimator:
  - Estimatoren skal være forventningsrett,
$$\mathsf{E}\left(\hat{\theta}\right) = \theta.$$
  - Variansen til estimatoren skal være minst mulig,
$$\mathsf{Var}\left(\hat{\theta}\right) \text{ minst mulig.}$$
- Merk: Vi kan sjekke disse kravene uten å kjenne den sanne verdien til θ.