NTNU

Kunnskap for ei betre verd

Forelesing 18
Konfidensintervall

Torstein Fjeldstad

Institutt for matematiske fag, NTNU

07.03.2018

I dag

- Konfidensintervall

Sannsynsmaksimeringsestimator (SME)

Oppskrift for å finne SME

1. Finn simultanfordelinga til $X_1, X_2, \ldots, X_n$:

$$f(x_1, x_2, \ldots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \ldots f(x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

Sannsynsmaksimeringsestimator (SME)

Oppskrift for å finne SME

1. Finn simultanfordelinga til $X_1, X_2, \ldots, X_n$:

$$f(x_1, x_2, \ldots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \ldots f(x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

2. Definer rimelighetsfunksjonen (eng: likelihood function)

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

Sannsynsmaksimeringsestimator (SME)

Oppskrift for å finne SME

1. Finn simultanfordelinga til $X_1, X_2, \ldots, X_n$:

$$f(x_1, x_2, \ldots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \ldots f(x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

2. Definer rimelighetsfunksjonen (eng: likelihood function)

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

3. For å forenkle rekninga ser me ofte på log-rimelighetsfunksjonen

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i; \theta)$$

Sannsynsmaksimeringsestimator (SME)

Oppskrift for å finne SME

1. Finn simultanfordelinga til $X_1, X_2, \ldots, X_n$:

$$f(x_1, x_2, \ldots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \ldots f(x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

2. Definer rimelighetsfunksjonen (eng: likelihood function)

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

3. For å forenkle rekninga ser me ofte på log-rimelighetsfunksjonen

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i; \theta)$$

4. Maksimer (log-)likelihood mhp θ:

$$\hat{\theta}_{SME} = \arg\max_\theta L(\theta) = \arg\max_\theta l(\theta)$$

Kritiske verdier i standard normalfordelingen

$$P(Z > z_\alpha) = \alpha$$

α	z_α
.2	0.842
.15	1.036
.1	1.282
.075	1.440
.05	1.645
.04	1.751
.03	1.881
.025	1.960
.02	2.054
.01	2.326
.005	2.576
.001	3.090
.0005	3.291
.0001	3.719
.00005	3.891
.00001	4.265
.000005	4.417
.000001	4.753

Konfidensintervall for μ i normalfordelinga (σ kjend)

Situasjon: Anta $X_1, X_2, \ldots, X_n$ tilfeldig utval frå $n(x; \mu, \sigma)$-populasjonen der σ er kjend. Ynskjer eit $(1 - \alpha) \cdot 100\%$ konfidensintervall for θ.

1. Estimator for μ : $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$
2. Veit at $\bar{X} \sim n(\bar{x}; \mu, \frac{\sigma}{\sqrt{n}})$, derfor er
$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim n(z; 0, 1)$$
3. Har at
$$P(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \leq z_{\alpha/2}) = 1 - \alpha$$
4. Løyser ulikskapane for μ og får
$$P(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$
5. Eit $(1 - \alpha) \cdot 100$ % konfidensintervall for μ er
$$\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Utledning av konfidensintervall

Situasjon: Anta $X_1, X_2, \ldots, X_n$ tilfeldig utval frå $f(x;\theta)$-populasjonen. Ynskjer eit $(1-\alpha)\cdot 100\%$ konfidensintervall for μ.

1. Finn ein estimator for θ, $\hat{\theta}$ (t.d. SME)
2. La $Z = h(\hat{\theta}, \theta)$ der $h(\cdot,\cdot)$ er ein funksjon s.a. Z har ei kjend fordeling.
3. Då har me at
$$P(z_{1-\alpha/2} \leq h(\hat{\theta}, \theta) \leq z_{\alpha/2}) = 1-\alpha$$
4. Løys ulikskapane (mhp. θ) kvar for seg og finn eit uttrykk med θ i midten
$$P(\hat{\theta}_L(X_1, X_2, \ldots, X_n) \leq \theta \leq \hat{\theta}_U(X_1, X_2, \ldots, X_n)) = 1-\alpha$$
5. Et $(1-\alpha)\cdot 100$ % konfidensintervall for θ er
$$\hat{\theta}_L(X_1, X_2, \ldots, X_n), \hat{\theta}_U(X_1, X_2, \ldots, X_n)$$

Neste veke

— Meir om konfidensintervall