Statistikk

Jo Eidsvik

Matematiske fag, NTNU

Estimering og konfidensintervall

- Data $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt med tetthet eller punktsannsynlighet $f(x; \theta)$.
- Estimator for modellparameter $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$.
- Konfidensintervall
  $P(\theta_L(X_1, X_2, \ldots, X_n) < \theta < \theta_U(X_1, X_2, \ldots, X_n)) = 1 - \alpha$.
- Ved innsetting av data har vi 'tall' $\hat{\theta}(x_1, x_2, \ldots, x_n)$, $\theta_L(x_1, x_2, \ldots, x_n)$, $\theta_U(x_1, x_2, \ldots, x_n)$, men som funksjoner av stokastiske variable vil de variere.

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. $\bar{X}$ estimerer μ.
2. $Z = \sqrt{n}\frac{\bar{X}-\mu}{\sigma}$, Z er normalfordelt med $E(Z) = 0$, $Var(Z) = 1$.
3. $P(z_{\alpha/2} < \sqrt{n}\frac{\bar{X}-\mu}{\sigma} < z_{1-\alpha/2}) = 1 - \alpha$
4. $z_{1-\alpha/2} = -z_{\alpha/2}$, $z_{0.025} = -1.96$, $z_{0.05} = -1.645$.
5. Konfidensintervall:
   $P(\bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n} < \mu < \bar{X} - z_{\alpha/2}\sigma/\sqrt{n}) = 1 - \alpha$

Varians i et utvalg

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

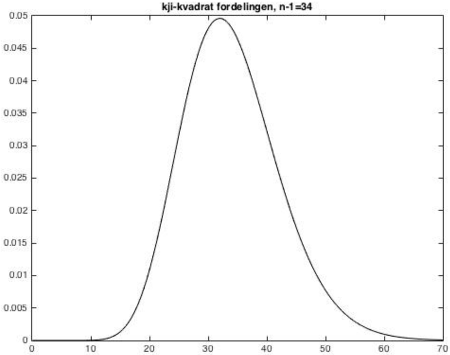
$$\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

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

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

Kji-kvadrat fordeling

$\chi^2_{34,0.05} = 22$, $\chi^2_{34,0.95} = 49$.



Eksempel - akslinger

$X_1, \ldots, X_{35}$, Estimat for σ^2 er $S^2 = 0.0029$. Hvor sikkert er det?

$$P(\chi^2_{34,0.05} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{34,0.95}) = 0.9$$

$$P(\frac{(n-1)S^2}{\chi^2_{34,0.95}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{34,0.05}}) = 0.9$$

Konfidensintervall blir $(\frac{34 \cdot 0.0029}{22}, \frac{34 \cdot 0.0029}{22}) = (0.0020, 0.0045)$.

T fordeling

- Antakelse: $X_i \sim N(\mu, \sigma^2)$, $i = 1, \ldots, n$, uavhengige.
- Vi søker konfidensintervall for μ. Vi vet ikke σ^2.
- Estimator $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- $s^2 = \sum_{i=1}^{n} \frac{1}{n-1}(X_i - \bar{X})^2$ estimerer σ^2.
- $T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$

T fordelingen ligner en normalfordeling. Symmetrisk om 0. Men mer sannsynlighet for noe veldig lite eller veldig stort (større haler).

Eksempel - deteksjonsavstander for flaggermus

- Avstander $X_1, X_2, \dots, X_n$. $n = 11$
- Antakelser $X_i \sim N(\mu, \sigma^2)$, $i = 1, \dots, 11$, uavhengige.
- Vi søker konfidensintervall for μ.
- Estimator $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
- $s^2 = \sum_{i=1}^{n} \frac{1}{n-1} (X_i - \bar{X})^2$
- $T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$

Eksempel - deteksjonsavstander for flaggermus

Punktestimat for μ er $\bar{x} = 48$ cm.
Konfidensintervall finnes fra fordeling.

$$P(-t_{10,\alpha/2} < \frac{\bar{X} - \mu}{s/\sqrt{n}} < t_{10,\alpha/2}) = 1 - \alpha$$

$$P(\bar{X} - t_{10,\alpha/2} s/\sqrt{n} < \mu < \bar{X} + t_{10,\alpha/2} s/\sqrt{n}) = 1 - \alpha$$

$S^2 = 18^2$ cm^2. $\alpha = 0.05$. $t_{10,0.025} = 2.23$. $(\mu_{Lower}, \mu_{Upper}) = (36, 61)$.