

Estimering og konfidensintervaller for sannsynligheten p og raten λ

Thea Bjørnland
Institutt for matematiske fag
NTNU

Repetisjon

Binomisk forsøksrekke

- Det gjøres n uavhengige forsøk
- Hvert forsøk har to utfall: “suksess og fiasko”
- Sannsynligheten for “suksess” er den samme i alle forsøk

$$P(\text{suksess}) = p$$

X : antall “suksesser” i n forsøk

$$X \sim \text{binomisk}(n, p)$$

$$E(X) = np$$

$$\text{Var}(X) = np(1 - p)$$

$$\hat{p} = \frac{X}{n}$$

Antall suksesser (pointing to X)

Antall forsøk (pointing to n)

Eks:
 $n = 10$ forsøk,
 $x = 7$ suksesser
 $\hat{p} = \frac{7}{10} = 0.7$

$$E(\hat{p}) = \frac{E(X)}{n} = \frac{np}{n} = p \quad \text{Forventningsrett!}$$

$$\text{Var}(\hat{p}) = \frac{\text{Var}(X)}{n^2} = \frac{np(1 - p)}{n^2} = \frac{p(1 - p)}{n}$$

Jo flere forsøk,
jo lavere varians

Repetisjon

Binomisk forsøksrekke

- Det gjøres n uavhengige forsøk
- Hvert forsøk har to utfall: “suksess og fiasko”
- Sannsynligheten for “suksess” er den samme i alle forsøk

$$P(\text{suksess}) = p$$

X : antall “suksesser” i n forsøk

$$X \sim \text{binomisk}(n, p)$$

$$E(X) = np$$

$$\text{Var}(X) = np(1 - p)$$

Tilnærmet normalfordelt

$$\hat{p} = \frac{X}{n} \quad E(\hat{p}) = p \quad \text{Var}(\hat{p}) = \frac{p(1 - p)}{n}$$

Sentralgrenseteoremet!

$$SE(\hat{p}) \approx \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

Tilnærmet $(1 - \alpha)100\%$ konfidensintervall for p :

$$\left[\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \quad \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right]$$

$$n\hat{p}(1 - \hat{p}) \geq 5$$

Eks:

$n = 50$ forsøk,

$$50 \cdot 0.4 \cdot (1 - 0.4) = 12$$

$x = 20$ suksesser

$$\hat{p} = \frac{20}{50} = 0.4 \quad SE(\hat{p}) \approx \sqrt{\frac{0.4 \cdot (1 - 0.4)}{50}} \approx 0.07$$

$$\alpha = 0.05 \quad z_{0.025} = 1.96$$

95% konfidensintervall for p : $[0.26, 0.54]$

Repetisjon

Binomisk forsøksrekke

- Det gjøres n uavhengige forsøk
- Hvert forsøk har to utfall: “suksess og fiasko”
- Sannsynligheten for “suksess” er den samme i alle forsøk

$$P(\text{suksess}) = p$$

X : antall “suksesser” i n forsøk

$$X \sim \text{binomisk}(n, p)$$

$$E(X) = np$$

$$\text{Var}(X) = np(1 - p)$$

Tilnærmet normalfordelt

$$\hat{p} = \frac{X}{n} \quad E(\hat{p}) = p \quad \text{Var}(\hat{p}) = \frac{p(1 - p)}{n}$$

Sentralgrenseteoremet!

$$SE(\hat{p}) \approx \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

Tilnærmet $(1 - \alpha)100\%$ konfidensintervall for p :

$$\left[\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \quad \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right]$$

$n\hat{p}(1 - \hat{p}) \geq 5$

Eks:

$n = 500$ forsøk,

$x = 200$ suksesser

$$\hat{p} = \frac{200}{500} = 0.4 \quad SE(\hat{p}) \approx \sqrt{\frac{0.4 \cdot (1 - 0.4)}{500}} \approx 0.02$$

$$\alpha = 0.05 \quad z_{0.025} = 1.96$$

95% konfidensintervall for p : $[0.36, 0.44]$

Repetisjon

Poissonprosess (i tid)

- Hendelser skjer uavhengig av hverandre
- To hendelser kan ikke skje akkurat samtidig
- Forventet antall hendelser er det samme når som helst:

λ hendelser / tidsenhet

Y : antall hendelser i et intervall av lengde t

$$Y \sim \text{Poisson}(\lambda t)$$

$$E(Y) = \lambda t$$

$$\text{Var}(Y) = \lambda t$$

$\hat{\lambda} = \frac{Y}{t}$

↖ Antall hendelser
↗ Lengden på intervallet

Eks:
 $t = 10$ minutter,
 $y = 7$ hendelser
 $\hat{\lambda} = \frac{7}{10} = 0.7$

$$E(\hat{\lambda}) = \frac{E(Y)}{t} = \frac{\lambda t}{t} = \lambda \quad \text{Forventningsrett!}$$

$$\text{Var}(\hat{\lambda}) = \frac{\text{Var}(Y)}{t^2} = \frac{\lambda t}{t^2} = \frac{\lambda}{t}$$

Jo lengre vi observerer prosessen,
jo lavere blir variansen

Repetisjon

Poissonprosess (i tid)

- Hendelser skjer uavhengig av hverandre
- To hendelser kan ikke skje akkurat samtidig
- Forventet antall hendelser er det samme når som helst:

λ hendelser / måleenhet

Y : antall hendelser i et intervall av lengde t

$$Y \sim \text{Poisson}(\lambda t)$$

$$E(Y) = \lambda t$$

$$\text{Var}(Y) = \lambda t$$

Tilnærmet normalfordelt

$$\hat{\lambda} = \frac{Y}{t} \quad E(\hat{\lambda}) = \lambda \quad \text{Var}(\hat{\lambda}) = \frac{\lambda}{t}$$

$$SE(\hat{\lambda}) \approx \sqrt{\frac{\hat{\lambda}}{t}}$$

Tilnærmet $(1-\alpha)100\%$ konfidensintervall for λ :

$$\left[\hat{\lambda} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{\lambda}}{t}}, \quad \hat{\lambda} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{\lambda}}{t}} \right]$$

$$\hat{\lambda} t \geq 10$$

Eks:

$t = 50$ min,

$y = 120$ suksesser

$$2.4 \cdot 50 = 120$$

$$\hat{\lambda} = \frac{120}{50} = 2.4 \quad SE(\hat{\lambda}) \approx \sqrt{\frac{2.4}{50}} \approx 0.219$$

$$\alpha = 0.1 \quad z_{0.05} = 1.645$$

90% konfidensintervall for λ : $[2.04, 2.76]$