Linnéa Gustafsson
linneag2@kth.se

F12 KONFIDENSINTERVALL

- Konfidensintervall för varians och standardavvikelse i normalfördelade data
- Konfidensintervall m.h.a. normalapproximation
- Felfortplantning

Kap. 12

Approximativa konfidensintervall

Se § 12.3 i F.S.: § 12.3 används när mätdata inte är normalfördelade men vårt θ^* är approximativt N-fördelad

§ 12.3: $I_\theta = \theta^*_{obs} \pm \underbrace{d(\theta^*)}_{\text{medelfelet } (D^*_{obs})} \cdot \lambda_{\alpha/2}$

Om vi vill använda § 12.3 ska θ^* vara approximativt N-fördelad
Då fås ett konfidensintervall med approximativ konfidensgrad $1-\alpha$

CGS-fallet

Anta att vi har två stickprov $x_1, \ldots, x_n$
$y_1, \ldots, y_n$

X_i:na och Y_j:na kan dock inte sägas vara N-fördelade

$E(X_i) = \mu_x \qquad E(Y_j) = \mu_y$

$D(X_i) = \sigma_x \qquad D(Y_j) = \sigma_y \qquad \sigma_x$ och σ_y okända

Vi vill nu bilda konfidensintervall för $\mu_x - \mu_y$ och vill använda § 12.3

$$I_\theta = \theta^*_{obs} \pm D^*_{obs} \cdot \lambda_{\alpha/2}$$

Villkoret för detta är enl. § 12.3 att θ^* är approximativt N-fördelad

<u>Vad är vårt θ^*?</u>

- $\theta = \mu_x - \mu_y$
- $\theta^*_{obs} = \bar{x} - \bar{y}$
- $\theta^* = \bar{X} - \bar{Y}$

$\bar{X} - \bar{Y} = \frac{1}{n} \underbrace{\sum x_i}_{Z} - \frac{1}{m} \underbrace{\sum y_j}_{W}$

Stickprov $\Rightarrow$ oberoende

Om nu X_i:na och Y_j:na är ober., N-fördelade och många $\Rightarrow$ CGS $\Rightarrow$

$\Rightarrow$ Z och W är approximativt N-fördelade

$$\theta^* = \frac{1}{n} \underset{\sim N}{Z} - \frac{1}{m} \underset{\sim N}{W}$$

Enl. § 4 är varje linjärkombination av ober. N-fördelade stok. var. också N-fördelad

$\Rightarrow$ § 12.3 får användas

$$I_\theta = \theta^*_{obs} \pm D^*_{obs} \cdot \lambda_{\alpha/2}$$

$\theta = \mu_x - \mu_y$

$\theta^*_{obs} = \bar{x} - \bar{y}$

D^*_{obs} ?

Enligt § 12.3 gäller att $D = \underbrace{D(\theta^*)}_{\text{standardavvikelsen för } \theta^*} = D(\bar{X} - \bar{Y})$

$$V(\bar{X} - \bar{Y}) = \ldots = \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}$$

$$D = \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}$$

$$D^*_{obs} = \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}^{\,*}_{obs} = \sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}$$

$$I_{\mu_x - \mu_y} = \bar{x} - \bar{y} \pm \sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}} \cdot \lambda_{\alpha/2}$$

Binomialfördelningsfallet

Vi har $X \in Bin(n,p)$. Vi skattar p med $p^*_{obs} = \frac{x}{n}$ och vill bilda ett konf-int för p.

Vi har ej N-fördelning $\Rightarrow$ Vi vill använda § 12.3 : $I_\theta = \theta^*_{obs} \pm D^*_{obs} \cdot \lambda_{\alpha/2}$

Enl. § 12.3 måste θ^* vara approximativt N-fördelad.

<u>Vad är θ^* i vårt fall?</u>

- $\theta = p$
- $\theta^*_{obs} = p^*_{obs} = \frac{x}{n}$
- $\theta^* = \frac{X}{n}$

Om X är approximativt N-fördelad är även θ^* det pga linjärkombination (se § 4)

$X \in Bin(n, p)$

Se § 6: om $np(1-p) \geq 10$, har ej $p \Rightarrow np^*_{obs}(1-p^*_{obs}) \geq 10$ gäller att:

X approximativt N-fördelad $\Rightarrow$ θ^* approximativt N-fördelad $\Rightarrow$ § 12.3 får användas

Nu kan vi börja bilda konf-int:

$$I_\theta = \theta^* \pm D^*_{obs} \cdot \lambda_{\alpha/2}$$

$$\theta = p$$
$$\theta^*_{obs} = p^*_{obs} = \frac{x}{n}$$

Enl. § 12.3 är $D = D(\theta^*) = D\left(\frac{X}{n}\right)$

$$V\left(\frac{X}{n}\right) = \frac{1}{n^2} V(X) = [\text{Bin-fördeln., se § 3}] = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$$

$$\Rightarrow D(p^*) = D\left(\frac{X}{n}\right) = \sqrt{\frac{p(1-p)}{n}}$$

$$D^*_{obs} = \sqrt{\frac{p(1-p)}{n}}^*_{obs} = \sqrt{\frac{p^*_{obs}(1-p^*_{obs})}{n}} \quad \left(p^*_{obs} = \frac{x}{n} \right)$$

$$\Rightarrow I_p = p^*_{obs} \pm \sqrt{\frac{p^*_{obs}(1-p^*_{obs})}{n}} \cdot \lambda_{\alpha/2}$$

Konf-int för skillnad mellan andelar

Vi har $X \in Bin(n_x, p_x)$, $Y \in Bin(n_y, p_y)$

Vi vill bilda konf-int för $p_x - p_y$

Om nu $n_x p^*_{x_{obs}}(1-p^*_{x_{obs}}) \geq 10$ och $n_y p^*_{y_{obs}}(1-p^*_{y_{obs}}) \geq 10$ så kan §12.3 användas

$$I_\theta = \theta^* \pm D^*_{obs} \cdot \lambda_{\alpha/2}$$

$$\theta = p_x - p_y$$

$$\theta^*_{obs} = p^*_{x_{obs}} - p^*_{y_{obs}} = \frac{x}{n_x} - \frac{y}{n_y}$$

$$D = [\text{enl. §12.3}] = D(\theta^*) = D\left(\frac{X}{n_x} - \frac{Y}{n_y}\right)$$

$$V\left(\frac{X}{n_x} - \frac{Y}{n_y}\right) = [\text{ober.}] = \frac{1}{n_x^2} V(X) + \frac{1}{n_y^2} V(Y) = [\text{Bin-fördelade}] =$$

$$= \frac{1}{n_x^2} n_x p_x (1-p_x) + \frac{1}{n_y^2} n_y p_y (1-p_y)$$

$$\Rightarrow I_{p_x - p_y} = p^*_{x_{obs}} - p^*_{y_{obs}} \pm \sqrt{\frac{p^*_{x_{obs}}(1-p^*_{x_{obs}})}{n_x} + \frac{p^*_{y_{obs}}(1-p^*_{y_{obs}})}{n_y}} \cdot \lambda_{\alpha/2}$$

Poissonfördelningsfallet

Vi antar att $X_i \in Po(\mu)$. Enl. § 3 är då $\mu = E(X)$. Vi skattar μ med $\bar{x}$.

Vi vill bilda konf-int nu för μ.

Här har vi inte N-fördelning så vi vill använda §12.3 : $I_\theta = \theta^*_{obs} \pm D^*_{obs} \cdot \lambda_{\alpha/2}$

Vi ska visa att vårt θ^* är approximativt N-fördelad för att få använda §12.3 (se §12.3)

Vad är vårt θ^*?

- $\theta = \mu$
- $\theta^*_{obs} = \mu^*_{obs} = \bar{x}$
- $\theta^* = \bar{X} = \frac{1}{n} \sum X_i$ måste vara approximativt N-fördelad

Pga lin. komb. etc. (se §4) gäller att om $\sum X_i$ approximativt N-fördelad så är även $\bar{X}$ det.

X_i:na är Poissonfördelade

Sats

Om $X_1 \in Po(\mu_1)$ och $X_2 \in Po(\mu_2)$ och X_1 och X_2 är ober $\Rightarrow X_1 + X_2$ ober.

$\Rightarrow \sum X_i \in Po(n \cdot \mu)$

Se §6: $Po(\mu) \sim N(\mu, \sqrt{\mu})$ om $\mu \geq 15$

I vårt fall är då $\sum X_i \sim N(n\mu, \sqrt{n\mu})$ om $n\mu \geq 15$

$\Rightarrow$ Om $n\mu^*_{obs} = n\bar{x} \geq 15$ så är $\sum X_i \sim$ N-fördelad.

$\Rightarrow \theta^* = \frac{1}{n} \sum X_i \sim$ N-fördelad

$\Rightarrow$ § 12.3 kan användas

$I_\theta = \theta^*_{obs} \pm D^*_{obs} \cdot \lambda_{\alpha/2}$

$\theta = \mu$

$\theta^*_{obs} = \mu^*_{obs} = \bar{x}$

D^*_{obs}?

$D = [\text{enl. § 12.3}] = D(\theta^*) = D(\bar{X})$

$V(\bar{X}) = V\left(\frac{1}{n} \sum X_i\right) = \frac{1}{n^2} V(\sum X_i) = [\text{se F.S. § 3}] = \frac{1}{n^2} \cdot n \cdot \mu = \frac{\mu}{n}$

$D(\bar{X}) = \sqrt{\frac{\mu}{n}}$

$D^*_{obs} = \sqrt{\frac{\mu}{n}}^*_{obs} = \sqrt{\frac{\mu^*_{obs}}{n}} = \sqrt{\frac{\bar{x}}{n}}$

$\Rightarrow I_\mu = \bar{x} \pm \sqrt{\frac{\bar{x}}{n}} \cdot \lambda_{\alpha/2}$ (I detta fall använder vi inte stickprovsvariansen s^2)

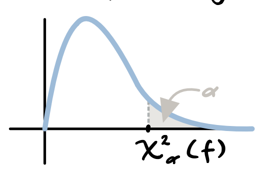
Konfidensintervall för σ och σ^2

"Chi-två"

Bakgrund

Se § 10: om $X_1, X_2, \dots, X_f \in N(0,1)$ och ober. $\Rightarrow \sum_{i=1}^{f} X_k^2 \in \chi^2(f)$

χ^2-fördelningen (se tab. 4):



$$P(X > \chi^2_\alpha(f)) = \alpha$$

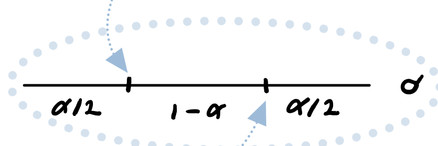
Anta att X_i:na är ober. och $N(\mu, \sigma)$ $\Rightarrow \sum_{i=1}^{n} \left(\frac{X_i - \bar{x}}{\sigma}\right)^2 \in \chi^2(n-1)$

§ 11.2 b: $\frac{1}{\sigma^2} \sum (X_i - \bar{X})^2 = \frac{(n-1)}{\sigma^2} \frac{1}{n-1} \sum (X_i - \bar{X})^2 = \frac{(n-1)}{\sigma^2} S^2 \in \chi^2(n-1)$

Titta nu på tab. 4: $P\left(\frac{(n-1)S^2}{\sigma^2} > \chi^2_\alpha(n-1)\right) = \alpha$

$$P\left(\sigma < \sqrt{\frac{(n-1)S^2}{\chi^2_\alpha(n-1)}}\right) = \alpha$$

$$P\left(\sigma < \sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}}\right) = \frac{\alpha}{2}$$



$$P\left(\sigma < \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}}\right) = 1 - \frac{\alpha}{2}$$

$$\Rightarrow I_\sigma = \left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}}, \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}}\right)$$

Hur man tar fram I_σ från §12.4:

$$I_\theta = \left(\theta^*_{obs}\sqrt{\frac{f}{\chi^2_{\alpha/2}(f)}},\ \theta^*_{obs}\sqrt{\frac{f}{\chi^2_{1-\alpha/2}(f)}}\right)$$

Jämför §12.4: $f\cdot\left(\frac{\theta^*}{\theta}\right)^2 \in \chi^2(f)$

med §11.1 b): $(n-1)\cdot\frac{S^2}{\sigma^2} \in \chi^2(n-1)$

$$\Rightarrow \begin{cases} \theta = \sigma \\ \theta^*_{obs} = s \\ f = n-1 \end{cases}$$

$$I_\sigma = \left(s\sqrt{\frac{n-1}{\chi^2_{\alpha/2}(n-1)}},\ s\sqrt{\frac{n-1}{\chi^2_{1-\alpha/2}(n-1)}}\right)$$

$$I_{\sigma^2} = \left(\frac{s^2(n-1)}{\chi^2_{\alpha/2}(n-1)},\ \frac{s^2(n-1)}{\chi^2_{1-\alpha/2}(n-1)}\right)$$

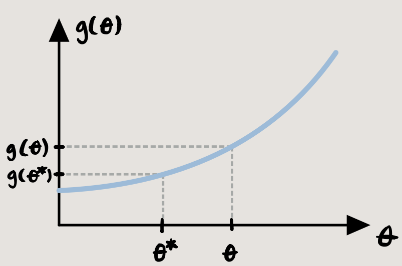
Felfortplantning

se kap. 11 sid. 273

Ex Övningsuppgift 11.13 b)

Vi har fått skattningen $\lambda^*_{obs} = \frac{1}{\bar{x}}$ där X_i:na är ober. och $Exp(\lambda)$.
Vi vill ha $E(\lambda^*)$ och medelfelet $D^*(\lambda^*)_{obs}$ men λ^* beror ej linjärt av X_i:na så vi kan inte använda oss av de vanliga räknereglerna för väntevärde och varians.

Härledning av § 9.4 i F.S.



Väntevärdesriktig

Vi antar att θ^*_{obs} är en vvr skattning av θ

Taylorutveckling:

$$g(\theta) = g(\theta) + g'(\theta)(\theta^*-\theta) + \underbrace{\frac{g''(\theta)}{2!}(\theta^*-\theta)^2 + \dots}_{\text{restterm}}$$

$$E[g(\theta^*)] \approx E[\underbrace{g(\theta)}_{\text{konstant}} + \underbrace{g'(\theta)}_{\text{konstant}}(\theta^*-\theta)] = E[g(\theta)] + g'(\theta)E[\theta^*-\theta] =$$

$$= E[g(\theta)] + g'(\theta)[\underbrace{E(\theta^*)}_{\theta \text{ p.g.a. vvr}} - \theta] = E[\underbrace{g(\theta)}_{\text{konstant (ur stokastisk synpunkt)}}] = g(\theta) \approx g(\theta^*_{obs})$$

§ 9.4 a): $\boxed{1)\ E[g(\theta^*)] \approx g(\theta^*_{obs})}$

$$V[g(\theta^*)] \approx V[\underbrace{g(\theta)}_{\text{konstant}} + \underbrace{g'(\theta)}_{\text{konstant}}(\theta^* - \theta)] = [g'(\theta)]^2 \cdot V(\theta^* - \underbrace{\theta}_{\text{konstant}}) =$$

$$= [g'(\theta)]^2 \cdot V(\theta^*)$$

$$D[g(\theta^*)] = |g'(\theta)| \cdot D(\theta^*)$$

$$d[g(\theta^*)] = D^*_{obs}[g(\theta^*)] = |g'(\theta^*_{obs})| \cdot D^*_{obs}(\theta^*)$$

§ 9.4 a): $\boxed{2)\ D[g(\theta^*)] \approx |g'(\theta^*_{obs})| \cdot D(\theta^*)}$

Tillbaka till övningsuppgiften 11.13 b)

$\lambda^*_{obs} = \frac{1}{\bar{x}}$, X_i:na oberoende och $Exp(\lambda)$

Sökt: Medelfelet för skattningen av λ

Sökt enl. § 9.3: $D^*(\lambda^*)_{obs}$ när vi har $\lambda^* = \frac{1}{\bar{X}}$ $X_i \in Exp(\lambda)$

Se § 9.4: $D(g(\theta^*)) \approx |g'(\theta^*_{obs})| \cdot D(\theta^*)$

I vårt fall är $g(\theta^*) = \lambda^*(\bar{X}) = \frac{1}{\bar{X}}$

$g(\theta^*_{obs}) = \frac{1}{\bar{x}}$

$g'(\theta^*_{obs}) = -\frac{1}{\bar{x}^2}$

$|g'(\theta^*_{obs})| = \frac{1}{\bar{x}^2}$

$D(\theta^*) = D(\bar{X})$

$V(\bar{X}) = V[\frac{1}{n}(x_1 + x_2 + \dots + x_n)] = \frac{1}{n^2} \cdot n \cdot V(X_i) = \frac{V(X_i)}{n} = [Exp(\lambda)] = \frac{1}{\lambda^2 n}$

$\Rightarrow D(\bar{X}) = \frac{1}{\lambda\sqrt{n}}$

$D(g(\theta^*)) \approx D(\lambda^*) = \frac{1}{\bar{x}^2} \cdot \frac{1}{\lambda\sqrt{n}}$

Medelfelet $= D^*(g(\theta^*))_{obs} = \frac{1}{\bar{x}^2} \cdot \frac{1}{\lambda^*_{obs}\sqrt{n}} = \frac{1}{\bar{x}^2} \cdot \frac{\bar{x}}{\sqrt{n}} = \frac{1}{\bar{x}\sqrt{n}}$