

Utgangspunkt for konfidensintervall i en del situasjoner

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland

Institutt for matematiske fag

Norges teknisk-naturvitenskapelige universitet

Utleddning av konfidensintervall

★ Prosedyre for å utlede et konfidensintervall for parameter θ basert på $x_1, x_2, \dots, x_n$

1. Bestemme pivotal, $Z = h(\theta, X_1, X_2, \dots, X_n)$
 - inneholder ingen ukjente parametre unntatt θ
 - har en kjent fordeling (som ikke avhenger av θ)

2. Finne kvantiler $z_{1-\frac{\alpha}{2}}$ og $z_{\frac{\alpha}{2}}$ slik at

$$P\left(z_{1-\frac{\alpha}{2}} \leq h(\theta, x_1, x_2, \dots, x_n) \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

3. Løse hver ulikhet med hensyn på θ
4. Sette ulikhetene sammen igjen med θ i midten

$$P\left(\hat{\theta}_L(x_1, x_2, \dots, x_n) \leq \theta \leq \hat{\theta}_U(x_1, x_2, \dots, x_n)\right) = 1 - \alpha$$

5. Konkludere ved å skrive opp $(1 - \alpha) \cdot 100\%$ -konfidensintervallet for θ

$$\left[\hat{\theta}_L(x_1, x_2, \dots, x_n), \hat{\theta}_U(x_1, x_2, \dots, x_n)\right]$$

★ Fokus i denne videoen: Valg av pivotal

Hvilke situasjoner skal vi se på?

- ★ $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $N(\mu, \sigma^2)$ -populasjon
 - konfidensintervall for μ , σ^2 er kjent
 - konfidensintervall for μ , σ^2 er ukjent
 - konfidensintervall for σ^2 , μ er kjent
 - konfidensintervall for σ^2 , μ er ukjent
- ★ X er binomisk fordelt med parametre n og p
 - konfidensintervall for p
- ★ To-utvalg: $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $N(\mu_1, \sigma_1^2)$ -populasjonen og $Y_1, Y_2, \dots, Y_m$ tilfeldig utvalg fra $N(\mu_2, \sigma_2^2)$ -populasjonen, X_i -ene og Y_i -ene uavhengige
 - konfidensintervall for $\mu_1 - \mu_2$, σ_1^2 og σ_2^2 kjente

Konfidensintervall i normalfordeling, en ukjent parameter

$$Z \sim N(0, 1) \Rightarrow Z^2 \sim \chi_1^2$$

$$V_i \sim \chi_{\nu_i}^2, i = 1, 2, \dots, n \text{ uavh.} \Rightarrow \sum_{i=1}^n V_i \sim \chi_{\nu}^2 \text{ med } \nu = \sum_{i=1}^n \nu_i$$

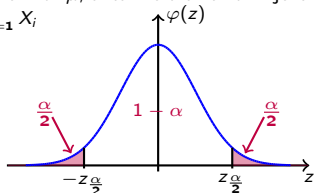
★ Situasjon:

- la $X_1, X_2, \dots, X_n$ være tilfeldig utvalg fra $N(\mu, \sigma^2)$ -populasjonen
- verdien til μ er ukjent, ønsker et konfidensintervall for μ , antar verdien til σ^2 kjent

★ Tar utgangspunkt i estimator for μ : $\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

★ Vi vet at

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0, 1)$$



★ Situasjon:

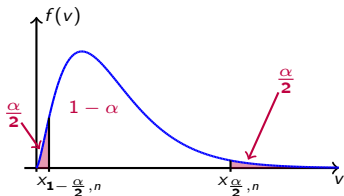
- la $X_1, X_2, \dots, X_n$ være tilfeldig utvalg fra $N(\mu, \sigma^2)$ -populasjonen
- verdien til σ^2 er ukjent, ønsker et konfidensintervall for σ^2 , antar verdien til μ kjent

★ Tar utgangspunkt i SME for σ^2 : $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$

★ Vi vet at

$$V = \frac{n\hat{\sigma}^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi_n^2$$

$$\frac{X_i - \mu}{\sigma} \sim N(0, 1), \quad \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi_1^2$$



Konfidensintervall for μ når også verdien til σ^2 er ukjent

Definisjon: Anta $Z \sim N(0, 1)$, $V \sim \chi^2_\nu$,
og Z og V uavh. Da er $\frac{Z}{\sqrt{\frac{V}{\nu}}} \sim t_\nu$

★ Situasjon:

- la $X_1, X_2, \dots, X_n$ være tilfeldig utvalg fra $N(\mu, \sigma^2)$ -populasjonen
- verdien til μ er ukjent, ønsker et konfidensintervall for μ , antar verdien til σ^2 ukjent

★ Da verdien til σ^2 var kjent tok vi utgangspunkt i

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0, 1)$$

★ Siden σ^2 er ukjent erstatter vi denne med estimator, og får

$$T = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} \quad \begin{array}{l} - S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \\ - \text{hvilken fordeling har } T? \end{array}$$

Teorem

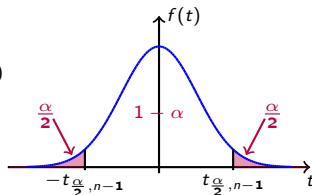
La $X_1, X_2, \dots, X_n$ være et tilfeldig utvalg fra $N(\mu, \sigma^2)$ -populasjonen. Da er

$$V = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \sim \chi^2_{n-1},$$

og dessuten er $\bar{X}$ og V uavhengige stokastiske variabler.

★ Ved å benytte dette får vi at (merk at $\frac{(n-1)S^2}{\sigma^2} = V$)

$$T = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} = \frac{\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2} \cdot \frac{1}{n-1}}} = \frac{Z}{\sqrt{\frac{V}{n-1}}} \sim t_{n-1}$$



Konfidensintervall for σ^2 når også verdien til μ er ukjent

Teorem: $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $N(\mu, \sigma^2)$ -populasjonen. Da er $V = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \sim \chi_{n-1}^2$.

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

★ Situasjon:

- la $X_1, X_2, \dots, X_n$ være tilfeldig utvalg fra $N(\mu, \sigma^2)$ -populasjonen
- verdien til σ^2 er ukjent, ønsker et konfidensintervall for σ^2 , antar verdien til μ ukjent

★ Da verdien til μ var kjent tok vi utgangspunkt i

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi_n^2$$

★ Siden μ er ukjent erstatter vi denne med en estimator, og får

$$V = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2$$

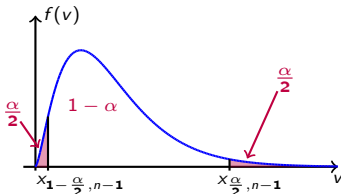
- $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
- hvilken fordeling har V ?

★ Teoremet på forrige side gir at

$$V = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \sim \chi_{n-1}^2$$

★ Fra forrige side har vi også

$$V = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$



Konfidensintervall for p i binomisk fordeling

★ Situasjon:

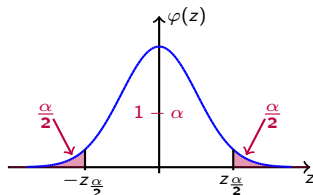
- la X være binomisk fordelt med n forsøk og sannsynlighet for suksess lik p
- verdien til p er ukjent, ønsker et konfidensintervall for p

★ Antar n stor slik at

$$X \approx N(np, np(1-p)) \Rightarrow \frac{X - np}{\sqrt{np(1-p)}} \approx N(0, 1)$$

★ Man må løse andregradulikheter ved videre regning

$$P\left(-z_{\frac{\alpha}{2}} \leq \frac{X - np}{\sqrt{np(1-p)}} \leq z_{\frac{\alpha}{2}}\right) \approx 1 - \alpha$$



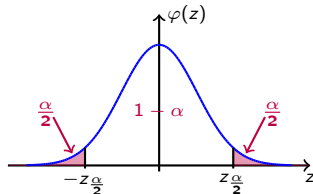
★ Alternativt: Enda en approksimasjon ved å erstatte p i nevner med estimator $\hat{p} = \frac{X}{n}$

- når n er stor kan man vise at

$$\frac{X - np}{\sqrt{n\hat{p}(1-\hat{p})}} \approx N(0, 1)$$

★ Slipper da å løse andregradulikheter

$$P\left(-z_{\frac{\alpha}{2}} \leq \frac{X - np}{\sqrt{n\hat{p}(1-\hat{p})}} \leq z_{\frac{\alpha}{2}}\right) \approx 1 - \alpha$$



To-utvalg: konfidensintervall for $\mu_1 - \mu_2$ når verdiene til σ_1^2 og σ_2^2 er kjente

★ Situasjon:

- la $X_1, X_2, \dots, X_n$ være tilfeldig utvalg fra $N(\mu_1, \sigma_1^2)$ -populasjonen
- la $Y_1, Y_2, \dots, Y_m$ være tilfeldig utvalg fra $N(\mu_2, \sigma_2^2)$ -populasjonen
- anta X_i -ene og Y_i -ene er uavhengig av hverandre
- verdiene til μ_1 og μ_2 er ukjente, ønsker et konfidensintervall for $\mu_1 - \mu_2$
- antar verdiene til σ_1^2 og σ_2^2 kjente

★ Estimatorer for μ_1 og μ_2

$$\hat{\mu}_1 = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{og} \quad \hat{\mu}_2 = \bar{Y} = \frac{1}{m} \sum_{i=1}^m Y_i$$

★ Vi vet at

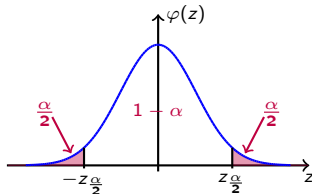
$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n}\right) \quad \text{og} \quad \bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{m}\right), \quad \bar{X} \text{ og } \bar{Y} \text{ er uavhengige}$$

★ Estimator for $\mu_1 - \mu_2$

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

$\Downarrow$

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0, 1)$$



Oppsummering

- ★ Har sett på hvilke pivotaler kan bruke for å lage konfidensintervall
- ★ $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $N(\mu, \sigma^2)$ -populasjon
 - konfidensintervall for μ , σ^2 er kjent
 - konfidensintervall for μ , σ^2 er ukjent
 - konfidensintervall for σ^2 , μ er kjent
 - konfidensintervall for σ^2 , μ er ukjent
- ★ X er binomisk fordelt med parametre n og p
 - konfidensintervall for p
- ★ To-utvalg: $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $N(\mu_1, \sigma_1^2)$ -populasjonen og $Y_1, Y_2, \dots, Y_m$ tilfeldig utvalg fra $N(\mu_2, \sigma_2^2)$ -populasjonen, X_i -ene og Y_i -ene uavhengige
 - konfidensintervall for $\mu_1 - \mu_2$, σ_1^2 og σ_2^2 kjente

