Plan for dagens undervisning

- Regne på eksempler/situasjoner som inkluderer
  - en regresjonsmodell som avviker litt fra vanlig enkel lineær regresjon
  - vurdering av modell: residualplott
  - hvilken fordeling har $\widehat{\alpha}$, $\widehat{\sigma}^2$?
  - konfidensintervall for konstantleddet α

Regresjonsmodell regnet på forrige uke

- Anta $Y_1, Y_2, \ldots, Y_n$ uavhengige og $Y_i \sim N(\alpha + \beta x_i, \sigma^2 x_i^2)$
- Vi fant sannsynlighetsmaksimeringsestimatorene for α, β og σ^2 til å være

$$\widehat{\alpha} = \frac{\sum_{i=1}^n \frac{Y_i}{x_i^2} - \frac{1}{n}\left(\sum_{i=1}^n \frac{Y_i}{x_i}\right)\left(\sum_{i=1}^n \frac{1}{x_i}\right)}{\sum_{i=1}^n \frac{1}{x_i^2} - \frac{1}{n}\left(\sum_{i=1}^n \frac{1}{x_i}\right)^2}$$

$$\widehat{\beta} = \frac{1}{n}\sum_{i=1}^n \frac{Y_i}{x_i} - \widehat{\alpha} \cdot \frac{1}{n}\sum_{i=1}^n \frac{1}{x_i}$$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \frac{(Y_i - \widehat{\alpha} - \widehat{\beta} x_i)^2}{x_i^2}$$

- Egenskaper for $\widehat{\alpha}$

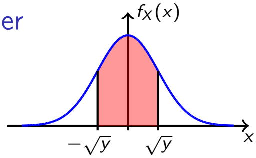
$$E[\widehat{\alpha}] = \alpha, \qquad \mathrm{Var}[\widehat{\alpha}] = \sigma^2 \cdot \frac{\sum_{i=1}^n \left(\frac{1}{x_i} - \frac{1}{n}\sum_{j=1}^n \frac{1}{x_j}\right)^2}{\left(\sum_{i=1}^n \frac{1}{x_i^2} - \frac{1}{n}\left(\sum_{i=1}^n \frac{1}{x_i}\right)^2\right)^2}$$

  - hvilken type fordeling har $\widehat{\alpha}$?
  - hvilken type fordeling har $\widehat{\sigma}^2$?

Lineær funksjon av uavhengige normalfordelte variabler

- Husk teorem: La $X_1, X_2, \ldots, X_n$ være uavhengige og $X_i \sim N(\mu_i, \sigma_i^2), i = 1, 2, \ldots, n$. La $a_1, a_2, \ldots, a_n$ og b være konstanter, og
$$Y = \sum_{i=1}^{n} a_i X_i + b = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n + b.$$
Da er
$$Y \sim N\left(\sum_{i=1}^{n} a_i \mu_i + b, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

Transformasjon av kontinuerlige stokastiske variabler



- Anta $X \sim f_X(x)$ kontinuerlig stokastisk variabel
- Definer ny stokastisk variabel $Y = u(X)$
  - $u(\cdot)$ er gitt matematisk funksjon
- Spørsmål: Hvordan kan vi da finne fordelingen til Y, $f_Y(y)$?
- Finn først et uttrykk for $F_Y(y) = P(Y \leq y)$
- Finn deretter $f_Y(y) = F_Y'(y)$
- Eksempel:
  - $X \sim N(0,1)$, dvs. $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty < x < \infty$
  - $Y = X^2$

$g(y) = \int_0^y \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$
$h(y) = \sqrt{y}$

$g'(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$
$h'(y) = \frac{1}{2\sqrt{y}}$
$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

- Mulige verdier for Y: $[0, \infty)$. Så for $y \geq 0$ har vi

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x) dx$$

$$= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 2 \int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 2g(h(y))$$

$$f_Y(y) = F_Y'(y) = 2g'(h(y)) \cdot h'(y) = 2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{y})^2}{2}} \cdot \frac{1}{2\sqrt{y}}$$

$$= \frac{1}{2^{\frac{1}{2}} \Gamma\left(\frac{1}{2}\right)} y^{\frac{1}{2}-1} e^{-\frac{y}{2}}$$

χ^2-distribution:
$f(x) = \frac{1}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} x^{-\frac{x}{2}}$

Teorem

Anta at $X \sim N(0,1)$. Da er $X^2 \sim \chi^2_1$.

Andre teoremer som kan bevises ved momentgenererende funksjoner

Teorem

La $X_1, X_2, \ldots, X_n$ være uavhengige og $X_i \sim \text{Poisson}(\lambda_i)$. La $Y = \sum_{i=1}^n X_i$. Da er Y poissonfordelt med parameter $\sum_{i=1}^n \lambda_i$.

Teorem

La $X_1, X_2, \ldots, X_n$ være uavhengige og $X_i \sim \chi^2_{\nu_i}$. La $Y = \sum_{i=1}^n X_i$. Da er Y kjikvadratfordelt med $\sum_{i=1}^n \nu_i$ frihetsgrader.

Teorem

La $X_1, X_2, \ldots, X_n$ være uavhengige og eksponensialfordelt med samme parameter λ. La $Y = \sum_{i=1}^n X_i$. Da er Y gammafordelt med parametre $\alpha = n$ og $\beta = \frac{1}{\lambda}$.

Teorem

La $X_1, X_2, \ldots, X_n$ være uavhengige og binomisk fordelt og $X_i \sim \text{bin}(n_i, p)$. La $Y = \sum_{i=1}^n X_i$. Da er Y binomisk fordelt med $n = \sum_{i=1}^n n_i$ forsøk og sannsynlighet for suksess lik p.

Teorem

La $X \sim \text{bin}(n, p)$ og la

$$Z = \frac{X - np}{\sqrt{np(1-p)}}.$$

I grensen når $n \to \infty$ er da $Z \sim N(0, 1)$.

Student t-fordeling

Definisjon (t-fordeling)

La Z og V være uavhengige stokastiske variabler, og la $Z \sim N(0,1)$ og $V \sim \chi^2_\nu$. La så T være en stokastisk variabel definert ved

$$T = \frac{Z}{\sqrt{V/\nu}}.$$

Da sies T å være (Student) t-fordelt med ν frihetsgrader.