Statistical Data Analysis

Dr. Jana de Wiljes

8. Dezember 2021

Universität Potsdam

Asymptotic Properties of the LS-Estimator

Proposition: Consider the setting

$$\mathbf{y}_n = \mathbf{X}_n\beta + \epsilon_n \quad \text{with } \mathbb{E}[\epsilon_n] = \mathbf{0} \quad \text{and } Cov(\epsilon_n) = \sigma^2\mathbf{I}_n \qquad (1)$$

with the following assumption being fulfilled:

$$\boxed{\lim_{n\to\infty} \frac{1}{n}\mathbf{X}_n^\top \mathbf{X}_n = \mathbf{V}} \qquad (2)$$

where $\mathbf{V}$ is positive definite. Then

$X_n \cdot \beta$

$\beta \in \mathbb{R}^{p+1}$

$X_n \in \mathbb{R}^{n \times p+1}$

$X^\top \in \mathbb{R}^{p+1 \times n}$

$\mathbb{R}^{p+1 \times n} \cdot \mathbb{R}^{n \times p+1}$

$V \in \mathbb{R}^{p+1 \times p+1}$

- The LS-estimator $\hat{\beta}_n$ for β as well as the ML- and REML-estimators $\hat{\sigma}_n^2$ for σ^2 are consistent. ($\text{MSE}_\theta(\hat{\theta}) \to 0$ $n \to \infty$)
- The LS-estimator $\hat{\beta}_n$ for β is asymptotically normally distributed:

$$\boxed{\sqrt{n}(\hat{\beta}_n - \beta)} \to \boxed{\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{V}^{-1})} \text{ (in distribution)} \qquad (3)$$

Asymptotic Properties of the LS-Estimator

Proposition: Hence, for sufficiently large n it follows that $\hat{\beta}_n$ is approximately normally distributed with

$$\hat{\beta}_n \to \mathcal{N}(\beta, \sigma^2 \mathbf{V}^{-1}/n) \text{ (almost surely)} \tag{4}$$

Asymptotic Properties of the LS-Estimator

Proposition:

$\hat{\varepsilon}$

- Similar to the error terms, also the residuals have expectation zero.
- In contrast to the error terms, the residuals are not uncorrelated.

Asymptotic Properties of the LS-Estimator

Proposition: Beside the usual assumptions, additionally assume that the error terms are normally distributed. Then the following properties hold:

- The distribution of the squared sum of residuals is given by:

$$\boxed{\frac{\hat{\epsilon}^\top \hat{\epsilon}}{\sigma^2} = (n-p-1)\frac{\hat{\sigma}^2}{\sigma^2}} \tag{5}$$

- The squared sum of residuals $\hat{\epsilon}^\top \hat{\epsilon}$ and the LS-estimator $\hat{\beta}$ are independent.

ε $\quad \underbrace{Y - X\hat{\beta}}_{=\hat{\varepsilon}}$

Prediction

Proposition:

1. The expected prediction error is zero i.e., $\mathbb{E}[\hat{\mathbf{y}}_0 - \mathbf{y}_0] = 0$, i.e., $\cancel{\mathbb{E}[\hat{\mathbf{y}}_0 - \mathbf{y}_0] = 0}$
2. Prediction error covariance matrix is given by:

$$\mathbb{E}[(\hat{\mathbf{y}}_0 - \mathbf{y}_0)(\hat{\mathbf{y}}_0 - \mathbf{y}_0)^\top] = \boxed{\sigma^2(\mathbf{X}_0(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}_0^\top + \mathbf{I}_{T_0})} \quad (6)$$

Predicting

Get: X_0

$n \times \cancel{p+1}$

$T_0 \times p+1$

$y_0 = X_0 \beta + \varepsilon_0$

$\boxed{\hat{y}_0 = X_0 \hat{\beta}}$

$X \in \mathbb{R}^{n \times p+1}$

y_0

$X \neq X_0$

Proof

$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$ (Reminder)

$$\hat{y}_0 = X_0 \hat{\beta} \sim N(X_0 \beta, \hat{\sigma}^2 X_0 (X^T X)^{-1} X_0^T)$$

$X \sim N_p(\mu, \Sigma)$

$AX \sim N_q(A\mu, A \Sigma A^T)$

$A \in \mathbb{R}^{(q \times p)}$

The true value is given by $y_0 = X_0 \beta + \boxed{\varepsilon_0}$

For the prediction error $\hat{y}_0 - y$ one obtains

$$\mathbb{E}[\hat{y}_0 - y_0] = \mathbb{E}[X_0 \hat{\beta} - X_0 \beta - \varepsilon_0]$$

$$= \mathbb{E}[X_0 (\hat{\beta} - \beta) - \varepsilon]$$

$$= X_0 \underbrace{\mathbb{E}[\hat{\beta} - \beta]}_{\underbrace{\left(\underbrace{\mathbb{E}[\hat{\beta}]}_{= \beta} - \beta\right)}_{= 0}} - \underbrace{\mathbb{E}[\varepsilon_0]}_{= 0} = 0$$

Proof

(ii) For the prediction error variance one obtains

$$\mathbb{E}[(\hat{y}_0 - y_0)^T (\hat{y}_0 - y_0)] = \mathbb{E}[(X_0(\hat{\beta} - \beta) - \varepsilon_0)(X_0(\hat{\beta} - \hat{\beta}) - \varepsilon_0)^T]$$

$$= \underline{X_0 \mathbb{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] X_0^T} + \underline{\boxed{\mathbb{E}[\varepsilon_0 \varepsilon_0^T]}}$$

$$- X_0 \mathbb{E}[(\hat{\beta} - \beta)\varepsilon_0^T] - \underbrace{\mathbb{E}[\varepsilon_0 (\hat{\beta} - \beta)^T]}_{\varepsilon_0 \text{ and } (\hat{\beta} - \beta) \text{ are independent}} X_0^T$$

$$= \underline{\sigma^2}(X_0 (X^T X)^{-1} X_0^T + I_{T_0})$$

$\square$

Hypotheses Testing and Confidence Intervals

$$Y = \beta_0 + \underset{\substack{\| \\ 0}}{\beta_1} x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5$$

Primaries

Assumptions: $\varepsilon \sim N_n(0, \sigma^2 I_n)$

$\leadsto y = X\beta + \varepsilon \sim N_N(X\beta, \sigma^2 I_n)$

$$\hat\beta = \underbrace{(X^T X)^{-1} X^T}_{A} y \sim N_p\left((X^T X)^{-1} X^T X \beta,\ \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1}\right)$$

$$= N\left(\beta, \underbrace{\sigma^2 (X^T X)^{-1}}\right)$$

Analogously for the ith compont of the $\hat\beta$-vector we get

$$\hat\beta_i \sim N\left(\beta_i, \sigma^2 (X^T X)^{-1}_{i+1, i+1}\right)$$

or

$$\frac{\hat\beta_i - \beta_i}{\sqrt{\sigma^2 (X^T X)^{-1}_{i+1,i+1}}} \sim N(0,1)$$

$\Rightarrow$ Gauss-test

z-test

$\hat\sigma^2$ $\sim$ t-distributed

Primaries

$R_1 := (r_0, r_1, \dots, r_p)$

$\boxed{R_1 \cdot \beta = r}$ $\rightsquigarrow$ $\beta_i = 0$

$(0 \dots 0\,1\,0 \dots 0)$ 0

i

$R_1 \hat{\beta} \sim N(R_1 \beta, \sigma^2 R_1 (X^T X)^{-1} R_1^T)$

$\Leftrightarrow \dfrac{R_1(\hat{\beta} - \beta)}{\sigma \boxed{\sqrt{R_1 (X^T X)^{-1} R_1^T}}} \sim N(0,1)$

- Problem: σ^2 is unknown $\rightsquigarrow$ solution: estimate σ^2 with

$\hat{\sigma}^2$ – Random variable

$\rightsquigarrow \hat{\sigma}^2 = \dfrac{\hat{\varepsilon}^T \hat{\varepsilon}}{n-p-1} = \dfrac{\boxed{\varepsilon^T M \varepsilon}}{n-p-1}$ where $M := (I_n - X(X^T X)^{-1} X^T)$

$\hookrightarrow$ • $rk(M) = n-p-1$

• symmetric

$\varepsilon \sim N_n(0, \sigma^2 I_n)$ (Assumption)

$\dfrac{\varepsilon}{\sigma} \sim N_n(0, I_n) \Leftrightarrow \dfrac{\varepsilon_i}{\sigma} \sim N(0,1) \quad \forall i = 1, \dots, n$

$\rightsquigarrow \dfrac{\varepsilon_i^2}{\sigma^2} \sim \chi_1^2 \Rightarrow \dfrac{1}{\sigma^2} \sum_{i=1}^{n} \varepsilon_i^2 = \dfrac{\varepsilon^T \varepsilon}{\sigma^2} \sim \boxed{\chi_n^2}$

$\dfrac{\varepsilon^T M \varepsilon}{\sigma^2} \sim \chi^2_{n-p-1}$ $\rightsquigarrow$ $\hat{\sigma}^2 = \dfrac{\sigma^2}{n-p-1} \dfrac{\varepsilon^T M \varepsilon}{\sigma^2}$

Primaries

$\hat{\sigma}^2 \sim \frac{\sigma^2}{n-p-1} \chi^2_{n-p-1}$ as well as $\frac{\hat{\sigma}^2 (n-p-1)}{\sigma^2} \sim \chi^2_{n-p-1}$

Primaries

Primaries

Gamma-distribution

Def: A continuous, non-negative random variable X is called gamma-distributed with parameters $a > 0$ and $b > 0$, abbreviated by the notation $X \sim \mathcal{G}(a, b)$, if it has a density function of the following form

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx), \quad x > 0 \tag{7}$$

with $\boxed{\Gamma(n) = (n-1)!}$. $\Leftarrow$ $\Gamma(1/2)$

$X^2 \sim \chi^2$ - distributed

$X \sim N(0,1)$

Gamma-distribution

Lemma: Let $X \sim \mathcal{G}(a, b)$ be a continuous, non-negative random variable. Then its expectation and variance are given by:

- $\mathbb{E}[X] = \frac{a}{b}$
- $Var(X) = \frac{a}{b^2}$

χ^2-distribution

Def: A continuous, non-negative random variable X with density

$$f(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \exp\Big(-\frac{1}{2}x\Big), \quad x > 0 \tag{8}$$

is called χ^2-distributed with n degrees of freedom, abbreviated by the notation $X \sim \chi^2_n$.

χ^2-distribution

Lemma: Let $X \sim \chi_n^2$ be a continuous, non-negative random variable. Then its expectation and variance are given by:

- $\mathbb{E}[X] = n$
- $Var(X) = 2n$

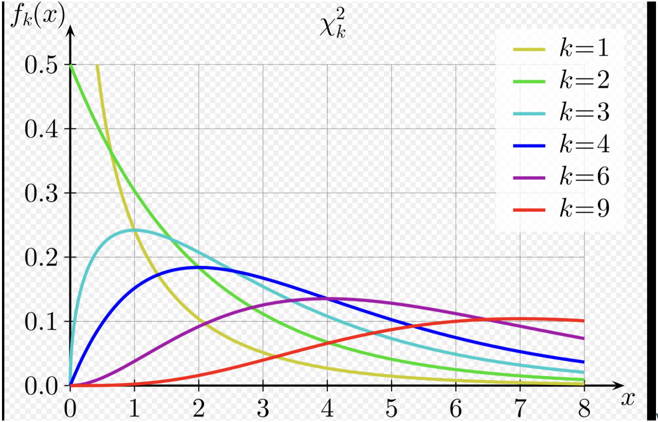
χ^2-distribution

Lemma: Let $X_1, ..., X_n$ be independent and identically standard normally distributed, then

$$Y_n = \sum_{i=1}^{n} X_i^2 \tag{9}$$

is χ^2- distributed with n degrees of freedom.

χ^2-distribution



wikipedia

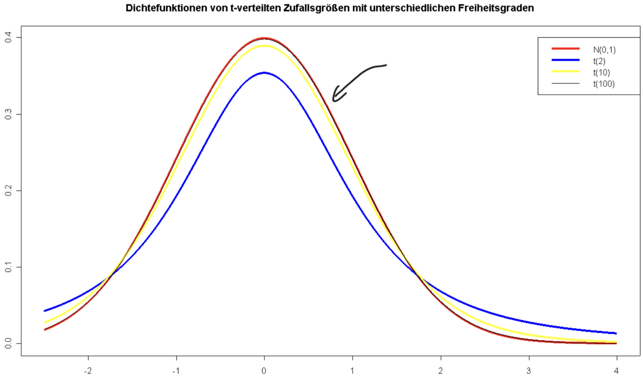
t-distribution

Def: A continuous random variable X with density

$$f(x) = \frac{\Gamma(n+1)/2}{\sqrt{n\pi}\Gamma(n/2)(1+x^2/n)^{(n+1)/2}} \tag{10}$$

is called t-distributed with n degrees of freedom, abbreviated by the notation $t \sim t_n$

t-distribution



wikipedia

t-distribution

Lemma: Let $X \sim t_n$ be a continuous, non-negative random variable. Then its expectation and variance are given by:

- $\mathbb{E}[X] = n \quad n > 1$
- $Var(X) = n/(n-2), \quad n > 2$

The t_1-distribution is also called Cauchy-distribution. If $X_1, ..., X_n$ are iid with $X_i \sim \mathcal{N}(\mu, \sigma^2)$, it follows that

check!

$$\frac{\bar{X} - \mu}{S}\sqrt{n} \sim \boxed{t_{n-1}} \tag{11}$$

with

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \text{ and } \bar{X} = \sum_{i=1}^{n} X_i \tag{12}$$

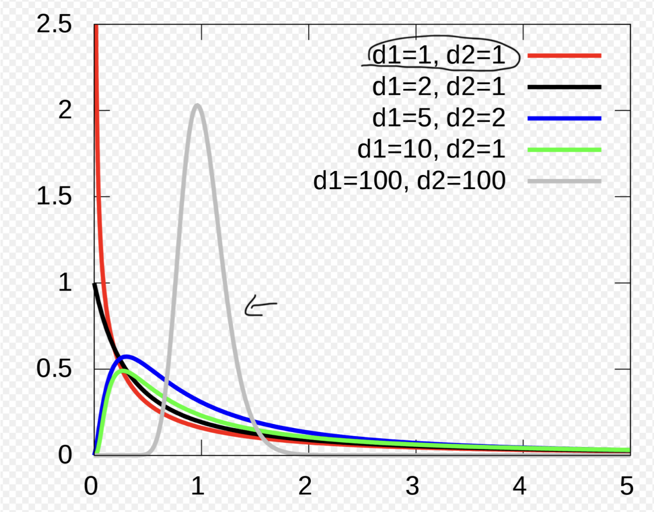
F-distribution

Def: Let X_1 and X_2 be independent random variables χ^2_n- and χ^2_m distributions respectively. Then the random variable

$$\boxed{F = \frac{X_1/n}{X_2/m}} \tag{13}$$

is called F-distributed with n and m degrees of freedom, abbreviated with the notation $\boxed{F \sim F_{n,m}.}$

F-distribution



wikipedia

Primaries

Primaries

Hypotheses Testing and Confidence Intervals

Let $Z \sim \mathcal{N}(0,1)$ and $X \sim \chi^2_k$ be independent random variables. Then therandom variable

$$T := \frac{Z}{\sqrt{\frac{X}{k}}} \tag{14}$$

is t-distributed with k degrees of freedom.

$$\frac{R_1(\hat{\beta}-\beta)}{\hat{\sigma}\sqrt{R_1(X^TX)^{-1}R_1^T}} = \frac{\frac{R_1(\hat{\beta}-\beta)}{\sigma\sqrt{R_1(X^TX)^{-1}R_1^T}}}{\sqrt{\frac{\hat{\sigma}^2(n-p-1)}{\sigma^2(n-p-1)}}} \sim \boxed{t_{n-p-1}}$$

Hypotheses Testing

1) $H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$ —

2 $H_0 = \beta_j - \beta_\ell = 0$ vs $H_1 : \beta_j - \beta_\ell \neq 0$ —

$R_1 \hat{\beta} \overset{H_0}{\sim} N(r, \sigma^2 R_1 (X^T X)^{-1} R_1^T)$

$\boxed{R_1 \beta = r}$

$$T = \frac{R_1 \hat{\beta} - r}{\hat{\sigma} \sqrt{R_1 (X^T X)^{-1} R_1^T}} \sim t_{n-p-1}$$

Hypotheses Testing

Hypotheses Testing

Hypotheses Testing