Statistical Data Analysis

Dr. Jana de Wiljes

7. Dezember 2021

Universität Potsdam

Asymptotic Properties of the LS-Estimator

Proposition: Consider the setting

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

with the following assumption being fulfilled:

$$\lim_{n\to\infty} \frac{1}{n}\mathbf{X}_n^\top \mathbf{X}_n = \mathbf{V} \tag{2}$$

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

- 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:

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

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:

- 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:

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

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

Prediction

Proposition:

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

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

Proof

Proof

Hypotheses Testing and Confidence Intervals

Primaries

Primaries

Primaries

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 $\Gamma(n) = (n-1)!$.

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.

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 $t \sim t_n$

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

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

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

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

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

Primaries

Primaries

Hypotheses Testing and Confidence Intervals

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

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

is t-distributed with k degrees of freedom.

Hypotheses Testing

Hypotheses Testing

Hypotheses Testing

Hypotheses Testing