Statistical Data Analysis

Dr. Jana de Wiljes

14. Dezember 2021

Universität Potsdam

Hypothesis testing for Regression parameters

In general, the following hypotheses can be tested:

1. $H_0 : \mathbf{R}_1\beta = r$ vs. $H_1 : \mathbf{R}_1\beta \neq r$
2. $H_0 : \mathbf{R}_1\beta \geq r$ vs. $H_1 : \mathbf{R}_1\beta < r$
3. $H_0 : \mathbf{R}_1\beta \leq r$ vs. $H_1 : \mathbf{R}_1\beta > r$

Under H_0:

$$\mathbf{R}_1\hat{\beta} \sim \mathcal{N}(r, \sigma^2\mathbf{R}_1(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{R}_1^\top) \tag{1}$$

holds. For unknown σ^2, a reasonable test-statistic is

$$T = \frac{\mathbf{R}_1\hat{\beta} - r}{\hat{\sigma}\sqrt{\mathbf{R}_1(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{R}_1^\top}} \sim t_{n-p-1} \tag{2}$$

The corresponding rejection areas are:

1. $|T| > t_{1-\alpha/2,n-p-1}$
2. $T < t_{1-\alpha,n-p-1}$
3. $T > t_{1-\alpha,n-p-1}$

$(1-\alpha)$-confidence intervals for $\mathbf{R}_1\hat{\beta}$ are:

$$\mathbf{R}_1\hat{\beta} \pm t_{n-p-1,1-\alpha/2}\hat{\sigma}\sqrt{\mathbf{R}_1(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{R}_1^\top} \tag{3}$$

Hypothesis for general parameter identification problems

Setting:

1. Let $(\mathbb{P})_{\theta \in \Theta}$ a family of probability measures on the sample space $(\mathcal{X}, \mathcal{A})$.
2. Find disjunct subsets Θ_1 and Θ_2 of parameter space $\Theta = \Theta_1 \cup \Theta_2$ and $\Theta_1 \cap \Theta_2 = \emptyset$
3.

Hypothesis:

1. Null hypothesis H_0: $\theta \in \Theta_0$
2. alternative hypothesis H_1: $\theta \in \Theta_1$

Example

Example

Example

Example

Neyman-Pearson-Theory

Setting: $\Theta_0 = \{\theta_0\}$, $\Theta_1 = \{\theta_1\}$, $\Theta = \{\Theta_0, \Theta_1\}$

Assumption: The associate probability measures $\mathbb{P}_{\theta_0}$ and $\mathbb{P}_{\theta_1}$ have densities h_0 and h_1 for a measure λ on $(\mathcal{X}, \mathcal{A})$

Def: Let $k \in [0, \infty]$ and $\gamma \in [0, 1]$. A likelihood-quotient-test (LQ-test) is of the form

$$\phi(x) = \begin{cases} 1, & \text{if } \frac{h_1(x)}{h_0(x)} > k \\ 1, & \text{if } \frac{h_1(x)}{h_0(x)} < k \\ \gamma, & \text{if } \frac{h_1(x)}{h_0(x)} = k. \end{cases} \tag{4}$$

Neyman-Pearson Lemma

Lemma: Let ϕ be a LQ-test with $\mathbb{E}_{\theta_0}[\phi(X)] = \alpha$. Then

$$\mathbb{E}_{\theta_1}[\phi(X)] = \sup_{\psi : \mathbb{E}_{\theta_0}[\psi(X)] \leq \alpha} \mathbb{E}_{\theta_1}[\psi(X)] \tag{5}$$

Further for every $\alpha \in (0, \infty)$ it is possible to find $k \in [0, \infty]$ and $\gamma \in [0, 1]$ so that for a predefined Test ϕ

$$\mathbb{E}_{\theta_0}[\phi(X)] = \alpha \tag{6}$$

Proof/Example

Proof/Example

Proof/Example

Proof/Example

Regularization

Ridge Regularization (L_2)

$$\hat{\beta}^{Ridge} = \arg\min_{\beta \mathbb{R}} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda\|\beta\|_2^2 \tag{7}$$

- decreases variance but increases bias (for increasing λ)
- Can improve predictive performance
- $$\hat{\beta}^{Ridge} = (\mathbf{X}^\top \mathbf{X} + \lambda I_p)^{-1}\mathbf{X}^\top \mathbf{y} \tag{8}$$

Lasso Regularization (L_1)

$$\hat{\beta}^{Lasso} = \arg\min_{\beta\mathbb{R}} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda\|\beta\|_1 \tag{9}$$

- LASSO=Least Absolute Shrinkage and Selection Operator
- This penalty allows coefficients to shrink towards exactly zero
- LASSO usually leads to sparse models, that are easier to interpret