Universität Bielefeld
Tanja Pasurek
Ehsan Abedi WiSe 2024/25

Foundations of Statistics

Homework 12

Topic I: Confidence intervals for ML-estimators

Exercise 1. For the dataset of Exercise 3 in HW 11 calculate a 99% confidence interval for the parameter vector (μ, σ^2) with the observed Fisher information matrix. To do this, create in `R` a function for the log-likelihood and follow pp. 9 and 15 of Ch. 4.4.

Topic II: Delta method

Exercise 2. Let $X_1, X_2, ..., X_n \overset{\text{i.i.d.}}{\sim} \text{Par}(\alpha)$ be a random sample from the continuous Pareto distribution with density

$$f(x) := \begin{cases} \dfrac{\alpha x_{\min}^{\alpha}}{x^{\alpha+1}}, & \text{if } x \geq x_{\min}, \\ 0, & \text{if } x < x_{\min}, \end{cases}$$

(cf. Ch. 1.6), where $x_{\min} > 0$ is a lower bound for possible values of the X_i's. The parameter $\alpha > 0$ is called the *tail index* or the *Pareto index*. Here we assume that $x_{\min}$ is given, while α is unknown. The goal of this exercise is to find estimators of α and study their asymptotic distributions as the number of samples n increases.

a Find the maximum likelihood estimator $\widehat{\alpha}_{ML}$ of α.

b Define $Y_i := \log \frac{X_i}{x_{\min}}$ and $\overline{Y}_n := \frac{1}{n}\sum_{i=1}^{n} Y_i$ and express $\widehat{\alpha}_{ML}$ in terms of $\overline{Y}_n$.

c Find the distribution of Y_i directly (*Hint*: write the CDF of Y_i and show that it follows an exponential distribution). Then, find $\mathbb{E}[Y_i]$ and $\text{Var}(Y_i)$. Having this result, what is your interpretation now of $\widehat{\alpha}_{ML}$?

d Find the asymptotic distribution of $\overline{Y}_n$ using the CLT and the results of task (c).

e Use tasks (b) and (d) to derive the asymptotic distribution of $\widehat{\alpha}_{ML}$:

$$\sqrt{n}(\widehat{\alpha}_{ML} - \alpha) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \alpha^2). \tag{1}$$

Here you can apply the so-called *delta method*, for its theoretical background see Ch. 4.5.

f Define $\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$. Apply the method of moments to obtain an estimator $\widehat{\alpha}_{MM}$ of α for the same stochastic model in terms of $\overline{X}_n$. Explain why this method requires that $\alpha > 1$ and compare $\widehat{\alpha}_{MM}$ and $\widehat{\alpha}_{ML}$.

g Find the asymptotic distribution of $\overline{X}_n$ using the CLT. (For this, you would need to compute $\mathrm{Var}(X_i)$ whose existence requires a stronger assumption, namely $\alpha > 2$.)

h With the delta method as before, use parts (f) and (g) to derive the asymptotic distribution of $\widehat{\alpha}_{MM}$:

$$\sqrt{n}(\widehat{\alpha}_{MM} - \alpha) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \frac{\alpha(\alpha-1)^2}{\alpha - 2}\right). \tag{2}$$

i Finally, under the assumption $\alpha > 2$, compare the asymptotic variance of the two estimators found in (1) and (2). Which one is smaller?

Topic III: Student's t-distribution

Exercise 3. Let $Z_1, ..., Z_n$ be an *i.i.d.* random sample from an $\mathcal{N}(0,1)$ distribution. Define $X_i := \mu + \sigma Z_i$ for some $\mu \in \mathbb{R}, \sigma > 0$ and for $i = 1, ..., n$. We know that $X_1, ..., X_n$ is a random sample from a $\mathcal{N}(\mu, \sigma^2)$ distribution.

Let $\overline{Z}_n$, $\overline{X}_n$ denote the sample averages and S_Z, S_X denote the sample standard deviations of the Z_is and X_is, respectively.

a Express $\overline{X}_n$ and S_X in terms of $\overline{Z}_n$, S_Z, μ, and σ.

b Verify that

$$\frac{\overline{X}_n - \mu}{S_X/\sqrt{n}} = \frac{\overline{Z}_n}{S_Z/\sqrt{n}},$$

and explain why this shows that the distribution of the *studentized* mean does not depend on μ and σ.

Topic IV: Bootstrap confidence intervals

Exercise 4. With a parametric bootstrap construct a 95% confidence interval for the parameter of a Poisson distribution using the sample below with size $n = 20$

$$5, 1, 2, 3, 1, 2, 1, 1, 2, 2, 3, 2, 1, 1, 4, 4, 3, 2, 4, 4.$$

Exercise 5. The dataset `acme` in `R` package `boot` has the following description:

"The excess returns for the Acme Cleveland Corporation along with those for all stocks listed on the New York and American Stock Exchanges were recorded over a five year period. These excess returns are relative to the return on a riskless investment such as U.S. Treasury bills."

The vector `acme$acme` has 60 values for the excess return for the Acme Cleveland Corporation. The median is negative. Construct a 95% bootstrap confidence interval for the median. Does it contain zero?

Exercise 6. Suppose the following sample of observations is available on the electricity consumption (in kilowatts), denoted as "Power", and day temperature (in degrees Celsius), denoted as "Temp", in a house over a period of 20 days:

Day	1	2	3	4	5	6	7	8	9	10
Power	198	184	245	223	263	246	206	216	191	237
Temp	30	25	37	38	27	36	33	29	26	34

Day	11	12	13	14	15	16	17	18	19	20
Power	208	244	221	209	256	276	226	208	198	207
Temp	24	35	37	28	37	36	33	31	26	34

The variables "Power" and "Temp" are correlated. Construct a 95% bootstrap confidence interval for the Pearson correlation coefficient between "Power" and "Temp". For that, apply the `boot` package and the function `boot.ci` with both options `type='perc'` and `type='norm'` for the percentile and normal intervals, respectively. (For example, use $B = 5000$ bootstrap samples and set a seed value 1234.)