Universität Bielefeld
Tanja Pasurek
Ehsan Abedi

WiSe 2023/24

Foundations of Statistics

Homework 10

Topic I: Point estimation (Chapter 3)

(Solve any 4 exercises of your choice from the 5 in this topic.)

Exercise 1. Consider the linear regression model in Ch. 3.6.

(a) Prove that the least square estimators $\hat{\alpha}$ and $\hat{\beta}$ (given there by formula (7)) are unbiased.

Hint: First, show that $\hat{\beta}$ can be equivalently rewritten in the following form

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \overline{x}_n)Y_i}{s_{xx}}, \quad \text{where} \quad s_{xx} := \sum_{i=1}^{n}(x_i - \overline{x}_n)^2.$$

Thereafter, represent $\hat{\alpha} = \overline{Y}_n - \hat{\beta} \cdot \overline{x}_n$ as a linear function of $Y_1, ..., Y_n$.

(b) Assuming that $\varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, prove that $\hat{\alpha}$ and $\hat{\beta}$ are both normally distributed with

$$\text{Var}(\hat{\beta}) = \frac{\sigma^2}{s_{xx}}, \quad \text{Var}(\hat{\alpha}) = \sigma^2 \left(\frac{1}{n} + \frac{(\overline{x}_n)^2}{s_{xx}} \right).$$

(c) Under the same conditions, prove that

$$\text{Cov}(\hat{\alpha}, \hat{\beta}) = -\sigma^2 \frac{\overline{x}_n}{s_{xx}}.$$

Hint: Use the following property of covariances

$$\text{Cov}\left(\sum_{i=1}^{n} a_i X_i, \ \sum_{j=1}^{m} b_j Y_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \text{Cov}(X_i, Y_j)$$

that holds for any random variables X_i, Y_j (with finite $\mathbb{E}(X_i^2)$, $\mathbb{E}(Y_j^2)$), any constants $a_i, b_j \in \mathbb{R}$ and $1 \le i \le n$, $1 \le j \le m$.

Exercise 2. The buit-in-dataset `trees` in `R` provides measurement of the girth, height and volume of timber in 31 felled black cherry trees.

(a) Draw a scatterplot of the measurements in `R`.

(b) For `x=trees$Girth` and `y=trees$Volume` the command `fit<-lm(y~x)` is read as `y` is modeled by `x` and prints out the estimates for the coefficients of the regression line. Use the command `summary()` to summarize regression model. Plot the regression line into the scatterplot of the measurements.

(c) A tree has a girth size of 16 inches. Predict its volume using your regression model using the command `predict()` with `interval = "prediction"` and include your prediction point in the plot. Compare the result with direct computation from the coefficients you obtained in task **(b)**.

Exercise 3 is aimed to illustrate the theoretical material of Ch. 3.9.

Consider a Bernoulli distribution $\mathrm{Ber}(\pi)$ with parameter $\pi \in (0,1)$. Its PMF can be represented by the following formula

$$f_\pi(x) = \mathbb{P}(X = x) = \begin{cases} \pi^x(1-\pi)^{1-x}, & x \in \{0,1\}, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Let $x_1, ..., x_n$ be a realization of a random sample $X_1, ..., X_n \overset{\text{iid}}{\sim} \mathrm{Ber}(\pi)$. Calculate the observed Fisher information for this dataset.

(b) Calculate the expected Fisher information for $X_1, ..., X_n \overset{\text{iid}}{\sim} \mathrm{Ber}(\pi)$.

(c) Show that the estimator $\hat{\pi}_n = \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ attains the explicit Cramér–Rao bound.

Exercise 4 is aimed to illustrate the theoretical material of Ch. 3.10.

For some $\lambda > 0$, suppose $X_1, ..., X_n$ is a random sample from the density

$$f(x) = \frac{\lambda}{2\sqrt{x}} e^{-\lambda\sqrt{x}} \quad \text{for } x > 0.$$

(a) Compute the ML-estimator $\widehat{\lambda}_n$ and the Fisher information $\mathcal{I}_n(\lambda)$.

(b) Use the Fisher information to approximate $\mathrm{Var}(\widehat{\lambda}_n)$ as $n \to \infty$.

(c) For a sample of size $n = 30$ and $\lambda = 1/2$, use simulation to get a better approimation of the true variance of $\widehat{\lambda}_n$, and compare this to the approximation using the Fisher information.

Hint: Recall the so-called inverse transform method for simulating random variables, described in Ch. 1.8, on pages 14–16. To simulate samples from probability density function f, we first calculate the CDF

$$F(x) := \int_{-\infty}^{x} f(y)\,\mathrm{d}y = 1 - \exp(-\lambda\sqrt{x}), \quad x > 0,$$

and then find its inverse

$$G(y) := F^{-1}(y) := \left[\tfrac{1}{\lambda}\log(1-y)\right]^2, \quad y \in (0,1).$$

Then we know that the following transformation

$$X_i := G(Y_i) = \left[\tfrac{1}{\lambda}\log(1-Y_i)\right]^2$$

of the random variable $Y_i \sim \mathrm{Unif}(0,1)$ has the desired PDF f.

Exercise 5. Suppose that a random sample X_i, $i \geq 1$, has a normal distribution $\mathcal{N}(0, \sigma^2)$ with mean 0 and unknown variance $\sigma^2 > 0$.

(a) Find the Fisher information $\mathcal{I}(\sigma)$ for a single variable X_i considering the standard deviation $\sigma > 0$ as the unknown parameter.

(b) Find the ML-estimator $\hat{\sigma}_n$ and describe approximately its sampling distribution as $n \to \infty$.

(c) Find the Fisher information $\widetilde{\mathcal{I}}(\theta)$ considering the variance $\theta := \sigma^2$ as the unknown parameter.

(d) Find the ML-estimator $\hat{\theta}_n$ directly and applying the invariance principle (see page 30 of Ch. 3.4). Describe the sampling distribution of $\hat{\theta}_n$ as $n \to \infty$. Check that $\hat{\theta}_n$ is unbiased, whereas $\hat{\sigma}_n$ is biased (by using Jensen's inequality, see the Addendum to HW 9 and Ex. 3b there).

(e) (optional*) Suppose that X is a random variable for which the PDF or the PMF is $f_\phi(x)$, where the value of the parameter $\phi \in \mathbb{R}$. Let $\mathcal{I}(\phi)$ denote the Fisher information in X. Suppose now that the parameter ϕ is replaced by a new parameter θ, where $\phi = g(\theta)$, and $g : \mathbb{R} \to \mathbb{R}$

is a differentiable function. Let $\widetilde{\mathcal{I}}(\theta)$ denote the Fisher information in X with respect to the parameter θ. Show that

$$\widetilde{\mathcal{I}}(\theta) = \left[g'(\theta)\right]^2 \mathcal{I}\left[g(\theta)\right].$$

Apply this general result to **(a)** and **(c)**.

Topic II: Confidence intervals for proportions (Chapters 4.1-4.2)

Exercise 6. Suppose we want to make a 95% confidence interval for the probability of getting heads with a Dutch 1 Euro coin, and it should be at most 0.01 wide. To determine the required sample size, we note that the probability of getting heads is about 0.5. Furthermore, if X has a $\mathrm{Bin}(n, p)$ distribution, with n large and $p \approx 0.5$, then

$$\frac{X - np}{\sqrt{n/4}} \text{ is approximately normal.}$$

(a) Use this statement to derive that the width of the 95% CI for p is approximately $z_{0.025}/\sqrt{n}$.

Use this width to determine how large n should be.

(b) The coin is thrown the number of times just computed, resulting in 19 477 times heads. Construct the 95% CI.

Exercise 7. Let's do more simulations to find coverage probabilities for a binomial proportion. Given a random sample $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{Ber}(p)$, we know that $\sum_{i=1}^n X_i \sim \mathrm{Bin}(n, p)$.

(a) As was shown in Ch. 4.2, the approximate 95% CI is

$$\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n},$$

where $\hat{p} = \overline{X}_n$. However, it is known that for p near 0 and 1 the true confidence level might be too low. Find the "true" coverage probability for the case $p = 0.05$ and $n = 60$ using a simulation in `R`.

(b) The **Wilson confidence interval** was published in 1927.

To derive the formula for $(1-\alpha)\,100\%$ CI, let $z_{1-\alpha/2}$ be the $(1-\alpha/2)$-quantile of $N(0,1)$. Then

$$\mathbb{P}\left(\hat{p} - z_{1-\alpha/2}\sqrt{p(1-p)/n} \ < p \ < \hat{p} + z_{1-\alpha/2}\sqrt{p(1-p)/n}\right) \approx 1-\alpha.$$

Now, we do not plug in $\hat{p}$ for p; insted we solve both inequalities for p. This involves solving a quadratic equation. The resulting formula gives the confidence interval

$$\left(\frac{\hat{p} + \frac{z^2_{1-\alpha/2}}{2n} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2_{1-\alpha/2}}{4n^2}}}{1 + \frac{z^2_{1-\alpha/2}}{n}},\ \frac{\hat{p} + \frac{z^2_{1-\alpha/2}}{2n} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2_{1-\alpha/2}}{4n^2}}}{1 + \frac{z^2_{1-\alpha/2}}{n}}\right).$$

Run a simulation in `R` to find the coverage probability for this improved 95% CI when $p = 0.05$ and $n = 60$.

Have a wonderful holiday season!