Exercises for the lecture on

Prof. Dr. Vladimir Spokoiny
Vladimir.Spokoiny@wias-berlin.de
Dr. Thorsten Dickhaus
Thorsten.Dickhaus@wias-berlin.de
www.wias-berlin.de/people/dickhaus/

Statistical Methods

Humboldt-University Berlin
Department of Mathematics
Winter term 2013 / 2014

Sheet 5

Solution are due on Monday, November 18th, 2013, 3:15pm.
Every completely and correctly solved exercise gives 4 points.

Exercises

17. **Confidence intervals for a binomial success probability.**
It is known that the cumulative distribution function of a beta distribution with parameters $a, b \in \mathbb{N}$ can be written in the following form.

$$F_{\mathrm{Beta}(a,b)}(x) = \sum_{j=a}^{a+b-1} \frac{(a+b-1)!}{j!(a+b-1-j)!} x^j (1-x)^{a+b-1-j},\ a, b \in \mathbb{N},\ x \in [0,1].$$

(a) Show the following duality: If X is binomially distributed with parameters n and p, then it holds

$$F_X(k) = F_{\mathrm{Beta}(n-k,k+1)}(1-p),\ k \in \{0, \ldots, n\}.$$

(b) We utilize the result of (a) for deriving an exact $(1-\alpha)$-confidence interval $[p_{\mathrm{lower}}, p_{\mathrm{upper}}]$ for the unknown success probability p in the case that $X = k$ is observed (Clopper-Pearson method). To this end, show that

$$\mathbb{P}_{p_{\mathrm{upper}}}(X \leq k) = \alpha/2 \iff p_{\mathrm{upper}} = F^{-1}_{\mathrm{Beta}(k+1,n-k)}(1-\alpha/2),$$
$$\mathbb{P}_{p_{\mathrm{lower}}}(X \geq k) = \alpha/2 \iff p_{\mathrm{lower}} = F^{-1}_{\mathrm{Beta}(k,n-k+1)}(\alpha/2).$$

(c) Construct an approximate $(1-\alpha)$-confidence interval for p based on the data $X = k$ by making use of the central limit theorem in the version of de Moivre and Laplace.

(d) For different parameter tuples (n, p), carry out a small simulation study with statistics software and assess the realized relative coverage frequencies of the concurring methods from (b) and (c).

18. **Descriptive statistics.** The body heights of 20 male pupils were measured and the following values (rounded to full cm) were obtained.

149	147	158	165	153	153	168	158	163	159
177	175	163	170	162	162	170	153	147	157

(a) Draw the empirical cumulative distribution function of this data sample (by hand!).

(b) Draw a histogram of the data sample using the following bins: $(145, 150], (150, 155], \ldots, (175, 180]$.

(c) Draw a floating histogram (i. e., the graph of a kernel density estimator with rectangular kernel) of the data sample. Choose the bandwidth h in accordance with (b).

(d) Compute the empirical skewness of the variable "body height of male pupils" by making use of the given data set and interpret the result by means of the graphics from (a) - (c).

19. **Programming exercise: Descriptive statistics.**
The dataset concerning the "Old Faithful" geyser, which has already been used in Chapter 2, is available in `R` (name of the dataset: `faithful`).

(a) Describe and visualize the univariate empirical distributions of the two variables "duration of eruption" and "waiting time until eruption" by means of appropriate parameters and graphics.

(b) Group both of the two variables into appropriate bins and draw pie charts.

(c) Compute a contingency table for the analysis of the grouped data.

(d) Draw and interpret a mosaic plot and an association plot for the analysis of interrelations between the grouped variables.

20. **Multiple Select.** Which of the following statements are true and which are false?
Please give reasons for your respective decisions (one short sentence each is sufficient).

1. If the real-valued random variable X possesses a finite first moment and the distribution of X is symmetric around zero, then the expected value of X and the median of X coincide.
2. Consistency of a kernel density estimator implies that the bandwidth h is a function of the sample size n.
3. If the sample size for constructing a confidence interval for the mean of a univariate Gaussian distribution (the variance of which is unknown) is doubled, then on average the length of the confidence interval gets halved.
4. The correspondence theorem (Theorem 1.39) implies that a false rejection of a point null hypothesis $\{\vartheta_0\}$ by a level-α test can only take place if ϑ_0 is not covered by the corresponding $(1-\alpha)$-confidence region.