Probability and Statistics

Exercise sheet 13

Exercise 13.1 Consider the null hypothesis $X \sim f(x)dx$ and the alternative $X \sim f(x-1)dx$ for the following cases:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$
$$f(x) = \frac{1}{\pi(1+x^2)}.$$

Compute the form of the rejection of the likelihood area ratio test (Neyman-Pearson Lemma). Comment the difference.

Exercise 13.2

Let $(X_i)_{i=1}^n$ be an i.i.d F-distributed sequence. Let F be absolutely continuous. The Sign test is a test where the null hypothesis is that the median of X is m, i.e.

$$F^{-1}(m) = \frac{1}{2}.$$

Use the Duality Theorem (cf. Theorem 6.4 LN, or Probability overview) to construct the test with significance level $\alpha = 0.05$.

Exercise 13.3 We want to investigate the effect of an outlier on confindence intervals. Let $X_1, \ldots, X_n$ i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$ with unknown σ.

(a) Give the two-sided confidence interval for the unknown parameter μ with level α.

(b) How does the confidence interval behaves for $x_1 \to \infty$ and fixed $x_2, \ldots, x_n$?
Hint: For every $c \in \mathbb{R}$ it holds that $\sum_{i=1}^n (x_i - c)^2 = \sum_{i=1}^n (x_i - \overline{x})^2 + n(c - \overline{x})^2$.

Exercise 13.4 In a study on the reliability of ball-bearing, two samples of 10 pieces each of two different types of ball-bearings were tested. The number of rotation (in millions) were

Typ I	3.03	5.53	5.60	9.30	9.92	12.51	12.95	15.21	16.04	16.84
Typ II	3.19	4.26	4.47	4.53	4.67	4.69	12.78	6.79	9.37	12.75

Before the realisation of this test, it was not clear which type was more reliable.

(a) Are we dealing with a paired sample ?

(b) Build a t-Test for the null-hypothesis “the expected number of rotations until break-down is the same for the two types of ball-bearing” with level 0.05%.

Exercise 13.5 Let $(X_i)_{i=1}^{2n+1}$ a sequence of i.i.d normal random variables with mean μ and variance σ^2 unknown. We take two different estimators for μ:

$$T_{2n+1}^{(1)} = \frac{1}{2n+1} \sum_{i=1}^{2n+1} X_i,$$
$$T_{2n+1}^{(2)} = X_{(n+1)},$$

where $X_{(1)} < X_{(2)} < ... < X_{(2n+1)}$ are the ordered results.

(a) With the help of the Central Limit Theorem find sequences $c_n^{(1)}$ and $c_n^{(2)}$ so that

$$\mathbb{P}\left(|T_{2n+1}^{(i)} - \mu| \le c_n^{(i)}\right) \to 0.95.$$

Hint: You may use as well the result of Example 4.6 from the lecture notes.

(b) Find $q \in \mathbb{R}^+$ so that

$$\frac{c_{nq}^2}{c_n^1} \to 1,$$

how can we interpret, in words, q?.

Exercise 13.6 Least-squares line.

(a) Let $(x_1, y_1), \ldots, (x_n, y_n)$ be a set of n points of $\mathbb{R}^2$ and the x_i's are not all the same. Show that the straight line defined by the equation $y(x) = \hat{\beta}_0 + \hat{\beta}_1 x$ that minimizes the sum of the squares of the vertical deviations of all the points from the line has the following slope and intercept, i.e. $(\hat{\beta}_0, \hat{\beta}_1)$ minimizes

$$I(\beta_0, \beta_1) := \sum_{i=1}^{n} (\beta_0 + \beta_1 x_i - y_i)^2$$

over all choices of $(\beta_0, \beta_1) \in \mathbb{R}^2$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2},$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$.

The minimizing line is called the *least-squares line*. Remark that the least-squares line passes through the point $(\bar{x}, \bar{y})$.

(b) Fit a straight line of the form $y = \beta_0 + \beta_1 x$ to these values by the method of least squares (with your calculator or Excel).

Table 1: Data for Ex 1.(b)

i	x_i	y_i
1	0.5	40
2	1.0	41
3	1.5	43
4	2.0	42
5	2.5	44
6	3.0	42
7	3.5	43
8	4.0	42

Exercise 13.7 Fitting a polynomial by Methode of Least Squares Suppose now that instead of simply fitting a straight line to n plotted points, we wish to fit a polynomial of degree k $(k \ge 2)$. such a polynomial will have the following form:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_k x^k.$$

The method of least squares specifies that the constants $\beta_0, \cdots, \beta_k$ should be chosen that the sum

$$Q(\beta_0, \cdots, \beta_k) = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i + \cdots + \beta_k x_i^k)]^2$$

of the squares of the vertical deviations of the points from the curve is a minimum.

(a) Which equation system should a minimizer $\hat{\beta}_0, \cdots, \hat{\beta}_k$ satisfy?

(b) Fit a parabola (polynomial of degree 2) to the 10 points given in the table.

Table 2: Data for Ex-2.(b)

i	x_i	y_i
1	1.9	0.7
2	0.8	-1.0
3	1.1	-0.2
4	0.1	-1.2
5	-0.1	-0.1
6	4.4	3.4
7	4.6	0.0
8	1.6	0.8
9	5.5	3.7
10	3.4	2.0

Exercise 13.8 GAUSS-MARKOV THEOREM We want to study linear regression models. We do m experiments with explanatory variables $(x_i)_{i=1}^m \subseteq \mathbb{R}^n$ and with a scalar dependent variable $(y_i)_{i=1}^n \subseteq \mathbb{R}$. We suppose that for all i, the underlying model is given by

$$y_i = \beta \cdot x_i + \epsilon_i \quad \beta \in \mathbb{R}^n \tag{1}$$

where (ϵ_i) is a i.i.d sequence such that $\mathbb{E}(\epsilon_i) = 0$ and $\mathrm{Var}(\epsilon_i) = \sigma^2$. We want to estimate β.

We say that $\tilde{\beta}$ is an unbiased estimator of β if

$$\mathbb{E}\left(\tilde{\beta}\right) = \beta.$$

Additionally we say that $\tilde{\beta}$ is linear if there exists a matrix, D, only depending on X such that $\tilde{\beta} = DY$ .We will also say that a matrix $A \lesssim B$ if $B - A$ is a positive semidefinite matrix.

(a) Show that (1) is equivalent to

$$Y = X\beta + \epsilon, \tag{2}$$

where $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$, $X = \begin{pmatrix} x_1^t \\ \vdots \\ x_m^t \end{pmatrix}$ and $\epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \end{pmatrix}$.

(b) Show that the normal linear regression model (example 3.1 of the Skript) is a linear unbiased estimator. We will call its associated matrix K.

(c) Compute the covariance matrix of $\bar{\beta}$, the estimator of the normal linear regression model. **Hint:** Remember that if $Z \in \mathbb{R}^n$ is a random variable and C is a matrix then $V(CZ) = CZC^t$, where $\mathrm{Var}(\cdot)$ is the covariance matrix.

(d) Show that if $\tilde{\beta} = (K + C)Y$ is an unbiased estimator, then $CX = 0$.

(e) Show that the covariance matrix of $\tilde{\beta}$ is such that

$$\mathrm{Var}(\tilde{\beta}) \gtrsim \mathrm{Var}(\bar{\beta}).$$