# Statistiek (WISB361)

**Final exam**

June 29, 2015

*Schrijf uw naam op elk in te leveren vel. Schrijf ook uw studentnummer op blad 1.*
The maximum number of points is 100.
Points distribution: 23–20–20–20–17

1. Consider the sample $\mathbb{X} = \{X_i\}_{i=1}^n$ of i.i.d. random variables with probability density function:

$$f(x|\theta) = \begin{cases} \dfrac{2\theta^2}{x^3} & \text{for } x \geq \theta, \\ 0 & \text{otherwise} \end{cases}$$

for $\theta > 0$.

   (a) [7pt] Determine the maximum likelihood estimator $\hat{\theta}_{MLE}$ and $\widehat{\theta^2}_{MLE}$ of $\theta$ and $\theta^2$ respectively.

   (b) [4pt] Determine the distribution of $\hat{\theta}_{MLE}$.

   (c) [3pt] Is $\hat{\theta}_{MLE}$ a biased estimator? Is $\hat{\theta}_{MLE}$ *asymptotically* unbiased?

   (d) [3pt] Find the method of moment estimator $\hat{\theta}_{MoM}$ of $\theta$.

   (e) [3pt] Find the variance of $\hat{\theta}_{MoM}$. Is it finite?

   (f) [3pt] Compare the *mean squared errors* (MSE) of $\hat{\theta}_{MLE}$ and $\hat{\theta}_{MoM}$. Which estimator is the most efficient?

2. Coffee abuse is often considered related to an increase of heart rate (i.e. number of poundings of the heart per unit of time). In order to support this claim, an experiment was planned on a sample of 10 subjects. The hearth rate was measured twice for each subject: the first time *at rest*, the second after having drank a cup of coffee. The measurements *at rest* are:

$$\mathbf{x} = \{x_1, \ldots, x_{10}\} = \{74, 68, 67, 71, 69, 65, 70, 70, 66, 67\}$$

and after a cup of coffe:

$$\mathbf{y} = \{y_1, \ldots, y_{10}\} = \{71, 72, 69, 66, 73, 77, 69, 68, 71, 78\}$$

We can assume that the random variables $Z_i = Y_i - X_i$, denoting the difference between the heart rate after the cup of coffee and the heart rate at rest, are i.i.d. normal random variables.

   (a) [12pt] Test the hypothesis that the coffee increase the heart rate at $\alpha = 0.05$ level of significance.

   (b) [5pt] Calculate the $p$-value of the test.

   (c) [3pt] In case the normality assumption does not hold, how you could test the hypothesis of point (a)? (It is enough to explain which test is the most appropriate, without performing the analysis).

3. Let $X_1$ and $X_2$ be i.i.d. random variables such that $X_i \sim \text{Unif}[\theta, \theta+1]$, for $i = 1, 2$, where $\text{Unif}[\theta, \theta+1]$ denotes the uniform distribution in the interval $[\theta, \theta+1]$. In order to test:

$$\begin{cases} H_0: & \theta = 0, \\ H_1: & \theta > 0 \end{cases}$$

we have two competing tests, with the following rejection regions:

$$\textbf{TEST1}: \text{Reject } H_0 \quad \text{if } X_1 > 0.95,$$
$$\textbf{TEST2}: \text{Reject } H_0 \quad \text{if } X_1 + X_2 > C,$$

with $C \in \mathbb{R}$.

(a) [2pt] Find the significance level $\alpha$ of **TEST1**.

(b) [4pt] Find the value of $C$ so that **TEST2** has the same significance level $\alpha$ of **TEST1**.

(c) [4pt] Calculate the power function of each test.

(d) [6pt] Is **TEST2** more powerful than **TEST1**?

(e) [4pt] Show how to get a test that has the same significance level but more powerful than **TEST2**.

4. Let the independent normal random variables $Y_1, Y_2, \dots, Y_n$ be such that $Y_i \sim N(\mu, \alpha^2 x_i^2)$, for $i = 1, \dots, n$, where the given constants $x_i$ are not all equal and no one of which is zero.

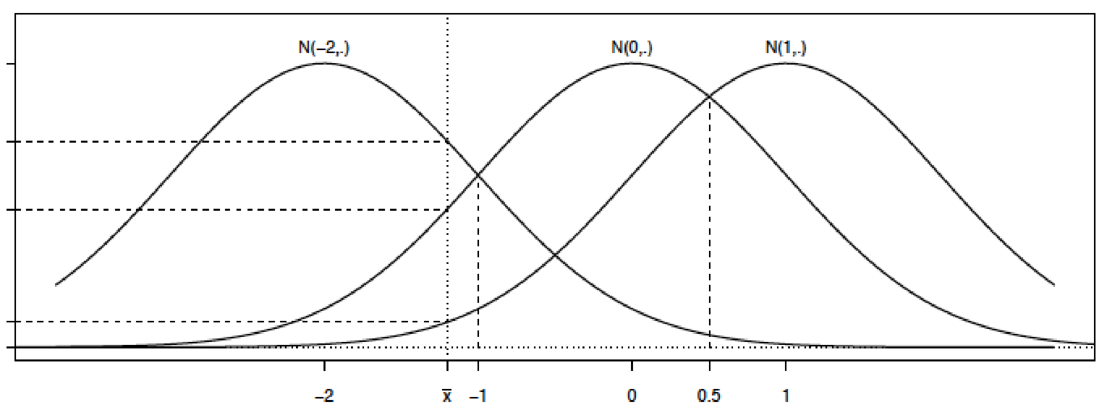
   (a) [13pt] Derive the least squares estimators of $\mu$ and $\alpha^2$, after you have properly rescaled the random variables $Y_i$.

   (b) [4pt] Which is the distribution of $(n-1)\hat{\alpha}^2/\alpha^2$?

   (c) [3pt] Discuss the test of hypotheses:
$$\begin{cases} H_0: & \alpha = 1, \\ H_1: & \alpha \neq 1. \end{cases}$$

5. Consider the sample $\mathbb{X} = \{X_i\}_{i=1}^n$ of i.i.d. random variables such that $X_i \sim N(\theta, \sigma^2)$ with $\sigma^2$ **known** and $\theta \in \Omega$, where the parameter space is $\Omega = \{-2, 0, 1\}$.

   (a) [2pt] Show that $\bar{X} = 1/n \sum_{i=1}^n X_i$ is a sufficient statistics for $\theta$ and that the likelihood $lik(\theta)$ can be factorized in $lik(\theta) = h(x) g_\theta(\bar{x})$, where $x$ is a realization of $\mathbb{X}$, $\bar{x}$ is a realization of $\bar{X}$ and
$$h(x) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \bar{x})^2\right\}, \qquad g_\theta(\bar{x}) = \exp\left\{-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2\right\}$$




In the figure above the functions $g_\theta(y)$ are plotted for the three possible values of the parameter $\theta$.

   (b) [8pt] Find a maximum likelihood estimator $\hat{\theta}_{MLE}$ of $\theta$.

   (c) [4pt] Find the probability mass function of $\hat{\theta}_{MLE}$.

   (d) [3pt] Is $\hat{\theta}_{MLE}$ a biased estimator?