Statistiek (WISB361)

Retake exam

July 20, 2015

Schrijf uw naam op elk in te leveren vel. Schrijf ook uw studentnummer op blad 1.
The exam is open–book and the use of the calculator is permitted.
The maximum number of points is 100.
Points distribution: 25–20–15–20–20.

1. Suppose that $X_1, X_2, \ldots, X_n$, are independent random variables with

$$X_i \sim N(i\,\theta, 1)$$

for $i = 1, \ldots, n$.

(a) [7pt] Find the maximum likelihood estimator $\hat{\theta}_{MLE}$ of θ

(b) [7pt] Find the variance of $\hat{\theta}_{MLE}$.

(c) [6pt] Compare the variance calculated in (b) with the Cramer–Rao lower bound for an unbiased estimator of θ. Is $\hat{\theta}_{MLE}$ an efficient estimator?

Suppose that we have now another sample $Y_1, \ldots, Y_n$ of i.i.d. random variables $Y_i \sim N(\mu, 1)$, with $i = 1, \ldots, n$ and where $\mu \in \mathbb{R}$ is an unknown parameter. Suppose we do not observe the exact values of Y_i but only their signs, i.e., we only observe $Z_i = \mathrm{sgn}(Y_i)$ for $i = 1, \ldots, n$.

(d) [5pt] Obtain the maximum likelihood estimator of μ

2. Let us suppose to have only **one** observation y from the discrete random variable Y, such that $Y \in \{10, 20, 30, 40, 50, 60\}$. The probability mass function (pmf) $p(y|\theta)$ of Y depends on the unknown parameter θ belonging to the discrete parameter space $\Omega := \{1, 2, 3, 4, 5, 6\}$. The pmf $p(y|\theta)$ is given by the following table:

y	10	20	30	40	50	60
$p(y\|\theta = 1)$	0.5	0.2	0.1	0.1	0.1	0
$p(y\|\theta = 2)$	0.2	0.5	0.1	0.1	0.1	0
$p(y\|\theta = 3)$	0.1	0.2	0.5	0.1	0.1	0
$p(y\|\theta = 4)$	0.1	0.1	0.2	0.5	0.1	0
$p(y\|\theta = 5)$	0.1	0.1	0.1	0.2	0.5	0
$p(y\|\theta = 6)$	0	0.1	0.1	0.1	0.2	0.5

(a) [7pt] Find the maximum likelihood estimator $\hat{\theta}_{MLE}$ of θ.

(b) [4pt] Is $\hat{\theta}_{MLE}$ unbiased?

(c) [5pt] Suppose we want to test:

$$\begin{cases} H_0: & \theta = 1, \\ H_1: & \theta \neq 1. \end{cases}$$

at $\alpha = 0.03$ level of significance. Propose a test statistic and find the rejection region of the test.

(d) [4pt] In case the observation $y = 20$, calculate an estimate of $\mathrm{Var}(\hat{\theta}_{MLE})$.

3. Let $\mathbb{Y} = \{Y_1, \dots Y_n\}$ a random sample of i.i.d. random variables Y_i with probability density function:

$$f(y|\theta) = (1 + \theta\, y)/2$$

with $-1 < y < 1$ and depending on the parameter θ such that $-1 < \theta < 1$.

(a) [7pt] Derive the rejection region B of the general most powerful test with significance α for testing:

$$\begin{cases} H_0 : & \theta = 0, \\ H_1 : & \theta = 1/2. \end{cases}$$

(b) [4pt] When $n = 1$, find the critical value for the test statistics of the test in point (a) such that the significance $\alpha = 0.05$.

(c) [4pt] When $n = 1$, find the power π of the test developed in point (b).

4. Two different types of injection–molding machines are used to form plastic parts. A part is considered defective if it has excessive shrinkage or is discolored. Two random samples, each of size 300, are selected, and 15 defective parts are found in the sample from **machine 1**, while 8 defective parts are found in the sample from **machine 2**.

(a) [9pt] Test the hypothesis that both machines produce the same fraction of defective parts (i.e. $p_1 = p_2$), at $\alpha = 0.05$ level of significance (You can consider the sample large enough for applying large sample results).

(b) [4pt] Find the (approximated) p–value for the test of point (a).

Suppose now that $p_1 = 0.05$ and $p_2 = 0.01$.

(c) [7pt] What is the (approximated) power of the test?

5. Consider the linear model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$, where X is the $n \times p$ design matrix, and $\mathbf{e}$ is the vector whose components e_i are i.i.d. random variables with $\mathbb{E}e_i = 0$ and $\mathrm{Var}(e_i) = \sigma^2$. The least squares estimator of β is given by $\hat{\beta}_{LS} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$. Let $\mathbf{P} := \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ and $\mathbf{Q} := \mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$, where $\mathbf{I}$ is the n–dimensional identity matrix. Let $\widehat{\mathbf{Y}} := \mathbf{X}\hat{\beta}_{LS}$ be the fitted model and $\hat{\mathbf{e}} := \mathbf{Y} - \widehat{\mathbf{Y}}$ the vector of the residuals.

(a) [7pt] Knowing that $\widehat{\mathbf{Y}} = \mathbf{Pe} + \mathbf{X}\beta$, $\hat{\mathbf{e}} = \mathbf{Qe}$ and that $\mathbf{PQ}$ is the zero matrix, prove that $\mathrm{Cov}(\widehat{\mathbf{Y}}, \hat{\mathbf{e}}) = \mathbb{E}(\widehat{\mathbf{Y}}\hat{\mathbf{e}}^\top)$.

(b) [8pt] Suppose now that we add an additional row vector (i.e. a vector of dimensions $1 \times p$) of variables x_{n+1} to the design matrix $\mathbf{X}$. The corresponding variable Y_{n+1} is then predicted by $\widehat{Y}_{n+1} = x_{n+1}\hat{\beta}_{LS}$. Calculate the expectation $\mathbb{E}(\widehat{Y}_{n+1})$ and the variance $\mathrm{Var}(\widehat{Y}_{n+1})$ of $\widehat{Y}_{n+1}$.

(c) [5pt] Suppose we have the additional information that $e_i \sim N(0, \sigma^2)$ with **known** variance σ^2. Construct a $(1 - \alpha)$–confidence interval for $x_{n+1}\beta$.