Statistiek (WISB361)

II Midterm exam

June 28, 2013

Schrijf uw naam op elk in te leveren vel. Ook schrijf uw studentnummer op blad 1.
The maximum number of points is 100.
Points distribution: 20–25–25–30

1. Suppose we have a random sample $\mathbb{Y} := \{Y_1, \ldots, Y_n\}$ of n i.i.d. random variables with density function:

$$f_Y(y;\theta) = \begin{cases} \theta\, y^{\theta-1}, & 0 < y < 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$ is a parameter. Consider the test:

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

(a) [10pt] Determine the generalized likelihood ratio for this testing problem, knowing that the maximum likelihood estimator of θ is: $\hat{\theta}_{MLE} = \frac{-n}{\sum_{i=1}^{n} \log(Y_i)}$

(b) [10pt] Suppose that we observe a random sample of size $n = 8$. Moreover, from the data we have $\sum_{i=1}^{8} \log y_i = -4$. If we consider the sample size $n = 8$ large enough for applying large sample results, test H_0 versus H_1 at the $\alpha = 0.10$ significance level.

2. A company has released a new battery B_{new} which is supposed to replace the standard one B_{old}. In order to compare the duration of the two type of batteries B_{new} and B_{old}, the following two independent samples $\mathbb{Y} = \{Y_i\}_{i=1}^{6}$ and $\mathbb{X} = \{X_i\}_{i=1}^{7}$ were collected, and the life times of the batteries expressed in hours are reported:

Table 1:

Battery B_{new} y_i $n = 6$	Battery B_{old} x_i $m = 7$
0.80	7.26
1.71	2.04
4,10	0.94
6.10	1.76
7.89	11.08
24.10	0.60
	9.04

(a) [10pt] Test the hypothesis that there is no difference of duration between B_{new} and B_{new} at $\alpha = 0.05$ level of significance, without any further assumption on the statistical model generating the data.

(b) [15pt] Suppose we know now that the observations are i.i.d and normally distributed. According to this additional knowledge, which statistical test would be more appropriate for testing the same hypothesis as in point (a)? Perform the test, trying to check the assumptions of the statistical model.
(Useful relation for the F distribution: $F_\alpha(n_1, n_2) = 1/F_{1-\alpha}(n_2, n_1)$).

3. Consider the multivariate regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, with $\boldsymbol{\beta}^\top = (\beta_0, \beta_1, \beta_2)$, $\mathbf{Y}^\top = (Y_1, \ldots, Y_n)$, where $n = 63$ is the sample size and $\mathbf{e}^\top = (\epsilon_1, \ldots, \epsilon_n)$ with ϵ_i i.i.d. $N(0, \sigma^2)$.
The least squares estimates $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ and the corresponding estimated covariance matrix are given by:

$\hat{\boldsymbol{\beta}}^\top = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = (2, 3, -1)$ and

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}) = \begin{pmatrix} 3 & -2 & 1 \\ -2 & 4 & 0 \\ 1 & 0 & 3 \end{pmatrix}$$

Test each of the following hypotheses at 0.05 level of significance and state the conclusion:

(a) [5pt]

$$\begin{cases} H_0: & \beta_1 = 0, \\ H_1: & \beta_1 \neq 0 \end{cases}$$

(b) [10pt]

$$\begin{cases} H_0: & \beta_0 + 2\beta_1 = 5, \\ H_1: & \beta_0 + 2\beta_1 \neq 5, \end{cases}$$

(c) [10pt]

$$\begin{cases} H_0: & \beta_0 - \beta_1 + \beta_2 = 4, \\ H_1: & \beta_0 - \beta_1 + \beta_2 \neq 4, \end{cases}$$

4. Consider the multivariate regression model in matrix form $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, with $\boldsymbol{\beta}^\top = (\beta_0, \beta_1, \ldots, \beta_{p-1})$, $\mathbf{Y}^\top = (Y_1, \ldots, Y_n)$, and $\mathbf{e}^\top = (\epsilon_1, \ldots, \epsilon_n)$ with ϵ_i i.i.d. $N(0, \sigma^2)$.

(a) [10pt] Show that, for any fixed σ^2, the maximum likelihood estimator $\hat{\boldsymbol{\beta}}_{MLE}$ is equal to the least squares estimator $\hat{\boldsymbol{\beta}}$.

(b) [8pt] Does the equality $\hat{\sigma}^2_{MLE} = \hat{\sigma}^2$ hold, where $\hat{\sigma}^2$ denotes the variance estimator based on the residual sum of squares (RSS)?,
(Recall that $\hat{\sigma}^2 = 1/(n-p)RSS = 1/(n-p)\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 = 1/(n-p)(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}))$

(c) [12pt] Consider now the case when the errors ϵ_i are not longer normally distributed but they are i.i.d. with probability density:

$$f(\epsilon) = \frac{1}{2\sigma}\exp(-|\epsilon|/\sigma), \quad \epsilon \in \mathbb{R}$$

Under this assumption, is the maximum likelihood estimator of $\boldsymbol{\beta}$ still equal to the least squares estimator?