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 unknown 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: $\hat{\theta}_{MLE} = \frac{-n}{\sum_{i=1}^n \log(Y_i)}$

Solution:

The GLR test has rejection region $\{\Lambda < c\}$, where

$$\Lambda = \frac{lik(\theta_0)}{lik(\hat{\theta}_{MLE})},$$

$lik(\theta)$ is the likelihood function:

$$lik(\theta) = \theta^n \left(\prod_{i=1}^n y_i \right)^{\theta-1}$$

and, in this case, $\theta_0 = 1$. Hence, we have:

$$\begin{aligned} \Lambda &= \frac{1}{\hat{\theta}^n_{MLE} \left(\prod_{i=1}^n y_i\right)^{\hat{\theta}_{MLE}-1}} = \frac{\left(\prod_{i=1}^n y_i\right)^{1-\hat{\theta}_{MLE}}}{\hat{\theta}^n_{MLE}} = \left(\frac{\sum_i \log y_i}{-n} \right)^n \left(\prod_i y_i \right)^{1+\frac{n}{\sum_i \log y_i}} \\ &= \left(\frac{\sum_i \log y_i}{-n} \right)^n \left(e^{\sum_i \log y_i} \right)^{1+\frac{n}{\sum_i \log y_i}} = n^{-n} \left(-\sum_i \log y_i \right)^n \exp\left(\sum_{i=1}^n \log y_i + n \right) \end{aligned}$$

(b) [10pt] Suppose that we observe a random sample of size $n = 8$. Moreover, from the data we have $\sum_{i=1}^8 \log y_i = -4$. Test H_0 at the *approximate* significance level $\alpha = 0.10$.

Solution:

We have:

$$-2\log\Lambda = -2\left(n \log\left(\frac{-\sum_i \log Y_i}{n} \right) + \sum_i \log Y_i + n \right)$$

By asymptotic results $-2\log\Lambda \approx \chi^2(1-0) = \chi^2(1)$. Thus, for the observed data we have:

$$-\log\Lambda = -2(8\log(4/8) - 4 + 8) \approx 3.09$$

and since $\chi^2_1(0.10) = 2.71$, we can reject H_0 at 0.10 level of significance.

Table 1:

Battery B_{new}	Battery B_{old}
y_i	x_i
$n = 6$	$m = 7$
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

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:

(a) [10pt] Test the hypothesis that there is no difference of duration between B_{new} and B_{old} at $\alpha = 0.05$ level of significance, without any further assumption on the statistical model generating the data.
Solution:
Since we do not have any information about the distribution of the data, we perform a non–parametric two–sided Mann Whitney test. The sum of ranks are: $T_x = 47$, $T_y = 44$. In order to use Table 8 of the textbook: we have $n_1 = 6$, $R = 44$, $R' = 84 - 44 = 40$, $R^\star = \min(40, 44) = 40$. From Table 8, the critical value for R^* is 27. Since $R^\star > 27$ we do not reject the null hypothesis of no difference at 0.05 level of significance.

(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 H_0 of point (a)? Perform this test, trying to check, if possible, the assumptions of the statistical model behind.
(Useful relation for the F distribution: $F_\alpha(n_1, n_2) = 1/F_{1-\alpha}(n_2, n_1)$).
Solution:

Since $\mathbb{X}$ and $\mathbb{Y}$ are two independent normal samples, we perform a F test at 0.05 level of significance for testing the assumption of equal variance of the t test. Thus, we test:

$$\begin{cases} H_0: & \sigma_X^2 = \sigma_Y^2, \\ H_1: & \sigma_X^2 \neq \sigma_Y^2 \end{cases}$$

We have:

$$\bar{x} = 4.67, \ \bar{y} = 7.45$$

$$S_X^2 = 2.65.7 - 7 \times 4.67^2/6 = 18.84, \quad S_Y^2 = 700.65 - 6 \times 7.45^2/5 = 73.53,$$

$$\frac{S_X^2}{S_Y^2} = 0.256, \quad F_{0.975}(6,5) = 6.98, \quad , F_{0.025}(6,5) = 1/F_{0.975}(5,6) = 0.197$$

Since $0,256 \in (0.167, 678)$ we do not reject H_0 of equal variance. We can proceed now to perform a two–sided t test for:

$$\begin{cases} H_0: & \mu_X = \mu_Y, \\ H_1: & \mu_X \neq \mu_Y, \end{cases}$$

We have:

$$s_p^2 = 43.7, \quad \frac{\bar{x} - \bar{y}}{s_p\sqrt{1/7 + 1/6}} = -0.756, \quad df = 7 + 6 - 2 = 11,$$

$$t_{0.0975}(11) = 2.201$$

Since $-0.756 \in (-2.201, 2.201)$ also in this case we do not reject H_0 at 0.05 level of significance.

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

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

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

(a) [5pt]

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

Solution:
In each case we use a two-sided t test with a 0.05 significance level. The critical values are given by $t_{0.025}(63 - 3) = t_{0.025}(60) = -2.000$ and $t_{0.975}(60) = 2.000$. The rejection region is $t < -2$ or $t > 2$.

$$t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}} = \frac{3}{\sqrt{4}} = 1.5$$

Since $-2 < 1.5 < 2$, we fail to reject H_0 and conclude that there is no sample evidence to suggest that $2 \neq 0$.

(b) [10pt]

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

Solution:
We use the statistic:

$$t = \frac{\hat{\beta}_0 + 2\hat{\beta}_1 - 5}{s_{\hat{\beta}_0 + 2\hat{\beta}_1}} = \frac{3}{s_{\hat{\beta}_0 + 2\hat{\beta}_1}}$$

with

$$s^2_{\hat{\beta}_0 + 2\hat{\beta}_1} = \text{Var}(\hat{\beta}_0 + 2\hat{\beta}_1) = \text{Var}(\hat{\beta}_0) + 4\text{Var}(\hat{\beta}_1) + 4\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = 3 + 4 \times 4 - 4 \times 2 = 11$$

Therefore,

$$t = \frac{3}{3.3166} = 0.9045$$

Since $-2 < 0.9045 < 2$, we do not reject H_0.

(c) [10pt]

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

Solution:
We use the statistic:

$$t = \frac{\hat{\beta}_0 - \hat{\beta}_1 + \hat{\beta}_2 - 4}{s_{\hat{\beta}_0 - \hat{\beta}_1 + \hat{\beta}_2}} = \frac{-6}{s_{\hat{\beta}_0 - \hat{\beta}_1 + \hat{\beta}_2}}$$

with

$$\begin{aligned} s^2_{\hat{\beta}_0 - \hat{\beta}_1 + \hat{\beta}_2} &= \mathrm{Var}(\hat{\beta}_0 - \hat{\beta}_1 + \hat{\beta}_2) \\ &= \mathrm{Var}(\hat{\beta}_0) + \mathrm{Var}(\hat{\beta}_1) + \mathrm{Var}(\hat{\beta}_2) - 2\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) + 2\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_2) - 2\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2) = 16 \end{aligned}$$

Thus,

$$t = \frac{-6}{4} = -1.5$$

Since $-2 < -1.5 < 2$, we fail to reject H_0.

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 $\hat{\boldsymbol{\beta}}$.
Solution:
Since the observations are independent, the likelihood is just the product of their density functions, so:

$$lik(\boldsymbol{\beta}, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2}(Y_i - (\mathbf{X}\boldsymbol{\beta})_i)^2 \right\}$$

Thus the log-likelihood is:

$$l(\boldsymbol{\beta}, \sigma) = -n \log \sqrt{2\pi} - n \log \sigma - \frac{1}{2\sigma^2} \sum_{1=1}^{n} (Y_i - (\mathbf{X}\boldsymbol{\beta})_i)^2 = -n \log \sqrt{2\pi} - n \log \sigma - \frac{1}{2\sigma^2} S(\boldsymbol{\beta})$$

where

$$S(\boldsymbol{\beta}) = \mathbf{e}^\top \mathbf{e} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

is the sum of squares. Hence, for any fixed σ^2, $l(\boldsymbol{\beta}, \sigma)$ is maximised when $S(\boldsymbol{\beta})$ is minimised, so that $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{MLE}$.

(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 $\hat{\sigma}^2 = 1/(n-p)\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 = 1/(n-p)(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$?
Solution:
Since we know that $\hat{\sigma}$ is unbiased and that $\hat{\sigma}_{MLE}$ is biased, we could have answered to this question. However, Differentiating the log-likelihood with respect to σ, and setting to zero, we get:

$$-\frac{n}{\sigma} + S(\boldsymbol{\beta})/\sigma^2 = 0$$

so we obtain $\hat{\sigma}^2_{MLE} = S(\hat{\boldsymbol{\beta}})/n$.

(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:

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

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

Solution:

If we assume that the errors have density f, then a similar argument to the one of point a), shows that MLE requires us to minimise

$$LD(\mathbf{Y};\boldsymbol{\beta}) = \sum_{i=1}^{n} |Y_i - (\mathbf{X}\boldsymbol{\beta})_i)|$$

in other words the ℓ_1 norm of the vector $\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$ that in general has different minimizers from the sum of squares.

For instance, consider the simple linear regression with only the intercept β_0. The model is $Y_i = \beta_0 + \epsilon_i$. Clearly, the least squares estimator is the mean $\hat{\beta}_0 = \bar{\mathbf{Y}} = \frac{\sum_{i=1}^n Y_i}{n}$, while the MLE estimator:

$$\hat{\beta}_{0MLE} = \text{argmin}_{\beta_0 \in \mathbb{R}} LD(\mathbf{Y};\beta_0) = \text{argmin}_{\beta_0} \sum_i |Y_i - \beta_0|$$

Let us consider the data $\mathbf{y}^\top = (1, 0, \dots, 0)$, with only the first component different fom 0. Hence, $\hat{\beta}_0 = 1/n$ and $LD(\mathbf{y}; 1/n) = 1 - \frac{1}{n} + (n-1)\frac{1}{n} = 2 - \frac{2}{n} > 1$ when $n > 2$. However, for the choice $\beta_0 = 0$:

$$LD(\mathbf{y};0) = 1 + 0 = 1 < LD(\mathbf{y};1/n) = LD(\mathbf{y};\hat{\beta}_0)$$