Statistiek (WISB263)

Final Exam

January 28, 2019

Schrijf uw naam op elk in te leveren vel. Schrijf ook uw studentnummer op blad 1.
(The exam is an *open–book* exam: notes and book are allowed. The scientific calculator is allowed as well).
The maximum number of points is 110 (10 extra BONUS points!!).
Grade= $min(100, \text{points})$.
Points distribution: 25–16–12-15–32 (+10 extra BONUS points!!)

1. Let $\mathbf{Y} = \{Y_1, \ldots, Y_N\}$ be a random sample such that $Y_i \overset{i.i.d.}{\sim} \text{Pois}(\theta)$ for $i \in \{1, \ldots, N\}$.
   (a) [7pt] Find the Maximum Likelihood Estimator (MLE) of θ and its asymptotic distribution.
   (b) [7pt] Find the MLE of e^θ and show that it is biased.
   (c) [11pt] In case we are able to measure only the first n ($n < N$) observations, and of the other $N - n$ observations we know the value x of the sum, determine the MLE of θ. Compare this result with point (a).

2. Let X be a discrete random variable whose probability mass function $p(x) = \mathbb{P}(X = x)$ under H_0 and H_1 is given by:

x	1	2	3	4	5	6	7
$p(x\|H_0)$	0.01	0.01	0.01	0.01	0.01	0.01	0.94
$p(x\|H_1)$	0.06	0.05	0.04	0.03	0.02	0.01	0.79

   (a) [8pt] Find the most powerful test for testing H_0 versus H_1 with significance level $\alpha = 0.4$.
   (b) [8pt] Compute the power for this test.

3. Consider the oneway ANOVA model:
$$Y_{ij} = \mu_i + \epsilon_{ij}$$
   with $i \in \{1, \ldots, I\}$, $j \in \{1, \ldots, J\}$ and $\epsilon_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$.
   (a) [6pt] Show that set of statistics $(\overline{Y}_{1\cdot}, \overline{Y}_{2\cdot}, \ldots \overline{Y}_{I\cdot}, S_p^2)$ is sufficient for $(\mu_1, \mu_2, \ldots, \mu_I, \sigma^2)$, where $\overline{Y}_{i\cdot} := \frac{1}{J}\sum_{j=1}^{J} Y_{ij}$ and $S_p^2 := \frac{SS_W}{I(J-1)}$.
   (b) [6pt] Show that S_p^2 is independent of each $\overline{Y}_{i\cdot}$.

4. An algorithm is developed for generating pseudo-random numbers. We want to test now this algorithm by performing an experiment in which it produces $n = 10000$ digits (i.e. each digit is an integer in $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$). Suppose the following frequencies are observed:

Digit	0	1	2	3	4	5	6	7	8	9
Frequency	1007	987	928	986	1010	1029	987	1006	1034	1026

   (a) [15pt] Choose an appropriate test for testing at $\alpha = 0.05$ level of significance whether the algorithm is a proper pseudo-random number generator. Explain clearly your choice and perform the test.

5. Consider the multiple linear regression model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

with $\boldsymbol{\beta}^\top = (\beta_0, \beta_1, \beta_2)$, $\mathbf{Y}^\top = (Y_1, \dots, Y_n)$, where $n = 40$ is the sample size and $\mathbf{e}^\top = (\epsilon_1, \dots, \epsilon_n)$ with ϵ_i i.i.d. $N(0, \sigma^2)$.

We obtain the following estimates of the least squares estimators and for the residual sum of squares: $\hat{\boldsymbol{\beta}}^\top = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = (3, 4, -3)$; $\hat{\mathbf{e}}^\top \hat{\mathbf{e}} = 37$, with $\hat{\mathbf{e}} := \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$. Moreover, we know that:

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \begin{pmatrix} 3 & -2 & 1 \\ -2 & 4 & 0 \\ 1 & 0 & 3 \end{pmatrix}$$

(a) [10pt] Find 95% CIs for each of $\beta_0, \beta_1, \beta_2$.

(b) [10pt] Test the following hypotheses at 0.05 level of significance and state the conclusion:

$$\begin{cases} H_0: & \beta_0 + 2\beta_1 = 10, \\ H_1: & \beta_0 + 2\beta_1 \neq 10, \end{cases}$$

(c) [5pt] Given the fitted linear model, we want now to make a prediction. We are interested indeed at the predicted value $\hat{Y}$ corresponding to the values of the predictors: $x_1 = 1; x_2 = -1$. Give a 95% CI for the *true* prediction (i.e. $\beta_0 + \beta_1 x_1 + \beta_2 x_2$).

(d) [7pt] Suppose now that σ^2 is known and that $\sigma^2 = 1$. Are the least square estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ independent? Prove your statement.

6. **BONUS Ex.1** [5pt]: By using properly the CLT, calculate the following limit:

$$\lim_{n \to +\infty} e^{-n} \sum_{k=0}^{n+\sqrt{n}} \frac{n^k}{k!}$$

7. **BONUS Ex.2** [5pt]: Let $\mathbf{X} = \{X_1, \dots, X_n\}$ be a random sample such that $X_i \overset{i.i.d.}{\sim} \text{Unif}(\frac{1}{\theta}, \theta)$ with $\theta > 1$.

(a) Find the Maximum Likelihood Estimator (MLE) of θ and state conditions on the sample $\mathbf{X}$ for guaranteeing its existence.