UNIVERSITY OF PADUA
SCHOOL OF ENGINEERING

Inferential Statistics (A.Y. 2020/2021)

July 16, 2021, time allowed: 2 hours

INSTRUCTIONS TO CANDIDATES

- This **is not an open book** exam.
- This examination paper contains **2** (two) questions and comprises two printed pages.
- Answer all questions in clear and readable English. The marks for each sub-question are indicated at the beginning of each sub-question.
- Candidates may use hand calculators or R. Each solution must be fully justified.

Question 1.

A computer scientist is interested in modelling the dynamics of cyber attacks (y_i, daily counts) against time t_i in days, i.e. $t_1 = 1$ for day 1, $t_2 = 2$ for day 2, etc.. She suspects a quadratic effect with time and uses the following linear regression model

$$\begin{aligned} Y_i &\sim \mathrm{N}(\mu_i, \sigma^2), \\ \mu_i &= \beta_0 + \beta_1 t_i + \beta_2 t_i^2, \quad i = 1, \dots, n. \end{aligned}$$

Observations from one day to the next are independent and $(\beta_0, \beta_1, \beta_2) \in \mathbb{R}^3$, $\sigma^2 > 0$ are unknown parameters. Setting $X = [1_n|T|T^2]$, with $1_n = (1, \dots, 1)$ an $n \times 1$ vector, $T = (t_1, \dots, T_n)$, $T^2 = (T_1^2, \dots, T_n^2)$ and $n = 150$,

$$X^T X = \begin{bmatrix} 67 & 5142 & 530034 \\ ?? & 530034 & 61462338 \\ ?? & ?? & 7601706498 \end{bmatrix}, \quad X^T y = \begin{bmatrix} 9348 \\ 740407 \\ 72923845 \end{bmatrix}$$

$$\widehat{\mathrm{cov}}(\hat{\beta}) = \begin{bmatrix} 37.9038847 & -0.9810866 & 0.0052895 \\ ?? & 0.0344132 & -0.0002098 \\ ?? & ?? & 0.0000014 \end{bmatrix}.$$

(i) **(1 pts)** Fill in all the spaces denoted by '??'.

(ii) **(4 pts)** Derive the minimum least squares estimator of $\beta = (\beta_0, \beta_1, \beta_2)$ and compute the least squares estimate.

(iii) **(2 pts)** Compute an approximate confidence interval for β_1 with confidence level 99% and interpret the result.

(iv) **(3 pts)** For the model above, set $\beta = \hat{\beta}_{LS}$. Write the likelihood function and the log-likelihood function of σ^2. Derive $\hat{\sigma}^2_\beta$, the maximum likelihood estimator for σ^2 at the fixed value of β.

Question 2.

Let $X_1, \ldots, X_n$, be i.i.d. r.v.'s with p.d.f. $f(y; r, \theta) = \binom{y+r-1}{y}\theta^r(1-\theta)^y$, where r and θ are parameters.

(i) **(3 pts)** Specify, the parameter space. Then specify, the statistical model, the likelihood function and the log-likelihood function.

(ii) **(3 pts)** Is there a sufficient statistic for the parameter (r, θ)? If so, find it. If not explain why. And if r is a known constant, is there a sufficient statistic for θ? If. so, find it. If not explain why

(iii) **(2 pts)** Assuming $r = 4$, compute a method of moments estimator for θ.

(iv) **(3 pts)** Compute $\hat{\theta}$, the maximum likelihood estimator for θ and find its exact distribution.

(v) **(3 pts)** Let $\tau = \log[\theta/(1-\theta)]$. Compute the MLE of τ without appealing to the equivariance principle of MLE and find a distribution for $\hat{\tau}$.

(vi) **(4 pts)** Assuming $r = 2$ and for a sample of size $n = 10$ compute a most powerful test for $H_0 : \theta = 1/2$ against $H_1 : \theta = 4/5$ of size or level $\alpha = 0.05765915$. With this test, let $\bar{x} = 12$, what do you conclude?

(vii) **(3 pts)** Let $Y_1, \ldots, Y_m$ be annother iid sample with each Y_j having density function $f(y; 2, \lambda)$, $j = 1, \ldots, m$. Recalling that X_i's are iid r.v.'s with X_i having density $f(y; 2, \theta)$, $i = 1, \ldots, n$, construct an approximate test for $H_0 : \theta = \lambda$ against $H_1 : \theta \neq \lambda$. Compute an approximate p-value for these hypothesis given that $n = 10$, $m = 12$, $\sum_{i=1}^{n} x_i = 13$ and $\sum_{j=1}^{m} y_j = 17$.