Exam Statistics 30^{th} August 2022 (a)

Maura Mezzetti

Exercise 1

Let $(X_1, \ldots, X_n)$ be a random sample of i.i.d. random variables. Let $f_\theta(x)$ and $F_\theta(x)$ be the density function and cumulative distribution function respectively. Let $\hat{\theta}$ be the MLE of θ, θ_0 be the true parameter, $L(\theta)$ be the likelihood function, $logL(\theta)$ be the loglikelihood function, and $I(\theta)$ be the Fisher information matrix

1. To verify the hypothesis $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, which of the following statements is NOT true:

   (a) Wald test requires only the unrestricted model to be estimated.

   (b) Score test requires only the restricted model to be estimated.

   (c) LR test requires both the restricted and unrestricted models to be estimated

   (d) The three statistical tests provide the same values in the observed sample

2. For each of the following statements indicate whether it is TRUE or FALSE:

If a test is accepted at the significance level α, the probability that the null hypothesis is true equals α	T	F
A statistic $T = T(X)$ is a sufficient statistic for θ if it can be computed without knowing the value of θ	T	F
The likelihood ratio is a random variable	T	F
The following function is a density function f(x)= $\begin{cases} 6x^2 - 1 & if x \in [0, 1] \\ 0 & otherwise \end{cases}$	T	F
Probability of Type I error is always greater than probability of Type II error	T	F
Let $(X_1, \ldots, X_n)$ be a random sample of i.i.d. r.v. with $E(X) = \theta$. The following estimator $T_n(a) = a \times X_n + (1 - a) \times \bar{X}_{n-1}$ where X_n is the $n - th$ observed random variable and $\bar{X}_{n-1}$ is the sample mean based on $n - 1$ observations is an unbiased estimator for θ if and only if $a = 1$	T	F

Exercise 2

Let $(X_1, \ldots, X_n)$ be a random sample of i.i.d. random variables distributed as follows:

$$f(x;\theta) = \frac{(\theta - 1)\, 2^{\theta-1}}{x^{\theta}} \qquad x > 2 \quad \theta > 1$$

1. Show that $\sum_i log(X_i)$ is a sufficient statistics for θ
2. Find $\hat{\theta}_{MLE}$ maximum likelihood estimator (MLE) for θ and discuss properties of this estimator.
3. Find $\hat{\theta}_{MOM}$ method of moment estimator (MOM) for θ.
4. Compute the score function and the Fisher information.
5. Specify asymptotic distribution of $\hat{\theta}_{MLE}$.
6. Suppose form a random sample of 120 random variables you obtain $\sum_{i=1}^{120} log(x_i) = 115$. Fix $\alpha = 0.01$ and complete ALL of the following questions:
   (a) Find the Likelihood Ratio test statistic for testing $H_0 : \theta = 5$ versus $H_1 : \theta \neq 5$, specify the distribution and verify the null hypothesis.
   (b) Find the Wald test statistic for testing $H_0 : \theta = 5$ versus $H_1 : \theta \neq 5$, specify the distribution and verify the null hypothesis.
   (c) Find the Score test statistic for testing $H_0 : \theta = 5$ versus $H_1 : \theta \neq 5$, specify the distribution and verify the null hypothesis.

Exercise 3

In a casino, you observe a table where the game involves tossing a die. You suspect that the die of the dealer may be tricked. You record the result of each game. For game i, n_i is the number of times the die was tossed and X_i the number of times a 6 appeared. You recorded N games in all.

1. What is the distribution of X_i?
2. Find $\hat{p}_{MLE}$ maximum likelihood estimator (MLE) for $p = P(get\, a\, 6)$
3. Find the Likelihood Ratio Test statistics of level $\alpha = 0.05$ for $H_0 : p = 1/6$ against the alternative $H_1 : p \neq 1/6$
4. You repeat the experiment 100 times and observe $\sum_i x_i = 375$ and $\sum_i n_i = 2000$, decide whether accept or reject the null hypothesis.

Exercise 4

You studied the lifetime of electronic components, but since your study can not last forever, some of the components were still in working order when you had to stop your data collection. Among the n components you studied, m failed at time $X_i, \quad i = 1, ..., m$ (in hours) and the $n - m$ others did last for the 10000 hours that the study lasted, so $X_{m+1}, \ldots, X_n$ are all greater than 10000, but could not be observed. The exponential distribution is a good model for the lifetime of these components. You would like to estimate the parameter λ of that exponential.

1. Find the survival function of of X,

$$S(x) = P(X > x),$$

where X follows an exponential with mean $1/\lambda$.

2. Find $\hat{\lambda}_{MLE}$ the maximum likelihood estimator of λ from the sample above. Because you did not observe the exact value of each variable, the likelihood is

$$L(\lambda) = \left\{ \prod_{i=1}^{m} f(Y_i) \right\} \left\{ \prod_{i=m+1}^{n} S(Y_i) \right\}$$

where $Y_i = min(X_i, 10000)$ are the values you actually observed.

Exercise 5

Total precipitation (in mm) has been recorded on a daily basis at the Vancouver airport for many years. The largest of these daily totals for a given years is called the annual maximum; it describes the amount of precipitation on the "wettest" day of the year. You have a data file listing the annual maxima for the last n years. These n annual maxima do not show any trend over time, so a statistical model in which these are modelled as independent and identically distributed random variables seems reasonable. Assume that the exponential distribution provides a reasonable model. So, if X_i denotes the annual maxima for year i, our statistical models is $X_1, X_2, \ldots, X_n$ is a simple random sample from the population with density function $f(x)$ given by:

$$f(x) = \lambda exp(-\lambda x) \qquad for\ x > 0$$

Our primary interest is in θ, the probability that next year's annual maximum will exceed 100mm.

1. Express θ as a function of λ

2. Find $\hat{\theta}_{MLE}$ maximum likelihood estimator (MLE) for θ and its asymptotic distribution

Exercise 5

Define the likelihood ratio test, specifying the use and specifying the distribution of the likelihood ratio test statistics under null hypothesis (as you have defines).