New York University
Applied Statistics and Econometrics I
Fall 2015

Problem Set 2

Cristian Fuenzalida
cristian.fuenzalida@nyu.edu

(I) Write an R routine that:

1. Calculates percentile values for *any given pair* of degrees of freedom in the numerator and denominator in an F distribution. Please note that your routine must work like a function.
2. Plots an exact $F(6,8)$ and an exact $F(6,20)$ in the same graph.
3. Simulates a sample of 500 observations from a distribution $F(10,24)$. Plot the histogram of your simulations and an exact F distribution in the same graph.

(II) We are interested in modelling the default time of a set of loans. Suppose that we observe the default times $X_1, \dots, X_N$ and we are told these default times come from an exponential distribution:

$$f(x|\lambda) = \lambda e^{-\lambda x}$$

1. A broker tells you he only needs the statistic $Y = \min\{X_1, \dots, X_N\}$ to make an estimation of the true parameter. Can you find an unbiased estimator of λ based only on Y?
2. Find a better estimator of λ than the one suggested by the broker. State clearly the criteria you use to claim it is indeed better.
3. The file Data.txt contains information on default times for different sets of loans. Each column represents one type of loan with an idiosyncratic λ parameter. Write a R code to compute the mean default time for each set of mortgages, and determine a 95% confidence interval for each estimator.

(III) Consider two X_1, X_2 uniform i.i.d. variables with support $[\theta, \theta+1]$. We want to test H_0 : $\theta = 0$ against $H_1 : \theta \neq 0$. Calculate the size and power of the following tests:

size of test = Pr(type I error)
power of test = 1- Pr(type II error)

I : rej of null hyp when it is true
II: fail to rej of null hyp when it is false

1. Always reject H_0, no matter what data are obtained.
2. Always accept H_0, no matter the data.
3. Reject H_0 if $X_1 > 0.95$.
4. Reject H_0 if $X_1 + X_2 > C$.

Find the value of C such that the last two tests have the same size. Plot in R each power function. Is the last test more powerful that the third one?

(IV) Consider two r.v. Z and Y. Assume that $Var(Y|Z) = \sigma^2 Z^2$, and let $\mu_t = \mathbb{E}(Z^t)$. Define:

$$W_1 = (Y - \mathbb{E}(Y|Z))Z^2$$

$$W_2 = (Y - \mathbb{E}(Y|Z))^2 Z^2$$

1. Express $\mathbb{E}(W_1)$ and $Var(W_1)$ in terms of σ^2 and μ_t. HINT: Use the law of iterated expectations here: $\mathbb{E}(\mathbb{E}(U|V)) = \mathbb{E}(U)$.
2. Express $\mathbb{E}(W_2)$ in terms of σ^2 and μ_t.

(V) Let $X \sim N(\mu, \sigma^2)$, with σ^2 unknown. Let $\bar{X}$ denote the simple average statistic, and $\hat{\sigma}^2$ denote the sample variance.

1. Let N be the sample size. Determine the probability distribution of:

$$z := \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{N}}$$

   What happens with this distribution when N is large?

2. Let $N = 25$. Find the rejection region for a two-tailed test of the null hypothesis $H_0 : \mu = 40$ against the alternative $H_1 : \mu \neq 40$ at the 0.05 level of significance.

(VI) Suppose that the variables $Z_1, \ldots, Z_N$ are independent and identically distributed random variables whose density function is:

$$f(z|\theta) = \frac{1}{\theta} e^{-z/\theta}$$

We observe a sample $z_1, \ldots, z_N$. Find the MLE of θ. Compute the Information matrix.

(VII) We are given a discrete random variable Y with support $S = \{1, 2, 3\}$ and probability distribution $\mathbb{P}(Y = 1) = \theta_1$, $\mathbb{P}(Y = 2) = \theta_2$. Let $Y_1, \ldots, Y_N$ be an i.i.d. sample from this population.

1. Find $\mathbb{E}(Y)$
2. Derive the log-likelihood function for this model.
3. Let $(\hat{\theta}_1, \hat{\theta}_2)$ denote the ML estimator of (θ_1, θ_2). Show that:

$$\hat{\theta}_1 = \frac{\sum_i \mathbb{1}(Y_i = 1)}{N}$$

$$\hat{\theta}_2 = \frac{\sum_i \mathbb{1}(Y_i = 2)}{N}$$

4. Use the Central Limit Theorem to find the limit distribution of:

$$\sqrt{N}\left[\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} - \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}\right]$$

.

(VIII) Let $Y_1, \ldots, Y_M$ a sequence of independent and identically distributed Bernoulli random variables with mean p. The statistic average is denoted by $\bar{Y}$:

$$\bar{Y} = \frac{1}{M} \sum_{i=1}^{M} Y_i$$

Define the following statistic:

$$Z := \bar{Y}(1 - \bar{Y})$$

And the following parameter:

$$\eta := p(1 - p)$$

1. Prove that Z is a downward biased estimator of η.
2. Assume that $p \neq 1/2$. Show that $\sqrt{N}(Z - \eta) \xrightarrow{d} N(0, \sigma^2)$. What is σ^2?
3. Assume now that $p = 1/2$. What is the asymptotic distribution of $N(Z - \eta)$?