INSTRUCTIONS

1. The exam contains multiple choice questions that must be answered in the orange code sheet we have provided you with, together with three exercises that must be answered in detail in separate sheets of paper.
2. To select an answer, all you need to do is mark in the orange code sheet, **filling the rectangle over which the selected answer is located appropriately**. Please make sure you know the answer you wish to mark before doing it. Even though you can always erase your mark if you have used a pencil (number 2 or similar), any mark that has not been completely erased could be read by the machine. Therefore, we advice you to first mark your selected answers in the exam and to use only the last ten minutes or so from the time assigned to the multiple choice questions-part of the exam to copy them into the code sheet.
3. In the multiple choice questions-part of the exam there is always **only one correct answer** for every question. Every question correctly answered is worth 1 point, while each question incorrectly answered will penalize your grade by 0.2 points. Questions that have not been answered do not penalize your grade in any form.
4. Each one of the exercises, A, B and C, must be answered in a separate sheet of paper. We will collect the different parts of the exam at the indicated times and in this order: first, we will collect the code sheet for the multiple choice questions part of the exam and, then, and in this order, exercises A, B and C will be collected.
5. The exam has six numbered sheets, going from 0.1 to 0.6. Please make sure that you have all sheets and contact your professor if this is not the case. There are different exam types. This exam is of type 0. Mark a 0 in the column labelled with I in your code sheet, just as is illustrated in the example.
6. The maximum final grade for each of the parts of the exam (i.e., the multiple choice questions part and the exercises part) is 30 points.
7. Please fill in your personal information in the appropriate places both in the code sheet and in the sheets provided for the exercises. In "Resit" (column II) you will write the number of times you have registered for an exam in this course *not including this one.*

Example:

$\underbrace{12545}$ PEREZ, Ernesto Exam type $\underbrace{0}$ Resit

MULTIPLE CHOICE QUESTIONS (Time: 1 hour and 30 minutes)

1. FREE-QUESTION. The capital of Spain is:

   (A) Paris (B) Sebastopol (C) Madrid (D) London (E) Pekin

Questions 2 to 4 refer to the following exercise:

In a vehicle technical inspection site (i.e., ITV) it is known, from previous experience, that 96% of the vehicles that go there for inspection pass it in a satisfactory manner. We assume independence between the distributions for the different vehicles.

2. If 10 vehicles go to this inspection site, what is the probability that one of them **does not pass** the technical inspection?

   (A) 0.277 (B) 0.487 (C) 0.344 (D) 0.165 (E) 0.665

3. If 200 vehicles go to this inspection site, what is the approximate probability that between 7 and 18 of them, including 7 and 18, **do not pass** the technical inspection?

   (A) 0.546 (B) 0.407 (C) 0.686 (D) 0.491 (E) 0.234

4. If 200 vehicles go to this inspection site, what is the **exact** distribution of the number of vehicles that **do pass** the technical inspection?

   (A) b(0.96, 200) (B) $\mathcal{P}(\lambda = 8)$ (C) All false (D) $N(m = 192, \sigma^2 = 7.68)$ (E) b(0.04, 200)

Questions 5 to 7 refer to the following exercise:

The number of vehicles that arrives at a toll paying site in a given highway follows a Poisson distribution with mean equal to 2 vehicles per minute. We assume independence between the distributions for the different minutes.

5. What is the probability that, in a given minute, no vehicle arrives at the highway toll paying site?

   (A) 0.4060 (B) 0.8647 (C) 0.1353 (D) 0.5940 (E) 0.9985

6. What is the probability that, in a given minute, at least one vehicle arrives at the highway toll paying site?

   (A) 0.8647 (B) 0.1353 (C) 0.3441 (D) 0.4060 (E) 0.5940

7. What is the approximate probability that, in a given hour, more than 130 vehicles arrive at the highway toll paying site?

   (A) 0.1685 (B) 0.9128 (C) 0.8315 (D) 0.0872 (E) 0.3227

Questions 8 and 9 refer to the following exercise:

Let X_1, X_2 and X_3 be three independent and identically distributed r.v. having a $\gamma(1, 1)$ distribution.

8. The distribution of the r.v. $Y = X_1 + X_2 + X_3$ is:

   (A) $\gamma(1, 3)$ (B) $\gamma(1, 1)$ (C) All false (D) $\gamma(1/3, 3)$ (E) $\gamma(1/3, 1)$

9. The distribution of the r.v. $Z = 3X_1$ is:

(A) $\gamma(1/3, 1)$ (B) $\gamma(1, 1)$ (C) $\gamma(1, 3)$ (D) $\gamma(1/3, 3)$ (E) All false

Questions 10 to 12 refer to the following exercise:

Let X and Y be independent r.v. such that $X \in N(3, \sigma^2 = 9)$ and $Y \in \chi^2_9$.

10. The distribution of the r.v. $Z = \frac{X-3}{\sqrt{Y}}$ is:

(A) t_9 (B) $F_{1,9}$ (C) All false (D) $F_{9,1}$ (E) t_1

11. The probability $P(Z \leq 2.82)$ is:

(A) 0.02 (B) 0.99 (C) 0.98 (D) 0.01 (E) 0.50

12. The probability $P(Z^2 \geq 3.36)$ is:

(A) 0.10 (B) 0.95 (C) 0.50 (D) 0.90 (E) 0.05

Questions 13 and 14 refer to the following exercise:

Let X be a r.v. with probability density function given by:

$$f_X(x) = \begin{cases} (\theta+2)x^{-(\theta+3)} & x > 1, \theta > 0 \\ 0 & \text{otherwise,} \end{cases}$$

and $X_1, \ldots, X_n$ a r.s. of size n taken from this distribution.

It is known that the mean of this distribution is $m = \frac{\theta+2}{\theta+1}$.

13. The method of moments estimator of the parameter θ is:

(A) $\left(\dfrac{\overline{X}-2}{1+\overline{X}}\right)$ (B) $(\overline{X}-2)$ (C) $\left(\dfrac{\overline{X}-2}{1-\overline{X}}\right)$ (D) $(2-\overline{X})$ (E) All false

14. The maximum likelihood estimator of the parameter θ is:

(A) $\frac{n}{\ln\left(\prod_i X_i\right)}$ (B) $\frac{n}{\ln\left(\prod_i X_i\right)} - 2$ (C) All false (D) $\frac{-n}{\ln\left(\prod_i X_i\right)}$ (E) $\frac{n-1}{\ln\left(\prod_i X_i\right)}$

Questions 15 to 17 refer to the following exercise:

We have a normal distribution whose mean is unknown and we wish to estimate. In order to do so, two estimators are proposed: $\hat{\theta}_1$ is the sample mean for a r.s. of size n and $\hat{\theta}_2$ is the average of the even components of this sample of size n (you can assume that n is even). That is,

$$\hat{\theta}_1 = \frac{X_1 + X_2 + \cdots + X_n}{n}, \qquad \hat{\theta}_2 = \frac{X_2 + X_4 + \cdots + X_n}{\frac{n}{2}}$$

15. Which one of the two estimators is unbiased?

(A) $\hat{\theta}_2$ (B) Both estimators are unbiased. (C) $\hat{\theta}_1$ (D) None of them is unbiased. (E) -

16. Which one of these estimators has smaller variance?

(A) $\hat{\theta}_1$ (B) They have the same variance. (C) - (D) It depends on the sample values. (E) $\hat{\theta}_2$

17. Which one of these estimators is consistent?

(A) $\hat{\theta}_2$ (B) Both estimators are consistent. (C) $\hat{\theta}_1$ (D) None of them is consistent. (E) -

Questions 18 to 20 refer to the following exercise:

Under the null hypothesis H_0, the r.v. X has the following probability mass function:

X	1	2	3	4	5	6
$P_0(X)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

whereas, under the alternative hypothesis H_1, its probability mass function is:

X	1	2	3	4	5	6
$P_1(X)$	$\frac{2}{15}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{2}{15}$

The null hypothesis H_0 is rejected if the observed values of the r.v. X are equal to either 3 or 4.

18. The probability of a type I error for this test is:

(A) $\frac{1}{3}$ (B) $\frac{1}{6}$ (C) 1 (D) $\frac{1}{9}$ (E) $\frac{2}{3}$

19. The probability of a type II error for this test is:

(A) $\frac{2}{3}$ (B) $\frac{1}{5}$ (C) $\frac{3}{5}$ (D) $\frac{2}{5}$ (E) 1

20. The power of this test is:

(A) $\frac{3}{5}$ (B) $\frac{1}{5}$ (C) 1 (D) $\frac{2}{5}$ (E) $\frac{2}{3}$

Questions 21 and 22 refer to the following exercise:

We wish to test the null hypothesis that the probability density function for a r.v. X with domain on the interval $(0, 1)$ (i.e., $x \in (0, 1)$) is $f(x) = 1$ against the alternative hypothesis that it is $f(x) = 2x$ instead. In order to do so, a random sample of size $n = 1$ is taken.

21. At the $\alpha = 0.05$ significance level, the most powerful critical region for this test is:

(A) $[0.95, 1)$ (B) $[0.975, 1)$ (C) $(0, 0.05]$ (D) $(0, 0.025]$ (E) All false

22. For the critical region in the previous item, the power of the test is:

(A) 0.0025 (B) 0.9975 (C) 0.9506 (D) 0.0975 (E) 0.9025

23. We wish to test the null hypothesis $H_0 : m = 5$ against the alternative hypothesis $H_1 : m = 8$. In order to do so, a r.s. of size $n = 100$ has been taken. It is assumed that the population follows a normal distribution with variance equal to 4. At the $\alpha = 5\%$ significance level, the decision of the test will be to reject the null hypothesis H_0 if the sample mean is larger than or equal to:

(A) 5.328 (B) 7.282 (C) 6.124 (D) 8.328 (E) 9.124

Questions 24 and 25 refer to the following exercise:

A firm devoted to the growing of apples wishes to start a new marketing option, which will be successful if the variance of the weight of the apples is at most 50gr^2, but it will not be successful if the variance is larger that this amount. In order to be able to make a decision on this issue, a r.s. of size $n = 10$ has been taken obtaining that $s^2 = 53$. We assume normality in the distribution of the weights of the apples.

24. The 95% confidence interval for the population variance is:

    (A) $(27.89, 196.30)$ (B) $(38.92, 162.12)$ (C) $(55.34, 140.19)$ (D) $(57.34, 138.19)$ (E) $(31.36, 159.16)$

25. If $H_0 : \sigma^2 \leq 50\text{gr}^2$ and, at the $\alpha = 5\%$ significance level, the firms's decision will be:

    (A) Do not reject H_0. (B) - (C) Reject H_0. (D) - (E) -

Questions 26 and 27 refer to the following exercise:

We wish to test if the attitude clients have towards a given product is the same in Spain and France. In order to do so, two r.s. (one from each country) of sizes 100 and 200, respectively, have been taken, where clients were classified according to regular, sporadic and no consumption of the product under study.

26. The most appropriate test for this setting is:

    (A) Test of homogeneity

    (B) All false

    (C) Test for the ratio of two variances

    (D) χ^2 goodness of fit test to a totally specified distribution

    (E) Test of independence

27. If, after having the same objective in mind, a single random sample of 100 individuals in Spain and 200 in France was taken, the most appropriate test for this new setting would now be:

    (A) All false

    (B) Test for the ratio of two variances

    (C) Test of homogeneity

    (D) χ^2 goodness of fit test to a totally specified distribution

    (E) Test of independence

Questions 28 to 30 refer to the following exercise:

We wish to estimate the mean milk weekly consumption per individual with a maximum absolute error of 0.5 liters and a confidence of 95% in a city having 2000 inhabitants. We assume that the distribution of the milk consumption is normal with variance equal to 16 liters^2.

28. If a simple random sample with replacement is taken, the required sample size is, approximately:

    (A) 280 (B) 220 (C) 246 (D) 215 (E) 325

29. What should be this sample size if the sampling is now taken without replacement?

    (A) 325 (B) 220 (C) 280 (D) 215 (E) 246

30. It is known that the milk consumption is different with regard to the age of the individuals. Because of this, two groups are considered: the first one will include the 500 inhabitants with age smaller than 16 years old, and the second one will include the remaining 1500 inhabitants with age larger than or equal to 16 years old. In addition, it is also known that the quasi standard deviations for the first and second groups are 2 and 3 liters, respectively. If stratified random sampling with n-optimal allocation has been used and the total sample size is 200 inhabitants, the number of inhabitants in the sample belonging to the first group (age smaller than 16 years old) will be:

    (A) 100 (B) 36 (C) 82 (D) 26 (E) 50

EXERCISES (Time: 75 minutes)

A. (10 points, 25 minutes)

Let X be a r.v. such that $X \in \gamma(a,3)$; that is, having probability density function given by :

$$f_X(x;a) = \frac{a^3}{2} x^2 e^{-ax}, \quad x > 0, \quad a > 0$$

We wish to test the null hypothesis $H_0 : a = \frac{1}{2}$ against the alternative hypothesis $H_1 : a = 2$. In order to do so, a random sample of size $n = 5$, $X_1, \ldots, X_5$, has been taken.

i) Obtain the general form of the most powerful critical region for this test.

ii) What is the distribution of the test statistic under the null hypothesis?

iii) For an $\alpha = 0.05$ significance level, find the specific most powerful critical region for this test.

B. (10 points, 25 minutes)

A given high school suggests that the probability that one of its students finds a job within the first three months after he/she has finalized his/her studies is larger than or equal to 0.7. In order to verify this hypothesis, the education department takes a r.s. of 100 alumni (i.e., individuals who studied in that school), obtaining that 65 found a job within that specific deadline.

i) Obtain an approximate 0.95 confidence interval for the proportion of students who found a job within the specified deadline.

ii) Based on the information provided by the sample, and at the approximate 5% significance level, what will be the education department's decision about the hypothesis claimed by the school?

C. (10 points, 25 minutes)

Let X be a r.v. with probability mass function given by:

$$P(x,\lambda) = \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} \qquad x = 1, 2, ..., \qquad \lambda > 0$$

It is known that the mean and variance of the r.v. X are $m = \lambda + 1$ and $\sigma^2 = \lambda$, respectively.

In order to be able to estimate the parameter λ, a r.s. of size n, $X_1, \ldots, X_n$ has been taken.

i) Obtain **providing all relevant details** the method of moments estimator of the parameter λ.

ii) Obtain **providing all relevant details** the maximum likelihood estimator of the parameter λ.

iii) Let $\hat{\lambda} = \overline{X} - 1$ be an estimator of the parameter λ. Is it unbiased? Is it consistent? Is it efficient?

Remark: All of the answers provided should be appropriately justified.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)

1: C	11: B	21: A
2: A	12: A	22: D
3: C	13: C	23: A
4: A	14: B	24: A
5: C	15: B	25: A
6: A	16: A	26: A
7: A	17: B	27: E
8: A	18: A	28: C
9: A	19: C	29: B
10: A	20: D	30: B

SOLUTIONS TO EXERCISES

Exercise A

i) Let X be a r.v. with probability density function given by

$$f_X(x;a) = \frac{a^3}{2}x^2e^{-ax}, \quad x>0, \quad a>0$$

We wish to test the null hypothesis $H_0 : a = \frac{1}{2}$ against the alternative hypothesis $H_1 : a = 2$.

For a given significance level, the most powerful critical region is obtained from the likelihood ratio test. That is,

$$\frac{L(\vec{x}|H_0)}{L(\vec{x}|H_1)} \leq k$$

In this case, and given that the likelihood function is given by:

$$L(\vec{x},a) = \frac{a^{3n}}{2^n}\left(\Pi_{i=1}^n x_i^2\right)e^{-a\sum_{i=1}^n x_i},$$

the most powerful critical region will be obtained from:

$$\frac{L(\vec{x}|H_0)}{L(\vec{x}|H_1)} = \frac{\frac{\frac{1}{2}^{3n}}{2^n}\left(x_1\dots x_n\right)^2 e^{-\frac{1}{2}(x_1+x_2+\ldots+x_n)}}{\frac{2^{3n}}{2^n}\left(x_1\dots x_n\right)^2 e^{-2(x_1+x_2+\ldots+x_n)}} =$$

$$= \frac{\frac{1}{2}^{3n}e^{-\frac{1}{2}(x_1+x_2+\ldots+x_n)}}{2^{3n}e^{-2(x_1+x_2+\ldots+x_n)}} = \frac{1}{2^{6n}}e^{\frac{3}{2}(x_1+x_2+\ldots+x_n)} \leq k_1$$

$$\frac{3}{2}(x_1+x_2+\ldots+x_n) \leq k_2$$

$$s = \sum_{i=1}^n x_i \leq C$$

Therefore, the most powerful critical region for the test statistic $S = \sum_{i=1}^n x_i$ is of the form $(0, C]$.

ii) Given that, under H_0, $X_i \in \gamma\left(\frac{1}{2}, 3\right)$, an that, in addition, $n = 5$, if we use the convolution property for the gamma distributions, we have that,

$$S = X_1 + X_2 + \ldots + X_5 \in \gamma\left(\frac{1}{2}, 15\right),$$

which is equivalent to stating that

$$S = X_1 + X_2 + \ldots + X_5 \in \chi^2_{30}$$

iii) We wish to find the value of C for the case of a test having a significance level $\alpha = 5\%$. That is, we wish to find the value of C such that

$$\alpha = 0.05 = P(S \leq C|H_0)$$

As we have already stated in the previous item, under the null hypothesis, $S \in \chi^2_{30|}$, so that the value of C that will satisfy the aforementioned condition is $C = \chi^2_{30|0.95} = 18.5$. Therefore, the most powerful critical region for an $\alpha = 5\%$ significance level for this test will be of the form CR $= (0, 18.5]$.

Exercise B

i) We wish to obtain the approximate 95% confidence level interval for the proportion of students finding a job within the first three months after they have finalized their studies. Given that the required conditions for the convergence of the binomial to the normal distribution hold, the interval will be given by

$$\mathrm{CI}_{1-\alpha}(p) \simeq \left(\frac{z}{n} \pm t_{\frac{\alpha}{2}}\sqrt{\frac{z(n-z)}{n^3}}\right)$$

In this case, we have that $n = 100$, $z = 65$ and $t_{0.025} = 1.96$, so that

$$\mathrm{CI}_{0.95}(p) = \left(0.65 \pm t_{0.025}\sqrt{\frac{65 \cdot 35}{100^3}}\right) = (0.65 \pm 0.093) = (0.557, 0.743)$$

ii) We wish to test the hypotheses:

$H_0 : p \geq 0.7$

$H_1 : p < 0.7$

The test statistic is given by $\frac{\frac{z}{n} - p_0}{\sqrt{\frac{p_0 q_0}{100}}}$. At the α significance level, the null hypothesis will be rejected if:

$$\frac{\frac{z}{n} - p_0}{\sqrt{\frac{p_0 q_0}{100}}} < -t_\alpha,$$

For the specific case where $\alpha = 0.05$, the null hypothesis is rejected if:

$$\frac{\frac{z}{n} - p_0}{\sqrt{\frac{p_0 q_0}{100}}} < -t_{0.05} = -1.64$$

Given that for the specific test under study, we have that $p_0 = 0.7$ and $\frac{z}{n} = 0.65$, so that

$$\frac{\frac{z}{n} - p_0}{\sqrt{\frac{p_0 q_0}{100}}} = \frac{0.65 - 0.7}{\sqrt{\frac{0.7(1-0.7)}{100}}} = -1.091$$

Therefore, at the approximate 5% significance level, the null hypothesis is not rejected.

Exercise C

The probability mass function for the r.v. X is:

$$P(x, \lambda) = \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} \qquad x = 1, 2, ..., \qquad \lambda > 0$$

It is known that the mean and variance of the r.v. X are $m = \lambda + 1$ and $\sigma^2 = \lambda$.

i) Method of moments estimator: In order to be able to obtain the method of moments estimator of the parameter λ, we need to equate the first population moment $\alpha_1 = \mathrm{E}(X) = m = \lambda + 1$ to the first sample moment. That is,

$$\alpha_1 = a_1 \Longrightarrow m = \overline{X} \Longrightarrow \lambda + 1 = \overline{X} \Longrightarrow \hat{\lambda}_{\mathrm{MM}} = \overline{X} - 1$$

ii) Maximum likelihood estimator: In order to be able to obtain the maximum likelihood estimator of the parameter λ, we need to obtain the likelihood function, which, in this case, is given by:

$$L(\vec{x};\lambda) = P(x_1;\lambda)\dots P(x_n;\lambda) = \frac{e^{-\lambda}\lambda^{x_1-1}}{(x_1-1)!}\cdots\frac{e^{-\lambda}\lambda^{x_n-1}}{(x_n-1)!} =$$

$$L(\vec{x};\lambda) = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^n x_i - n}}{\Pi_{i=1}^n (x_i-1)!}$$

If we take its natural logarithm, we have that:

$$\ln L(\vec{x};\lambda) = -n\lambda + \left(\sum_{i=1}^n x_i - n\right)\ln\lambda - \ln\left[\Pi_{i=1}^n (x_i-1)!\right]$$

If we take derivatives with respect to λ and make it equal to zero, we have that:

$$\frac{\partial \ln L(\vec{x},\lambda)}{\partial\lambda} = -n + \frac{\sum_{i=1}^n x_i - n}{\lambda} = 0,$$

so that

$$\frac{\sum_{i=1}^n x_i - n}{\lambda} = n \Longrightarrow \hat{\lambda}_{\text{ML}} = \frac{\sum_{i=1}^n X_i}{n} - 1 = \overline{X} - 1$$

iii) Unbiasedness: In order to see if the estimator $\hat{\lambda}$ is unbiased for the parameter λ, we need to verify if $\mathrm{E}(\hat{\lambda}) = \lambda$. In this case, we have that

$$\mathrm{E}(\hat{\lambda}) = \mathrm{E}\left(\overline{X} - 1\right) = \mathrm{E}\left(\overline{X}\right) - 1 = m - 1 = (\lambda+1) - 1 = \lambda$$

Therefore, $\hat{\lambda}$ is an unbiased estimator of λ.

Consistency: We first compute the variance of the estimator $\hat{\lambda}$:

$$\mathrm{Var}(\hat{\lambda}) = \mathrm{Var}(\overline{X}-1) = \mathrm{Var}(\overline{X}) = \frac{\mathrm{Var}(X)}{n} = \frac{\lambda}{n}$$

Given that it is an unbiased estimator and that, in addition, its variance tends to zero as n goes to infinity, we can state that $\hat{\lambda}$ is a consistent estimator of λ.

Efficiency: In order to see if the estimator is efficient, we need to compute the Cramer-Rao lower bound for unbiased and regular estimators. That is,

$$L_c = \frac{1}{n\mathrm{E}\left(\frac{\partial \ln P(x;\lambda)}{\partial\lambda}\right)^2}$$

We start by taking the logarithm of the probability mass function and, then, we take its derivative with respect to λ:

$$\ln P(x;\lambda) = -\lambda + (x-1)\ln\lambda - \ln(x-1)! \Longrightarrow \frac{\partial \ln P(x;\lambda)}{\partial\lambda} = -1 + \frac{x-1}{\lambda} = \frac{x-(\lambda+1)}{\lambda}$$

Therefore

$$\mathrm{E}\left(\frac{\partial \ln P(x;\lambda)}{\partial \lambda}\right)^2 = \mathrm{E}\left[\frac{X-(\lambda+1)}{\lambda}\right]^2 = \left(\frac{1}{\lambda^2}\right)\ \mathrm{E}\left[X-(\lambda+1)\right]^2 = \frac{1}{\lambda^2}\mathrm{Var}(X) = \frac{1}{\lambda^2}(\lambda) = \frac{1}{\lambda}$$

If we now replace this in the formula for the Cramer-Rao lower bound, we have that:

$$L_c = \frac{1}{n\left(\frac{1}{\lambda}\right)} = \frac{\lambda}{n} = \mathrm{Var}(\hat{\lambda})$$

Thus, $\hat{\lambda}$ is an efficient estimator of λ.