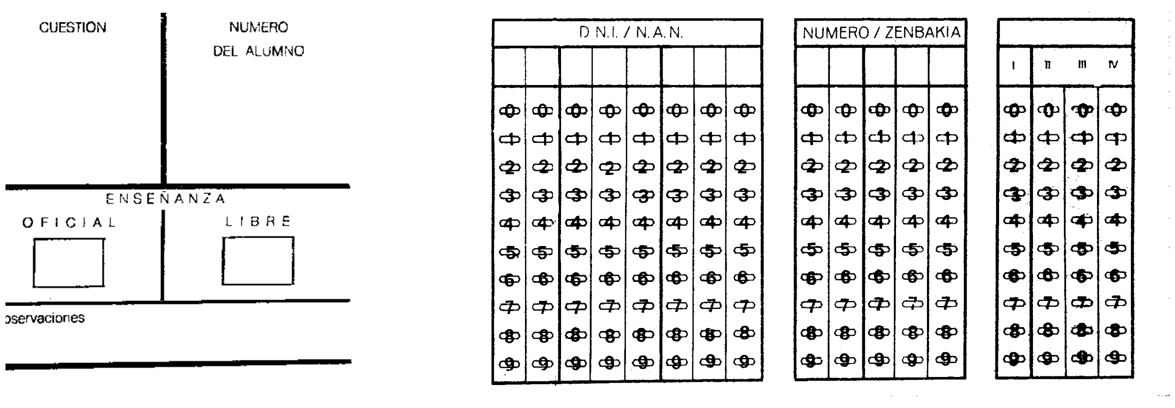
INSTRUCTIONS

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 is 30 points. You will need to obtain 15 points in each part of the exam to pass it.
7. Please fill in your personal information in the appropriate places both in the code sheet and in the sheets provided for the exercises.

Example:

ID: $\underbrace{12345678}$ Exam type $\underbrace{0}$



MULTIPLE CHOICE QUESTIONS (Time: 1 hour and 45 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:

The probability that a given supermarket's cashier makes an error when preparing the bill for a given client is 0.01. We assume independence between the different clients' bills.

2. If a client buys 6 times in that supermarket, and his/her bill is prepared by the aforementioned cashier, what is the probability of a bill's error in one of his/her six visits to the supermarket?

   (A) $\binom{6}{1}$ (0.01) $(0.99)^5$ (B) (0.01) $(0.99)^5$ (C) All false

   (D) $\sum_{x=0}^{6} \binom{6}{x}$ $(0.01)^x$ $(0.99)^{6-x}$ (E) $1 - \binom{6}{1}$ (0.01) $(0.99)^5$

3. If the aforementioned cashier prepares 200 bills, what is the approximate probability of an error in exactly two of those bills?

   (A) 0.27 (B) 0.59 (C) 0.68 (D) 0.33 (E) 0.84

4. If the aforementioned cashier prepares 5000 bills, what is the approximate probability of an error in 46 of those bills?

   (A) 0.05 (B) 0.01 (C) 0.74 (D) 0.10 (E) 0.26

5. Let X be a r.v. with a Poisson distribution such that $P(X = 1) = P(X = 2)$. Then, the value of $P(X = 4)$ is:

   (A) 0.2707 (B) 0.1465 (C) 0.0902 (D) 0.0122 (E) 0.3847

6. Let X_1, X_2, X_3 and X_4 be four independent random variables with a Poisson distribution. The first two of those r.v. have a mean equal to 0.4, and the last two have a mean equal to 0.6. Let us now consider the random variable $Z = \sum_{i=1}^{4} X_i$. The probability that $Z \geq 5$ is equal to:

   (A) 0.053 (B) 0.004 (C) 0.273 (D) 0.017 (E) 0.195

Questions 7 to 9 refer to the following exercise:

Let X be a random variable having a $\gamma(4, 1)$ distribution.

7. The mean and variance of the r.v. X are, respectively:

   (A) $\frac{1}{4}$ and $\frac{1}{16}$ (B) 4 and $\frac{1}{4}$ (C) 4 and 16 (D) 4 and 8 (E) $\frac{1}{4}$ and $\frac{1}{8}$

8. If we let $Y = 8X$, then the distribution of the r.v. Y is:

   (A) $\gamma(4, 8)$ (B) $\exp(\lambda = 0.5)$ (C) $\gamma(\frac{1}{2}, 2)$ (D) $\exp(\lambda = 4)$ (E) $\gamma(32, 2)$

9. The value of $P(Y > 4.61)$ is:

   (A) 0.25 (B) 0.01 (C) 0.10 (D) 0.75 (E) 1

10. Let X be a r.v. with a $\gamma(\frac{1}{2},\frac{3}{2})$ distribution, and Y be another r.v., independent of X, with a $N(0,1)$ distribution. If we let $Z = X + Y^2$, then the distribution of the r.v. Z is:

   (A) Snedecor's F with $(1,2)$ degrees of freedom

   (B) χ^2 with 2 degrees of freedom

   (C) χ^2 with 4 degrees of freedom

   (D) Student's t with 2 degrees of freedom

   (E) All false

Questions 11 to 13 refer to the following exercise:

Let X_1, X_2, X_3 and X_4 be independent random variables with distributions such that: $X_1 \in N(1, \sigma^2 = 1)$, $X_2 \in N(2, \sigma^2 = 4)$, $X_3 \in N(3, \sigma^2 = 4)$ and $X_4 \in N(4, \sigma^2 = 4)$.

11. The probability that the r.v. $Y = \left(\frac{X_2-2}{2}\right)^2 + \left(\frac{X_3-3}{2}\right)^2 + \left(\frac{X_4-4}{2}\right)^2$ belongs to the interval $(1.21, 7.81)$ is:

   (A) 0.05 (B) 0.70 (C) 0.75 (D) 0.80 (E) 0.30

12. If we let $Z = \dfrac{\left(\frac{X_2-2}{2}\right)^2 + \left(\frac{X_3-3}{2}\right)^2 + \left(\frac{X_4-4}{2}\right)^2}{3\,(X_1-1)^2}$, then the value of k such that $P(Z < k) = 0.1$ is equal to:

   (A) 53.60 (B) 0.18 (C) -2.35 (D) 5.54 (E) -1.64

13. The probability that the r.v. $V = \dfrac{\sqrt{3}\,(X_1-1)}{\sqrt{\left(\frac{X_2-2}{2}\right)^2 + \left(\frac{X_3-3}{2}\right)^2 + \left(\frac{X_4-4}{2}\right)^2}}$ is less than or equal to -2.35 is:

   (A) 0.90 (B) 0 (C) 0.95 (D) 0.10 (E) 0.05

Questions 14 and 15 refer to the following exercise:

We have a population with the following probability distribution: $P(1) = \lambda$, $P(2) = \frac{\lambda}{2}$, $P(3) = 1 - \frac{3\lambda}{2}$.

In order to be able to obtain the estimate of the parameter λ, a r.s. of size $n = 10$ has been taken providing the following results: 1, 1, 3, 1, 2, 1, 2, 1, 1, 3.

14. The method of moment estimate of λ is:

   (A) 0.56 (B) 0.16 (C) 0.32 (D) 0.12 (E) 0.28

15. The maximum likelihood estimate of λ is:

   (A) 0.27 (B) 0.20 (C) 0.53 (D) 0.60 (E) 0.40

Questions 16 and 17 refer to the following exercise:

Let $X_1, \ldots, X_n$ a r.s. from a r.v. X, with probability density function given by:

$$f(x,\theta) = \begin{cases} \theta x^{\theta-1} & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

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

(A) $\dfrac{-n}{\ln \prod_{i=1}^{n} X_i}$ (B) $\sum_{i=1}^{n} X_i$ (C) $\prod_{i=1}^{n} X_i$ (D) $\dfrac{-n}{\prod_{i=1}^{n} \ln X_i}$ (E) $\ln \prod_{i=1}^{n} X_i$

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

(A) $\dfrac{\overline{X}}{1-\overline{X}}$ (B) $\dfrac{1}{1-\overline{X}}$ (C) $\overline{X}$ (D) $\overline{X}(1-\overline{X})$ (E) $\dfrac{1-\overline{X}}{\overline{X}}$

Questions 18 to 20 refer to the following exercise:

Let $X_1, \ldots, X_n$ be a r.s. from population having a $N(m, \sigma^2 = 16)$ distribution. In order to estimate the mean m, we consider the following estimators:

$\hat{m}_1 = \dfrac{X_1 + \ldots + X_n}{n-1}$ and $\hat{m}_2 = \dfrac{X_1 + \ldots + X_n}{n+1}$

18. We can state that:

(A) Both estimators are unbiased (B) $\hat{m}_1$ is unbiased, but $\hat{m}_2$ is biased (C) - (D) $\hat{m}_2$ is unbiased, but $\hat{m}_1$ is biased (E) Both estimators are biased

19. Which one of the aforementioned estimators has that smallest variance?

(A) $\hat{m}_2$ (B) They have the same variance (C) $\hat{m}_1$ (D) The variance cannot be computed (E) -

20. We can state that:

(A) Both estimators are efficient (B) $\hat{m}_1$ is efficient, but $\hat{m}_2$ is not (C) - (D) $\hat{m}_2$ is efficient, but $\hat{m}_1$ is not (E) None is efficient

Questions 21 and 22 refer to the following exercise:

We wish to test the null hypothesis that the proportion of people favoring a given proposal is of at least 60%. In order to do so, a random sample of size $n = 5$ has been taken, and the null hypothesis is rejected if the number of people favoring the proposal is smaller than two.

21. The significance level for this test is approximately equal to:

(A) 0.337 (B) 0.002 (C) 0.087 (D) 0.186 (E) 0.543

22. The probability of a type II error, when the true value of the proportion is equal to 0.4, is approximately equal to:

(A) 0.663 (B) 0.843 (C) 0.359 (D) 0.176 (E) 0.337

Questions 23 and 24 refer to the following exercise:

Let X be a r.v. with probability density function $f(x) = \theta x^{\theta-1}$ for $x \in (0,1)$. In order to be able to test the null hypothesis $H_0 : \theta = 1$ against the alternative hypothesis $H_1 : \theta = 4$, a random sample of size $n = 1$ has been taken.

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

(A) $(0, 0.05)$ (B) $(0.025, 0.975)$ (C) $(0.95, 1)$ (D) $(0.025, 0.975)^c$ (E) $(0.987, 1)$

24. The power of this test is:

(A) 0.1855 (B) 0 (C) 0.9500 (D) 0.0625 (E) 0.8145

Questions 25 and 26 refer to the following exercise:

Let X and Y be two independent r.v. with distributions $X \in N(m_X, \sigma_X^2 = 25)$ and $Y \in N(m_Y, \sigma_Y^2 = 36)$, respectively. In order to test the hypothesis $H_0 : m_X = m_Y$ against $H_1 : m_X \neq m_Y$, two random samples, each of size 30 are taken in each one of the two populations, providing $\overline{x} = 82$ and $\overline{y} = 80$.

25. A 95% confidence interval for $(m_X - m_Y)$ is, approximately:

(A) $(-0.900,\ 4.900)$ (B) $(-0.346,\ 4.346)$ (C) $(-0.795,\ 4.795)$
(D) $(0.574,\ 3.426)$ (E) $(-1.985,\ 5.985)$

26. At the $\alpha = 5\%$ significance level, the decision of the test will be:

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

Questions 27 and 28 refer to the following exercise:

A given town has 5000 inhabitants and is thinking about building a sports centre. It would be a profitable investment if at least 20% of its inhabitants make use of it on a regular basis. In order to be able to make a decision, a r.s. of 500 people is taken, and 90 of them would use it on a regular basis.

27. A 95% confidence interval for proportion of inhabitants that would eventually use the sports centre is, approximately:

(A) $(0.175,\ 0.185)$ (B) $(0.146,\ 0.214)$ (C) $(0.191,\ 0.247)$ (D) $(0.179,\ 0.180)$ (E) $(0.152,\ 0.208)$

28. At the 5% significance level, what will be the decision the town should make so that it is in agreement with the inhabitants' view on this issue?

(A) -

(B) It cannot be determined with the available information.

(C) Do not build the sports centre.

(D) Build the sports centre.

(E) All false

Questions 29 and 30 refer to the following exercise:

In a given population of 5000 families, we wish to estimate the proportion of families having at least one member who has spent the last summer holidays in a foreign country.

29. If the sample is taken with replacement, and the proportion is estimated with an error equal to 0.03, and a confidence level of 0.95, the required sample size will be:

(A) 385 (B) 880 (C) 1068 (D) 32 (E) 357

30. If the sample is taken without replacement, and the proportion is estimated with the same error and confidence level objectives, the required sample size will be :

(A) 1068 (B) 880 (C) 357 (D) 32 (E) 385

EXERCISES (Time: 75 minutes)

A. (10 points, 25 minutes)

It is known that, in a given city, the family annual expense in dairy products, in euros, follows a normal distribution. From a r.s. of 31 families from this city, we have obtained a mean and a variance equal to 535 and 1600, respectively.

i) Find the 95% confidence interval for the mean of the family annual expense in dairy products in that city.

ii) Al the 5% significance level, test the null hypothesis that the family mean annual expense in dairy products in that city is of at least 550.

iii) Find the 90% confidence interval for the variance of the family annual expense in dairy products in that city.

iv) At the 10% significance level, test the null hypothesis that the variance of the family annual expense in dairy products in that city is equal to 1500, against the alternative that it is not equal to that value.

B. (10 points, 25 minutes)

Let $X_1, \ldots, X_n$ be a r.s. from a random variable X, measuring the time, in minutes, that is required for an employee to compile and verify the documents handed in a given counter in the City Hall office. It is known that this variable follows an exponential distribution with probability density function given by:

$$f(x, \theta) = \frac{1}{3\theta} e^{-\frac{1}{3\theta}x}, \qquad x \geq 0, \qquad \theta > 0$$

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

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

iii) Is the method of moments estimator of the parameter θ unbiased? Consistent? **You should appropriately justify all of your answers for this specific item**.

iv) A given morning, 10 people were at the specific City Hall office counter, and the times required for the employee to compile and verify the documents handed in were: 3.17, 2.54, 2.63, 1.87, 0.23, 6.53, 4.28, 7.92, 1.03, 2.80. Obtain a method of moments estimate of the parameter θ.

C. (10 points, 25 minutes)

The owners of a wholesale tourism firm wish to design their travel offers for the new summer season. In order to do so, they are interested in testing if the probability distribution for the different tourist destinations is the following one:

$P(\text{Cost of Spain}) = 2\theta$, $P(\text{Inland Spain}) = \theta$, $P(\text{Rest of Europe}) = \theta$, $P(\text{Rest of the world}) = 1 - 4\theta$

In order to test the null hypothesis that the probability distribution is the one they have claimed, they have decided to take a poll among their regular clients. In order to do so, a r.s. of 1000 clients is taken, and the firm asked their clients about their destinations for the upcoming summer season, providing the following results: 450 clients will go to the cost of Spain, 180 to the Spanish inland, another 170 to the rest of Europe, and the remaining 200 to the rest of the world.

i) Obtain, **providing all relevant details**, the maximum likelihood estimate of the parameter θ.

ii) At the 5% significance level, is the firm's belief the correct one?

iii) Obtain the 95% confidence interval for the proportion of clients that will go to the rest of the world.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)

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

SOLUTIONS TO EXERCISES

Exercise A

i) Given that we have a normal distribution with unknown variance, the corresponding confidence interval will be of the form:

$$\mathrm{CI}_{1-\alpha}(m) = \left(\overline{x} \pm t_{\overline{n-1}|\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}} \right)$$

$$\mathrm{CI}_{0.95}(m) = \left(535 \pm 2.04 \sqrt{\frac{1600}{30}} \right) = (535 \pm 14.898) = (520.102, 549.898)$$

ii)

$H_0 : m \geq 550$

$H_1 : m < 550$

At the α significance level, we reject the null hypothesis if:

$$\frac{\overline{x} - m_0}{\frac{s}{\sqrt{n-1}}} \leq -t_{\overline{n-1}|\alpha}$$

In our case, we have that:

$$\frac{\overline{x} - m_0}{\frac{s}{\sqrt{n-1}}} = \frac{535 - 550}{\sqrt{\frac{1600}{30}}} = \frac{-15}{7.303} = -2.054 < -1.70 = t_{\overline{30}|0.05}$$

Therefore, at the 5% significance level, we reject the null hypothesis.

iii)

$$\mathrm{CI}_{1-\alpha}(\sigma^2) = \left(\frac{ns^2}{\chi^2_{\overline{n-1}|\frac{\alpha}{2}}} \ , \ \frac{ns^2}{\chi^2_{\overline{n-1}|1-\frac{\alpha}{2}}} \right)$$

$$\mathrm{CI}_{0.9}(\sigma^2) = \left(\frac{31 \cdot 1600}{43.8} \ , \ \frac{31 \cdot 1600}{18.5} = \right)$$
$$= (1132.42, \ 2681.08)$$

iv)

$H_0 : \sigma^2 = 1500$

$H_1 : \sigma^2 \neq 1500$

We have a bilateral test, so that it suffices to verify if $\sigma_0^2 = 1500$ belongs or not to the aforementioned confidence interval in item (iii). In this case, we have that:

$$1500 \in (1132.42 \ , \ 2681.08) = \mathrm{CI}_{0.9}(\sigma^2)$$

Therefore, at the 10% significance level, we do not reject the null hypothesis.

Exercise B

Given that the r.v. X follows an exponential distribution, its probability density function is given by:

$$f(x;\theta) = \frac{1}{3\theta}\ e^{-\frac{1}{3\theta}x}, \qquad x > 0 \qquad \theta > 0$$

Therefore, X will follow an exponential distribution with parameter $\lambda = \frac{1}{3\theta}$. Moreover, $\mathrm{E}(X) = \frac{1}{\lambda} = 3\theta$ and $\mathrm{Var}(X) = \frac{1}{\lambda^2} = 9\theta^2$

i) **Method of moments estimator:**

We need to equate the first population moment to the first sample moment. That is,

$$\alpha_1 = \mathrm{E}(X) = a_1 = \frac{\sum_{i=1}^n X_i}{n} = \overline{X}$$

$$3\theta = \overline{X} \qquad \Rightarrow \qquad \hat{\theta}_{\mathrm{MM}} = \frac{\overline{X}}{3}$$

ii) **Maximum likelihood estimator:**

The likelihood function for the sample is given by:

$$\begin{aligned} L(\vec{x};\theta) &= f(x_1;\theta)\dots f(x_n;\theta) = \\ &= \left(\frac{1}{3\theta}\ e^{-\frac{1}{3\theta}x_1}\right)\ \left(\frac{1}{3\theta}\ e^{-\frac{1}{3\theta}x_2}\right)\ \dots\ \left(\frac{1}{3\theta}\ e^{-\frac{1}{3\theta}x_n}\right) = \\ &= \left(\frac{1}{3\theta}\right)^n\ e^{-\frac{1}{3\theta}\sum_{i=1}^n x_i} \\ &= \frac{1}{3^n\theta^n}\ e^{-\frac{1}{3\theta}\sum_{i=1}^n x_i} \end{aligned}$$

If we take its natural logarithm, we have that:

$$\ln L(\vec{x};\theta) = -n\ \ln 3 - n\ln\theta - \frac{1}{3\theta}\sum_{i=1}^n x_i$$

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

$$\frac{\partial \ln L(\vec{x},\theta)}{\partial\theta} = \frac{-n}{\theta} + \frac{1}{3\theta^2}\sum_{i=1}^n x_i = 0$$

so that,

$$n3\theta = \sum_{i=1}^n x_i$$

$$\hat{\theta}_{\mathrm{ML}} = \frac{\sum_{i=1}^n X_i}{3n} = \frac{\overline{X}}{3}$$

iii) **Unbiasedness**

In order to check if the method of moments estimator is unbiased, we need to verify if $\mathrm{E}(\hat{\theta}_{\mathrm{MM}}) = \theta$ holds. In this case, we have that:

$$\mathrm{E}(\hat{\theta}_{\mathrm{MM}}) = \mathrm{E}\left(\frac{\overline{X}}{3}\right) = \frac{1}{3}\,\mathrm{E}(\overline{X}) = \frac{1}{3}\,\mathrm{E}(X) = \frac{1}{3}\ (3\theta) = \theta$$

Therefore, $\hat{\theta}_{\text{MM}}$ is an unbiased estimator of θ.

Consistency

In order to be able to check if the method of moments estimator is consistent, we need to compute its variance.

$$\text{Var}(\hat{\theta}_{MM}) = \text{Var}\left(\frac{\overline{X}}{3}\right) = \frac{1}{9}\,\text{Var}(\overline{X}) = \frac{1}{9}\,\frac{\text{Var}(X)}{n} = \frac{\text{Var}X}{9n} = \frac{1}{9}\,\frac{9\theta^2}{n} = \frac{\theta^2}{n}$$

Given that the method of moments estimator is unbiased and that, in addition, its variance tends to zero as n goes to infinity, the required conditions for its consistency hold and, thus, $\hat{\theta}_{\text{MM}}$ is a consistent estimator of θ.

iv) **Method of moments estimate:**

In item (ii), we have obtained the method of moments estimator of θ:

$$\hat{\theta}_{\text{MM}} = \frac{\sum_{i=1}^{n} X_i}{3n} = \frac{\overline{X}}{3}$$

If we now make use of the available information from the sample, we have that:

$$\overline{x} = \frac{\sum_{i=1}^{10} x_i}{n} = \frac{3.17 + 2.54 + 2.63 + 1.87 + 0.23 + 6.53 + 4.28 + 7.92 + 1.03 + 2.80}{10} = 3.3$$

Therefore,

$$\hat{\theta}_{\text{MM}} = \frac{\overline{x}}{3} = \frac{3.3}{3} = 1.1$$

Exercise C

The information provided in the exercise can be summarized in the following table:

	p_i	n_i
Cost of Spain	2θ	450
Inland Spain	θ	180
Rest of Europe	θ	170
Rest of the World	$1-4\theta$	200
Total	1	$n = 1000$

i) **Maximum likelihood estimation:**

The likelihood function for the sample is given by:

$$\begin{aligned} L(\vec{x};\theta) &= f(x_1;\theta)\ldots f(x_n;\theta) = \\ &= (2\theta)^{450}\ \theta^{180}\ \theta^{170}\ (1-4\theta)^{200} = \\ &= 2^{450}\ \theta^{800}\ (1-4\theta)^{200} \end{aligned}$$

If we take its natural logarithm, we have that:

$$\ln L(\vec{x};\theta) = 450\ln 2 + 800\ln\theta + 200\ln(1-4\theta)$$

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

$$\frac{\partial \ln L(\vec{x},\theta)}{\partial\theta} = \frac{800}{\theta} + 200\left[\frac{-4}{(1-4\theta)}\right] = 0$$

$$\frac{800}{\theta} = \frac{800}{(1-4\theta)} \quad \Rightarrow \quad \frac{1}{\theta} = \frac{1}{(1-4\theta)} \quad \Rightarrow \quad (1-4\theta) = \theta \quad \Rightarrow \quad 1 = 5\theta$$

so that:

$$\hat{\theta}_{\text{ML}} = \frac{1}{5} = 0.2$$

ii) **Test of hypothesis:**

We have a goodness-of-fit test to a partially specified distribution. $H_0 : X$ follows the aforementioned specified distribution, where θ is an unknown parameter that has previously been estimated by the maximum likelihood method ($\hat{\theta}_{\text{ML}} = 0.2$) in item (i).

We now build the corresponding table for the test, computing the estimates for the theoretical probabilities $\hat{p}_i$, from the probability mass function provided by the firm and the estimate obtained for the parameter θ:

	p_i	n_i	$\hat{p}_i$	$n\hat{p}_i$	$\frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$
Cost of Spain	2θ	450	0.4	400	6.25
Inland Spain	θ	180	0.2	200	2.00
Rest of Europe	θ	170	0.2	200	4.50
Rest of the World	$1-4\theta$	200	0.2	200	0.00
Total	1	$n = 1000$	1	$n = 1000$	$z = 12.75$

Under the null hypothesis, the test statistic $\sum_i \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$ converges to a $\chi^2_{\overline{k-h-1}|}$ distribution, where k is the number of classes into which the sample is divided ($k = 4$), and h is the numbers of parameters that have been estimated ($h = 1$).

The decision indicates that, at the approximate 5% significance level, the null hypothesis should be rejected if:

$$z > \chi^2_{\overline{4-1-1}|0.05} = \chi^2_{\overline{2}|0.05}$$

In this case, we have that:

$$12.75 > 5.99 = \chi^2_{\overline{4-1-1}|0.05} = \chi^2_{\overline{2}|0.05},$$

so that, at the approximate 5% significance level, the null hypothesis that the distribution for the different tourist destinations is the one the firm claims it to be is rejected.

iii) **Confidence interval:**

Given that the number of clients indicating that their destination for the upcoming summer season is the rest of the world is $z = 200$, the approximate 95% confidence interval for the proportion of clients having this destination for the summer will be:

$$\begin{aligned}
\mathrm{CI}_{1-\alpha}(p) &= \left(\frac{z}{n} \pm t_{\alpha/2}\sqrt{\frac{z(n-z)}{n^3}}\right) \\
\mathrm{CI}_{0.95}(p) &= \left(\frac{z}{n} \pm t_{0.025}\sqrt{\frac{z(n-z)}{n^3}}\right) = \\
&= \left(\frac{200}{1000} \pm 1.96\sqrt{\frac{200 \cdot 800}{1000^3}}\right) = \\
&= (0.2 \pm 0.0248) = \\
&= (0.1751, 0.2248)
\end{aligned}$$