Exam

Duration: 3h

No documents except for one hand-written page, both sides.
No electronic device.

The exam conssts of 6 independent exercises. Quantile tables of some distributions are given in the appendix.

We recall the Beta and Gamma distributions, as well as the Gamma function:

Beta distribution. A random variable X has a Beta distribution with parameters $a, b > 0$, if it has the density:

$$\forall\, x \in \mathbb{R}, \quad f_X(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}\mathbb{1}_{[0,1]}(x).$$

We have:

$$\mathbb{E}(X) = \frac{a}{a+b}, \quad \mathrm{var}(X) = \frac{ab}{(a+b)^2(a+b+1)}.$$

In the particular case $a = b = 1$, it is a uniform distribution over $[0, 1]$.

Gamma distribution. A random variable X has a Gamma distribution with parameters $a, \lambda > 0$, if it has the density:

$$\forall\, x \in \mathbb{R}, \quad f_X(x) = \frac{1}{\Gamma(a)} \lambda^a x^{a-1} e^{-\lambda x}\mathbb{1}_{[0,+\infty)}(x).$$

We have:

$$\mathbb{E}(X) = \frac{a}{\lambda}, \quad \mathrm{var}(X) = \frac{a}{\lambda^2}.$$

In the particular case $a = 1$, it is an exponential distribution with parameter λ.

Gamma function. The Gamma function is defined by:

$$\forall\, a > 0, \quad \Gamma(a) = \int_0^{+\infty} t^{a-1} e^{-t} \mathrm{d}t.$$

We have $\Gamma(1) = 1$ and $\Gamma(a+1) = a\Gamma(a)$.

1 Point estimation

The occupancy rate of a show room is supposed to have a Beta distribution with parameters $\theta, 1$, where $\theta > 0$ is some unknown parameter. We would like to estimate θ using n i.i.d. samples $X_1, \dots, X_n$. We denote by X the vector $(X_1, \dots, X_n)$.

1. Show that the random vector X has the following probability density function:

$$\forall\, x \in \mathbb{R}^n, \quad p_\theta(x) = \prod_{i=1}^{n} \theta x_i^{\theta-1} 1_{[0,1]}(x_i).$$

2. Deduce the maximum likelihood estimator for θ.
3. Is this estimator biased?
4. Propose another estimator using the method of moments.
5. Is this new estimator biased?

2 Cramer-Rao's bound

For the same model as in the previous exercise, we now aim at estimating the average occupancy rate of the show room, that is:

$$g(\theta) = \frac{\theta}{\theta + 1}.$$

We propose the following estimator:

$$\widehat{g}(x) = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

where $x_1, \dots, x_n$ are the n observed values.

1. Show that the estimator $\widehat{g}$ is unbiased.
2. Calculate its quadratic risk.
3. Which value(s) of θ maximize the quadratic risk?
4. Calculate the Fisher information of the considered statistical model.
5. Is the estimator $\widehat{g}$ efficient in the sense of Cramer-Rao's bound?

3 Bayesian approach

We consider the same model as in Exercice 1 but with a Bayesian approach, the parameter θ having an exponential distribution with parameter $\lambda > 0$ (prior distribution, in the absence of any observations).

1. Show that the posterior distribution of θ given $X = x$ has a Gamma distribution, with parameters to be specified.
2. Deduce an estimator of θ minimizing Bayes' quadratic risk.
3. Show that this estimator is consistent, in the sense that it tends almost surely to θ when the number of observations n tends to $+\infty$.
4. How the parameter λ should be chosen in order to minimize the effect of the prior distribution on the estimator?

4 Hypothesis testing

A dice with 6 sides is suspected to be unfair, giving twice as more 6 as usual (that is, 1 out of 3), the other sides being equally likely. We propose to test this hypothesis through a statistical test.

1. Describe the problem in the form of a statistical test. In particular, specify the null hypothesis and the alternative hypothesis.
2. Propose a uniformly most powerful test of level α for this problem.
3. The level of the test is set to $\alpha = 5\%$ and we proceed with 80 throws of the dice. Using a Gaussian approximation, give the number of 6 beyond which the dice is declared unfair.

5 χ^2 test

Three candidates run for an election, say A, B, C. A first pool has given the distribution $\frac{1}{2}, \frac{1}{3}, \frac{1}{6}$ for candidates A, B, C, respectively. The candidate C wants to invalidate this pool through a χ^2 test applied to the results of a second pool with n electors.

1. What is the minimum value of n beyond which a χ^2 test is legitimate?
2. Give the result of the χ^2 test at level 5% for the following pool based on $n = 60$ electors:

Candidate	A	B	C	Total
Number of votes	24	22	14	60

6 Confidence interval

A voltmeter provides an estimation of the voltage with some error distributed according to a centered Gaussian distribution with unknown variance. To measure a voltage μ (unknown), we proceed with n measurements, whose errors are supposed to be i.i.d. We denote by $X_1, \ldots, X_n$ the obtained values.

1. Give the distribution of $X_1, \ldots, X_n$.
2. Deduce a confidence interval of level $1 - \alpha$ for μ, with $\alpha \in (0, 1)$.
3. Make this interval explicit for $\alpha = 5\%$ and the following 3 measurements: $200, 210, 220$.
4. We would like to test the null hypothesis $H_0 = \{\mu = 220\}$ against the alternative hypothesis $H_1 = \{\mu \neq 220\}$. Give a test of level α, then the result of this test for the values of the previous question.

Appendix

We give quantile tables $Q(\alpha)$ for some distributions with $\alpha \in [0.5, 1[$. The values are given with precision 10^{-2}.

α	0.5	0.9	0.95	0.975	0.99
$Q(\alpha)$	0.0	1.28	1.64	1.96	2.33

Table 1: Standard normal distribution.

k	α	0.5	0.9	0.95	0.975	0.99
1	$Q(\alpha)$	0.67	1.64	1.96	2.24	2.58
2	$Q(\alpha)$	1.18	2.15	2.45	2.72	3.03
3	$Q(\alpha)$	1.54	2.50	2.80	3.06	3.37

Table 2: χ^2 distribution with k degrees of freedom ($k = 1, 2, 3$).

k	α	0.5	0.9	0.95	0.975	0.99
1	$Q(\alpha)$	0.0	3.08	6.31	12.71	31.82
2	$Q(\alpha)$	0.0	1.89	2.92	4.30	6.96
3	$Q(\alpha)$	0.0	1.64	2.35	3.18	4.54

Table 3: Student distribution with k degrees of freedom ($k = 1, 2, 3$).