Exam

November 2021
Duration : 3h

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

We recall some usual probability distributions :

Pareto distribution. A random variable X has a Pareto distribution with parameter (a, θ), with $a, \theta > 0$ if it has the density :

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

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

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

where Γ is the Gamma function. 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 $a = b = 1$, the Beta distribution is the uniform distribution over $[0, 1]$.

Loi Gamma. A random variable X has a Gamma distribution with parameter (a, λ), with $a, \lambda > 0$ 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 :

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

and if $a > 2$,

$$\mathbb{E}\left(\frac{1}{X}\right) = \frac{\lambda}{a-1}, \quad \mathrm{var}\left(\frac{1}{X}\right) = \frac{\lambda^2}{(a-1)^2(a-2)}.$$

In the particular case $a = 1$, the Gamma distribution is the exponential distribution.

1 Point estimation

A physical quantity is supposed to have a uniform distribution over $[-\theta, \theta]$, where $\theta > 0$ is some unknown parameter. We use n i.i.d. samples $X_1, \ldots, X_n$ from this distribution to estimate θ. We consider the following estimator :

$$\forall x \in \mathbb{R}^n, \quad \widehat{\theta}(x) = \max(|x_1|, \ldots, |x_n|).$$

1. Show that $\widehat{\theta}$ is the maximum likelihood estimator of θ.
2. Using the fact that the random variables $|X_1|, \ldots, |X_n|$ have the uniform distribution over $[0, \theta]$, show that $\frac{\widehat{\theta}(X)}{\theta}$ has a Beta distribution with parameter $(n, 1)$.
3. Is the estimator $\widehat{\theta}$ biased ? Calculate its quadratic risk.

In the following, we denote for any $\alpha > 0$ the estimator :

$$\forall x \in \mathbb{R}^n, \quad \widehat{\theta}^{(\alpha)}(x) = \alpha \widehat{\theta}(x).$$

4. Give the value of α for which the estimator $\theta^{(\alpha)}$ is unbiased.
5. Give the value of α minimizing the quadratic risk of $\theta^{(\alpha)}$.

Finally, we consider the method of moments.

6. Propose an estimator $\tilde{\theta}$ derived from the method of moments, applied to $\mathrm{E}(|X_1|)$.
7. Is the estimator $\tilde{\theta}$ biased ? Calculate its quadratic risk.
8. Show that this estimator is worse that the maximum likelihood estimator $\widehat{\theta}$ in terms of quadratic risk whenever $n \geq 3$.

2 Cramer-Rao bound

We consider a Gamma distribution with parameter (a, θ), where $a > 2$ is known and $\theta > 0$ is unknown. We have n i.i.d. samples $X_1, \ldots, X_n$ from that distribution.

1. Show that the maximum likelihood estimator of θ is given by :

$$\forall x \in \mathbb{R}^n, \quad \widehat{\theta}(x) = \frac{na}{\sum_{i=1}^n x_i}.$$

2. Is this estimator biased ?
3. Propose an estimator $\tilde{\theta}$ using the method of moments, applied to $\mathrm{E}\left(\frac{1}{X}\right)$.
4. Show that $\tilde{\theta}$ is unbiased and calculate its quadratic risk.
5. Are the estimators $\widehat{\theta}$ and $\tilde{\theta}$ efficient in the sense of Cramer-Rao ? Justify the answer.

3 Bayesian model

Consider the previous statistical model, with n i.i.d. samples $X_1, \ldots, X_n$ of a Gamma distribution with parameter (a, θ). We now consider a Bayesian setting where θ is itself random, with an exponential prior distribution with parameter $\lambda > 0$.

1. Show that the posterior distribution of θ given $x = (x_1, \ldots, x_n)$ has a Gamma distribution, with parameters to be specified.
2. Deduce an estimator δ of θ minimizing the Bayesian quadratic risk.
3. Show that this estimator is consistent, in the sense that it converges almost surely to θ when n goes to $+\infty$, for n i.i.d. samples of a Gamma distribution of parameter (a, θ).

4 Statistical test

The distribution of salaries in a company is represented by a Pareto distribution with parameter (a, θ), where $a > 0$ is known (this is the minimum salary) and $\theta > 0$ is unknown. We propose to test the null hypothesis $H_0 = \{\theta \leq 1\}$ against the alternative hypothesis $H_1 = \{\theta > 1\}$ from n i.i.d. samples $X_1, \dots, X_n$. For a given constant $c > 0$, we denote by δ the statistical test :

$$\forall x \in \mathbb{R}^n, \quad \delta(x) = \mathbb{1}_{\{\frac{1}{n}\sum_{i=1}^n \ln\left(\frac{x_i}{a}\right) < c\}},$$

1. Show that δ is a uniformly most powerful test of H_0 against H_1.
2. Show that the random variables $\ln\left(\frac{X_1}{a}\right), \dots, \ln\left(\frac{X_n}{a}\right)$ are i.i.d. exponential with parameter θ.
3. Assuming n large enough and using a Gaussian approximation, calculate the constant c so that the test δ is of level α. The constant c should be expressed as a function of n and a quantile of the standard normal distribution.
4. A statistical study gives for $n = 25$ samples $x_1, \dots, x_n$:

$$\frac{1}{n}\sum_{i=1}^n \ln\left(\frac{x_i}{a}\right) = 1.$$

   The level of the test is $\alpha = 5\%$. What is the result of the test ?

5. Using again a Gaussian approximation, give the power of the test for $\theta = 2$.
   Show that this power exceeds 95% for $\alpha = 5\%$ and $n = 25$.
   Some quantiles of the standard normal distribution are given in the Appendix.

5 χ^2 test

The results of a football match is represented by a random variable X taking values in $\{0, 1, 2\}$, corresponding to a draw, a home win and an away win, respectively. We want to test the null hypothesis that the distribution of X is $p = (p_0, p_1, p_2)$ with $P(X = k) = p_k$ for $k \in \{0, 1, 2\}$. We observe the results of n independent matches.

1. Propose a χ^2 test of level α to test this hypothesis.
2. Give the result of the test for the following values :
   — $p = (\frac{3}{10}, \frac{5}{10}, \frac{2}{10})$,
   — 36 draws, 60 home wins and 14 away wins,
   — $\alpha = 2\%$.

   Some quantiles of the χ^2 distribution are given in the Appendix.

6 Confidence intervals

We consider the statistical model of exercise 1 with n i.i.d. samples of a uniform distribution over $[-\theta, \theta]$, where $\theta > 0$ is some unknown parameter. We still use the notation :

$$\forall x \in \mathbb{R}^n, \quad \widehat{\theta}(x) = \max(|x_1|, \dots, |x_n|).$$

1. Propose a confidence interval at level $1-\alpha$ of the form $[\widehat{\theta}(x), \widehat{\theta}(x)(1+c)]$ for θ. The constant c will be expressed as a function of n and α.
2. Calculate c for $n = 10$ and $\alpha = 0.01$. We give $100^{0.1} \approx 1.58$.

Appendix

Standard normal distribution. The following table gives some values of the quantile function Q of the standard normal distribution, with 10^{-2} precision.

α	0.5	0.6	0.7	0.8	0.9	0.95	0.99
$Q(\alpha)$	0.0	0.25	0.52	0.84	1.28	1.64	2.33

χ^2 **distribution.** The following table gives some quantiles at 98% of the χ^2 distribution with k degrees of freedom, for different values of k, at 10^{-2} precision.

k	1	2	3	4
quantile at 98%	2.33	2.80	3.14	3.42