OS02 - Written Exam 1 - Theory of Decision - Solutions

Attention : the only documents authorized during exam are the copies of slides of OS02 and memory aids (probability, statistics, mathematics).

Problem 1 : [$\simeq 12$ points]

A log-normal distribution, denoted by $\log\mathcal{N}$, is a continuous probability distribution of a random variable whose logarithm is normally distributed. A random variable which is log-normally distributed takes only positive real values. If ζ is a random variable with a normal distribution $\mathcal{N}(\mu,\sigma^2)$, then $\xi = \exp\{\zeta\}$ has a log-normal distribution $\log\mathcal{N}(\mu,\sigma^2)$. If ξ is log-normally distributed $\log\mathcal{N}(\mu,\sigma^2)$, then $\zeta = \log(\xi)$ has a normal distribution $\mathcal{N}(\mu,\sigma^2)$:

$$\zeta = \log(\xi) \sim \mathcal{N}(\mu,\sigma^2) \quad \Leftrightarrow \quad \xi = e^{\zeta} \sim \log\mathcal{N}(\mu,\sigma^2).$$

The density $f(x;\mu,\sigma)$ of the log-normal distribution $\log\mathcal{N}(\mu,\sigma^2)$ is :

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\dfrac{(\log(x)-\mu)^2}{2\sigma^2}} \quad \text{for} \quad x>0.$$

The cumulative distribution function of $\log\mathcal{N}(\mu,\sigma^2)$ is :

$$F(x) = \Phi\left(\frac{\log(x)-\mu}{\sigma}\right), \quad x>0, \quad \text{where} \quad \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\dfrac{u^2}{2}}\, du.$$

Let $(\xi_1,\ldots,\xi_n)$ be i.i.d. random variables from the distribution $\log\mathcal{N}(\mu,\sigma^2)$, with the parameters μ and σ^2 [1]. It is assumed that the parameter $\sigma^2>0$ is known but the parameter μ is unknown. We wish to test the null hypothesis

$$\mathcal{H}_1 = \{\xi_1,\ldots,\xi_n \sim \log\mathcal{N}(\mu_1,\sigma^2)\} \text{ against the alternative hypothesis } \mathcal{H}_2 = \{\xi_1,\ldots,\xi_n \sim \log\mathcal{N}(\mu_2,\sigma^2)\}.$$

Part 1 :

1. Write down the expression of the log likelihood ratio $\log\Lambda(\xi_1,\ldots,\xi_n)$ for testing $\mathcal{H}_1$ against $\mathcal{H}_2$.
2. Define the log likelihood ratio test δ for testing $\mathcal{H}_1$ against $\mathcal{H}_2$.
3. Is the function $c \mapsto R(c) = \mathbb{P}_1(\log\Lambda(\xi_1,\ldots,\xi_n) \geq c)$ continuous ?
4. Is the test δ optimal in some sense ? If, the answer is *"yes"*, then explain what is the sense of optimality.
5. Calculate the probabilities of Type I and II error $\alpha_1(\delta)$ and $\alpha_2(\delta)$ of the test δ.

1. Attention : the mean and the variance of a log-normal law are not equal to μ and σ^2, respectively. In fact, the mean and the variance are functions of μ and σ^2.

Part 2 : Let $(\xi_1, \ldots, \xi_n)$ be i.i.d. random variables from the distribution $\log\mathcal{N}(\mu, \sigma^2)$, with the parameters μ and σ^2. It is assumed that the parameter $\sigma^2 > 0$ is known but the parameter μ is unknown. We wish to test the null hypothesis

$$\mathcal{H}_1 = \{\xi_1, \ldots, \xi_n \sim \log\mathcal{N}(\mu \le \mu_0, \sigma^2)\} \text{ against the alternative } \mathcal{H}_2 = \{\xi_1, \ldots, \xi_n \sim \log\mathcal{N}(\mu > \mu_0, \sigma^2)\}.$$

1. Show that the parameterized family $\mathcal{P}_\mu = \{\log\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}\}$, where μ is a parameter and σ^2 is a constant, admits a monotone likelihood ratio $\Lambda(x_1, \ldots, x_n) = g(T(x_1, \ldots, x_n))$ (for n observations !). Write down the expression of the statistics $T(x_1, \ldots, x_n)$. Draw (approximately) the representative curve of the function $T \mapsto g(T)$.
2. Define a UMP test δ for testing between $\mathcal{H}_1 = \{\mu \le \mu_0\}$ and $\mathcal{H}_2 = \{\mu > \mu_0\}$ in the class $\mathcal{K}_\alpha$.

Part 1. Solution :

1. (1.5 points) The log likelihood ratio is

$$\log\Lambda(\xi_1, \ldots, \xi_n) = \frac{\mu_2 - \mu_1}{\sigma^2}\sum_{i=1}^{n}\log(\xi_i) + n\frac{\mu_1^2 - \mu_2^2}{2\sigma^2}.$$

2. (1 point) The likelihood ratio test is

$$\delta(\xi_1, \ldots, \xi_n) = \begin{cases} \mathcal{H}_1 & \text{if} \quad \dfrac{\mu_2 - \mu_1}{\sigma^2}\displaystyle\sum_{i=1}^{n}\log(\xi_i) + n\dfrac{\mu_1^2 - \mu_2^2}{2\sigma^2} < h \\ \mathcal{H}_2 & \text{if} \quad \dfrac{\mu_2 - \mu_1}{\sigma^2}\displaystyle\sum_{i=1}^{n}\log(\xi_i) + n\dfrac{\mu_1^2 - \mu_2^2}{2\sigma^2} \ge h \end{cases}.$$

where the constant h is such that :

$$\mathbb{P}_1(\Lambda(\xi_1, \ldots, \xi_n) \ge h) = \alpha.$$

3. (2 points) The probability of false alarm is given by

$$\mathbb{P}_1\left(\log\Lambda(\xi_1, \ldots, \xi_n) \ge c\right) = \mathbb{P}_1\left(\frac{\mu_2 - \mu_1}{\sigma^2}\sum_{i=1}^{n}\log(\xi_i) + n\frac{\mu_1^2 - \mu_2^2}{2\sigma^2} \ge c\right).$$

As it follows from the definition of the log-normal distribution (under $\mathcal{H}_1$), $\log(\xi_i) \sim \mathcal{N}(\mu_1, \sigma^2)$. Hence, we get

$$\sum_{i=1}^{n}\log(\xi_i) \sim \mathcal{N}(n\mu_1, n\sigma^2) \text{ and } \log\Lambda(\xi_1, \ldots, \xi_n) \sim \mathcal{N}\left(n\frac{(\mu_2 - \mu_1)\mu_1}{\sigma^2} + n\frac{\mu_1^2 - \mu_2^2}{2\sigma^2}, \frac{(\mu_2 - \mu_1)^2}{\sigma^2}\right)$$

and, finally,

$$\mathbb{P}_1\left(\log\Lambda(\xi_1, \ldots, \xi_n) \ge c\right) = 1 - \Phi\left(\frac{c + n\frac{(\mu_2 - \mu_1)^2}{2\sigma^2}}{\sqrt{n}\frac{|\mu_1 - \mu_2|}{\sigma}}\right)$$

The function $x \mapsto \Phi(ax + b)$ is continuous on $]-\infty, \infty[$. Hence, the function $c \mapsto R(c) = \mathbb{P}_1(\log\Lambda(\xi_1, \ldots, \xi_n) \ge c)$ is also continuous.

4. (1 point) The likelihood ratio test $\delta(\xi_1, \ldots, \xi_n)$ is the MP test in the class $\mathcal{K}_\alpha$.

5. (3.5 points) The probabilities of Type I and II error $\alpha_1(\delta)$ and $\alpha_2(\delta)$ of the test δ are

$$\alpha_1 = \mathbb{P}_1\left(\log \Lambda(\xi_1,\ldots,\xi_n) \geq c\right) = 1 - \Phi\left(\frac{h + n\frac{(\mu_2-\mu_1)^2}{2\sigma^2}}{\sqrt{n}\frac{|\mu_1-\mu_2|}{\sigma}}\right)$$

and

$$\alpha_2 = \mathbb{P}_2\left(\log \Lambda(\xi_1,\ldots,\xi_n) < c\right) = \Phi\left(\frac{h - n\frac{(\mu_2-\mu_1)^2}{2\sigma^2}}{\sqrt{n}\frac{|\mu_1-\mu_2|}{\sigma}}\right).$$

Part 1. Solution :

1. (2 points) The parameterized family $\mathcal{P}_\mu = \{\log \mathcal{N}(\mu,\sigma^2), \mu \in \mathbb{R}\}$, where μ is a parameter and σ^2 is a constant, admits a monotone likelihood ratio $\Lambda(x_1,\ldots,x_n) = g(T(x_1,\ldots,x_n))$. Indeed, the likelihood ratio is

$$\Lambda(\xi_1,\ldots,\xi_n) = \exp\left\{\frac{\mu_2-\mu_1}{\sigma^2}\sum_{i=1}^{n}\log(\xi_i) + n\frac{\mu_1^2-\mu_2^2}{2\sigma^2}\right\} = \exp\left\{\frac{\mu_2-\mu_1}{\sigma^2}T(\xi_1,\ldots,\xi_n) + n\frac{\mu_1^2-\mu_2^2}{2\sigma^2}\right\},$$

where $T(\xi_1,\ldots,\xi_n) = \sum_{i=1}^{n}\log(\xi_i)$. The function $T \mapsto g(T)$ is monotone increasing (when $\mu_2 > \mu_1$). Hence, the parameterized family $\mathcal{P}_\mu = \{\log \mathcal{N}(\mu,\sigma^2), \mu \in \mathbb{R}\}$ admits a monotone likelihood ratio.

2. (1 point) The UMP test is defined as follows

$$\delta(\xi_1,\ldots,\xi_n) = \begin{cases} \mathcal{H}_1 & \text{if} \quad T(\xi_1,\ldots,\xi_n) < \lambda \\ \mathcal{H}_2 & \text{if} \quad T(\xi_1,\ldots,\xi_n) \geq \lambda \end{cases}$$

where the constant λ is such that :

$$\mathbb{P}_{\mu_0}(T(\xi_1,\ldots,\xi_n) \geq \lambda) = \alpha.$$

Problem 2 : [$\simeq 8$ points]

Let ξ be a sample from the distribution $\mathcal{F}$ (the sample size is $n = 1$). Denote by f its probability density. We wish to test the null hypothesis $\mathcal{H}_1 : \{f = f_1\}$ against the alternative one $\mathcal{H}_2 : \{f = f_2\}$. The densities are defined as follows

$$f_1(x) = \begin{cases} x + 0.5 & \text{if} \quad x \in [0,1], \\ 0 & \text{if} \quad x \notin [0,1] \end{cases} \qquad f_2(x) = \begin{cases} -x + 1.5 & \text{if} \quad x \in [0,1], \\ 0 & \text{if} \quad x \notin [0,1] \end{cases}$$

1. Write down the likelihood ratio $\Lambda(\xi)$ and the likelihood ratio test $\delta(\xi)$ for testing between $\mathcal{H}_1$ and $\mathcal{H}_2$.
2. Calculate the probabilities $\alpha_1(\delta) = \mathbb{P}_1(\delta = \mathcal{H}_2), \alpha_2(\delta) = \mathbb{P}_2(\delta = \mathcal{H}_1)$ as functions of the test threshold h. Is the function $h \mapsto R(h) = \mathbb{P}_1(\Lambda(\xi) \geq h)$ continuous over the left-closed interval $[0, \infty[$?

   Hint. *Study the monotonicity of the function* $y = \dfrac{-x+1.5}{x+0.5}$ *defined on the interval* $[0,1]$ *and rewrite the inequalities* $\{\Lambda(\xi) \geq h\}$ *and* $\{\Lambda(\xi) < h\}$ *directly for the observation* ξ.
3. Is the test $\delta(\xi)$ optimal in the class $\mathcal{K}_\alpha$ in some sense ? If the answer is *"yes"* then explain what is the sense of optimality.
4. Let us consider now the Bayesian approach. It is assumed that the hypotheses $\mathcal{H}_1$ and $\mathcal{H}_2$ are random events and that $\mathbb{P}(\mathcal{H}_1) = q$, $\mathbb{P}(\mathcal{H}_2) = 1 - q$, $Q = \{q, 1-q\}$. Determine the Bayes test $\overline{\delta}$ which minimizes the average error probability

   $$\overline{\alpha}_Q(\delta) = q\alpha_1(\delta) + (1-q)\alpha_2(\delta).$$
5. Calculate the average error probability $\overline{\alpha}_Q(\overline{\delta})$ of the Bayes test for $q = \frac{1}{2}$.

Solution :

1. (1 point) The likelihood ratio is

$$\Lambda(\xi) = \frac{f_2(\xi)}{f_1(\xi)} = \frac{-\xi + 1.5}{\xi + 0.5}.$$

The likelihood ratio test is

$$\delta(\xi) = \begin{cases} \mathcal{H}_1 & \text{if} \quad \Lambda(\xi) = \dfrac{-\xi + 1.5}{\xi + 0.5} < h \\ \mathcal{H}_2 & \text{if} \quad \Lambda(\xi) = \dfrac{-\xi + 1.5}{\xi + 0.5} \geq h \end{cases}.$$

2. (3 points) Because the function $x \mapsto \dfrac{-x + 1.5}{x + 0.5}$ is monotone decreasing on the interval $[0, 1]$, the events $\{\Lambda(\xi) \geq h\}$ and $\{\Lambda(\xi) < h\}$ can be re-written as follows

$$\{\Lambda(\xi) \geq h\} \Leftrightarrow \{\xi \leq x_0\} \text{ and } \{\Lambda(\xi) < h\} \Leftrightarrow \{\xi > x_0\},$$

where $x_0 = \dfrac{-0.5h + 1.5}{h + 1}$. Hence,

$$\begin{aligned} \alpha_1 &= \mathbb{P}_1(\delta = \mathcal{H}_2) = \mathbb{P}_1\left(\frac{-\xi + 1.5}{\xi + 0.5} \geq h\right) = \mathbb{P}_1(\xi \leq x_0) \\ &= \int_0^{x_0} (x + 0.5)dx = \frac{1}{2}\left(\frac{-0.5h + 1.5}{h + 1}\right)^2 + \frac{1}{2}\left(\frac{-0.5h + 1.5}{h + 1}\right). \end{aligned}$$

$$\begin{aligned} \alpha_2 &= \mathbb{P}_2(\delta = \mathcal{H}_1) = \mathbb{P}_2\left(\frac{-\xi + 1.5}{\xi + 0.5} < h\right) = \mathbb{P}_2(\xi > x_0) \\ &= \int_{x_0}^{1} (-x + 1.5)dx = 1 + \frac{1}{2}\left(\frac{-0.5h + 1.5}{h + 1}\right)^2 - \frac{3}{2}\left(\frac{-0.5h + 1.5}{h + 1}\right). \end{aligned}$$

The function $x \mapsto \dfrac{-0.5x + 1.5}{x + 1}$ is continuous on the interval $[1/3, 3]$ as a polynomial fraction with $x + 1 > 0$. The function $h \mapsto R(h) = \mathbb{P}_1(\Lambda(\xi) \geq h)$, where

$$R(h) = \begin{cases} 1 & \text{if} \quad 0 \leq h < \frac{1}{3} \\ \frac{1}{2}\left(\frac{-0.5h+1.5}{h+1}\right)^2 + \frac{1}{2}\left(\frac{-0.5h+1.5}{h+1}\right) & \text{if} \quad \frac{1}{3} \leq h \leq 3 \\ 0 & \text{if} \quad h > 3 \end{cases}.$$

is continuous over the left-closed interval $[0, \infty[$.

3. (1 points) The likelihood ratio test $\delta(\xi)$ is the MP test.

4. (1 points) The Bayes test is

$$\delta(\xi) = \begin{cases} \mathcal{H}_1 & \text{if} \quad \Lambda(\xi) = \dfrac{-\xi + 1.5}{\xi + 0.5} < \dfrac{q}{1 - q} = 1 \\ \mathcal{H}_2 & \text{if} \quad \Lambda(\xi) = \dfrac{-\xi + 1.5}{\xi + 0.5} \geq \dfrac{q}{1 - q} = 1 \end{cases}.$$

5. (2 points)

$$\overline{\alpha}_Q(\delta) = q\alpha_1(\delta) + (1 - q)\alpha_2(\delta) = \frac{1}{2}\left[\frac{1}{2}\left(\frac{1}{2}\right)^2 + \frac{1}{2}\left(\frac{1}{2}\right)\right] + \frac{1}{2}\left[1 + \frac{1}{2}\left(\frac{1}{2}\right)^2 - \frac{3}{2}\left(\frac{1}{2}\right)\right] = 0.375.$$