University of Ljubljana, School of Economics

Quantitative finance and actuarial science

Probability and statistics

Written examination

January 26^{th}, 2023

Name and surname: ____________________ ID: | | | | | | | | |

Instructions

Read the problems carefull before staring your work. There are four problems. You can have a sheet with formulae and a mathematical handbook. Write your solutions on the paper provided. You have two hours.

Problem	a.	b.	c.	d.	
1.			•	•	
2.					
3.			•	•	
4.					
Total					

1. (25) A population of size N is divided into K groups of equal size $M = N/K$. A sample is selected in such a way that k groups are selected by simple random sampling and then all the units in the selected groups are selected.

a. (10) Show that the sample average $\bar{Y}$ is an unbiased estimate of the population mean.

   Solution: Let μ_i be the population mean in the i-th group. In the sampling procedure described we are choosing a simple random sample of groups and we observe μ_i for this group. The estimator $\bar{Y}$ is just a sample average of the μ_i selected. The expectation is therefore the average of all μ_is which is μ.

b. (15) Let μ_i be the population mean in group i for $i = 1, 2, \ldots, K$ and let μ be the population mean. Define

$$\sigma_b^2 = \frac{1}{K} \sum_{i=1}^{K} (\mu_i - \mu)^2 .$$

   Show that

$$\operatorname{se}\left(\bar{Y}\right) = \frac{\sigma_b}{\sqrt{k}} \cdot \sqrt{\frac{K-k}{K-1}} .$$

   Solution: Think of groups as units selected and to each group assign the value μ_i. The formula is then the formula for the standard error of such a sample average. But $\bar{Y}$ is equal to this sample average.

2. (25) Assume the observed values $x_1, x_2, \ldots, x_n$ were generated as random variables $X_1, X_2, \ldots, X_n$ with density

$$f(x) = \frac{1}{\sqrt{2\pi x^3}}\, e^{-\frac{(1-\mu x)^2}{2x}}$$

for $x, \mu > 0$.

a. (5) Find the maximum likelihood estimate of μ.

Solution: the log-likelihood function is

$$\ell = \frac{n}{2}\log 2\pi - \frac{3}{2}\sum_{k=1}^{n}\log x_k - \sum_{k=1}^{n}\frac{(1-\mu x_k)^2}{2x_k}\,.$$

Taking derivatives with respect to μ gives

$$\sum_{k=1}^{n}(1-\mu x_k) = 0\,.$$

The estimate os μ is

$$\hat{\mu} = \frac{n}{x_1 + x_2 + \cdots + x_n} = \frac{1}{\bar{x}}\,.$$

b. (5) Can you fix the maximum likelihood estimator to be unbiased? Assume as known:

- The density of $X_1 + \cdots + X_n$ is

$$f_n(x) = \frac{n}{\sqrt{2\pi x^3}}\, e^{-\frac{(n-\mu x)^2}{2x}}$$

  for $x > 0$.

- Assume as known that for $a, b > 0$ we have

$$\int_0^\infty x^{-5/2} e^{-ax-\frac{b}{x}}\,\mathrm{d}x = \frac{\sqrt{\pi}\left(1 + 2\sqrt{ab}\right)}{2b^{3/2}}\, e^{-2\sqrt{ab}}\,.$$

Solution: let X have density $f_n(x)$. We compute

$$\begin{aligned}
E\left(\frac{n}{X}\right) &= n\int_0^\infty \frac{1}{x}\, f_n(x)\,\mathrm{d}x \\
&= n^2 \frac{e^{n\mu}}{\sqrt{2\pi}} \int_0^\infty x^{-5/2} e^{-\frac{\mu^2}{2}x - \frac{n^2}{2x}}\,\mathrm{d}x \\
&= n^2 \frac{e^{n\mu}}{\sqrt{2\pi}} \sqrt{2\pi}\,\frac{1+n\mu}{n^3}\, e^{-n\mu} \\
&= \mu + \frac{1}{n}\,.
\end{aligned}$$

An unbiased estimator is given by

$$\tilde{\mu} = \frac{1}{\bar{X}} - \frac{1}{n}\,.$$

c. (10) Compute the variance of the maximum likelihood estimator of μ. Assume as known that for $a, b > 0$ we have

$$\int_0^\infty x^{-7/2} e^{-ax-\frac{b}{x}}\,\mathrm{d}x = \frac{\sqrt{\pi}\big(3 + 6\sqrt{ab} + 4ab\big)}{4b^{5/2}}\, e^{-2\sqrt{ab}}\,.$$

Solution: for X with density $f_n(x)$ we compute

$$\begin{aligned} E\left(\frac{n^2}{X^2}\right) &= \int_0^\infty \frac{n^2}{x^2}\, f_n(x)\,\mathrm{d}x \\ &= n^3 \frac{e^{n\mu}}{\sqrt{2\pi}} \int_0^\infty x^{-7/2} e^{-\frac{\mu^2}{2}x - \frac{n^2}{2x}}\,\mathrm{d}x \\ &= n^3 \frac{e^{n\mu}}{\sqrt{2\pi}} \frac{\sqrt{2\pi}(3 + 3n\mu + n^2\mu^2)}{n^5}\, e^{-n\mu} \\ &= \frac{3}{n^2} + \frac{3\mu}{n} + \mu^2\,. \end{aligned}$$

The variance is

$$\mathrm{var}(\hat{\mu}) = E(\hat{\mu}^2) - \big(E(\hat{\mu})\big)^2 = \frac{\mu}{n} + \frac{2}{n^2}\,.$$

d. (5) What approximation the the standard error of the maximum likelihood estimator do we get if we use the Fisher information? Assume as known that

$$\int_0^\infty x^{-1/2} e^{-ax-\frac{b}{x}}\,\mathrm{d}x = \frac{\sqrt{\pi}}{\sqrt{a}}\, e^{-2\sqrt{ab}}\,.$$

Solution: taking derivatives for $n = 1$ we get

$$\ell'' = -x\,.$$

It follows that

$$\begin{aligned} I(\mu) &= E(X) \\ &= \frac{e^{\mu}}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{x}}\, e^{-\frac{\mu^2 x}{2} - \frac{1}{2x}}\,\mathrm{d}x \\ &= \frac{e^{\mu}}{\sqrt{2\pi}} \cdot \sqrt{2\pi\mu e^{\mu}} \\ &= \frac{1}{\mu}\,. \end{aligned}$$

Fisher's approximation for the variance is

$$\frac{\mu}{n}\,.$$

3. (25) Suppose the observed values are pairs $(x_1, y_1), \ldots, (x_n, y_n)$. Assume the pairs are an i.i.d. sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from the density

$$f(x,y) = e^{-x} \cdot \frac{1}{\sigma\sqrt{2\pi x}}\, e^{-\frac{(y-\theta x)^2}{2\sigma^2 x}}$$

for $x > 0$ and $-\infty < y < \infty$ and $\sigma^2 > 0$. The testing problem is

$$H_0\colon \theta = 0 \quad \text{proti} \quad H_1\colon \theta \neq 0\,.$$

a. (15) Find the Wilks's test statistics for the testing problem.

Solution: the log-likelihood function is

$$\ell\left(\theta, \sigma|\mathbf{x}, \mathbf{y}\right) = -\frac{n}{2}\log 2\pi - n\log\sigma - \frac{1}{2}\sum_{k=1}^{n}\Big[-\log x_k - \frac{(y_k-\theta x_k)^2}{\sigma^2 x_k}\Big]\,.$$

Computing partial derivatives, we get

$$\frac{\partial \ell}{\partial \theta} = \sum_{k=1}^{n}\frac{(y_k-\theta x_k)}{\sigma^2}$$

$$\frac{\partial \ell}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{k=1}^{n}\frac{(y_k-\theta x_k)^2}{\sigma^3 x_k}$$

Equating with 0*, we get*

$$\hat{\theta} = \frac{\sum_{k=1}^n y_k}{\sum_{k=1}^n x_k}$$

and the second equation gives

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n}\frac{(y_k-\hat{\theta}x_k)^2}{x_k}\,.$$

When we maximize only over σ^2*, taking derivatives gives*

$$\frac{\partial \ell}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{k=1}^{n}\frac{y_k^2}{\sigma^3 x_k}\,.$$

It follows

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n}\frac{y_k^2}{x_k}\,.$$

After some calculations we get

$$\lambda = -2n\log\hat{\sigma} + 2n\log\tilde{\sigma}\,.$$

b. (10) Assume that H_0 is rejected when $\lambda > \lambda_\alpha$ where λ_α is chosen in such a way that the size of the test is $\alpha \in (0, 1)$. Give an approximate value for λ_α?

Solution: Wilks's theorem gives the rejection region as $\{\lambda > \lambda_\alpha\}$ *where* λ_α *is the* $(1 - \alpha)$*-th percentile of the* $\chi^2(1)$ *distribution.*

4. (25) Assume the regression equations are

$$\begin{aligned} Y_{k1} &= \alpha + \beta x_{k1} + \epsilon_{k1} \\ Y_{k2} &= \alpha + \beta x_{k2} + \epsilon_{k2} \end{aligned}$$

for $k = 1, 2, \ldots, n$. The error terms satisfy the assumptions that

$$E(\epsilon_{k1}) = E(\epsilon_{k2}) = 0$$

$$\mathrm{var}(\epsilon_{k1}) = \mathrm{var}(\epsilon_{k2}) = 2\sigma^2$$

for $k = 1, 2, \ldots, n$, and

$$\mathrm{cov}(\epsilon_{k1}, \epsilon_{k2}) = \sigma^2$$

for $k \neq l$. Assume that $\sum_{k=1}^{n}(x_{k1} + x_{k2}) = 0$. The vectors $(\epsilon_{k1}, \epsilon_{k2}), \ldots, (\epsilon_{n1}, \epsilon_{n2})$ are independent.

a. (5) Show that

$$\mathrm{cov}((3+\sqrt{3})Y_{k1} + (-3+\sqrt{3})Y_{k2}, (-3+\sqrt{3})Y_{k1} + (3+\sqrt{3})Y_{k2}) = 0$$

for $k = 1, 2, \ldots, n$.

Solution: compute

$$\begin{aligned} &\mathrm{cov}((3+\sqrt{3})Y_{k1} + (-3+\sqrt{3})Y_{k2}, (-3+\sqrt{3})Y_{k1} + (3+\sqrt{3})Y_{k2}) \\ &= \sigma^2 \left(-12 - 12 + (3+\sqrt{3})^2 + (-3+\sqrt{3})^2\right) \\ &= 0\,. \end{aligned}$$

b. (5) Compute

$$\mathrm{var}\left((3+\sqrt{3})Y_{k1} + (-3+\sqrt{3})Y_{k2}\right)$$

and

$$\mathrm{var}\left((-3+\sqrt{3})Y_{k1} + (3+\sqrt{3})Y_{k2}\right)\,.$$

Solution: both variances are the same by symmetry. For the first we compute

$$\begin{aligned} &\mathrm{var}\left((-3+\sqrt{3})Y_{k1} + (3+\sqrt{3})Y_{k2}\right) \\ &= (-3+\sqrt{3})^2\mathrm{var}(Y_{k1}) + (3+\sqrt{3})^2\mathrm{var}(Y_{k1}) \\ &\qquad +2(-3+\sqrt{3})(3+\sqrt{3})\mathrm{cov}(Y_{k1}, Y_{k2}) \\ &= \sigma^2(48 - 12) \\ &= 36\sigma^2\,. \end{aligned}$$

c. (10) Compute the best unbiased linear estimator $\hat{\alpha}$ of α as explicitly as possible.

Solution: we replace the pair (y_{k1}, y_{k2}) *by the pair*

$$(\tilde{y}_{k1}, \tilde{y}_{k2}) = \big((3+\sqrt{3})y_{k1} + (-3+\sqrt{3})y_{k2}, (-3+\sqrt{3})y_{k1} + (3+\sqrt{3})y_{k2}\big)$$

and the pair (x_{k1}, x_{k2}) *by*

$$(\tilde{x}_{k1}, \tilde{x}_{k2}) = \big((3+\sqrt{3})x_{k1} + (-3+\sqrt{3})x_{k2}, (-3+\sqrt{3})x_{k1} + (3+\sqrt{3})x_{k2}\big)\,.$$

The regression model is transformed into

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}}$$

where

$$\tilde{\mathbf{X}} = \begin{pmatrix} 2\sqrt{3} & \tilde{x}_{11} \\ 2\sqrt{3} & \tilde{x}_{12} \\ \vdots & \vdots \\ 2\sqrt{3} & \tilde{x}_{n1} \\ 2\sqrt{3} & \tilde{x}_{n2} \end{pmatrix}$$

The transformed model satisfies the assumptions of the Gauss-Markov theorem so the best unbiased estimator is

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \left(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}^T\tilde{\mathbf{Y}}\,.$$

The assumptions imply that

$$\tilde{\mathbf{X}}^T\tilde{\mathbf{X}} = \begin{pmatrix} 4\sqrt{3}n & 0 \\ 0 & \sum_{k=1}^n(\tilde{x}_{k1}^2 + \tilde{x}_{k2}^2) \end{pmatrix}\,.$$

Further we get

$$\tilde{\mathbf{X}}^T\tilde{\mathbf{Y}} = \begin{pmatrix} 2\sqrt{3}\sum_{k=1}^n(\tilde{y}_{k1} + \tilde{y}_{k2}) \\ \sum_{k=1}^n(\tilde{x}_{k1}\tilde{y}_{k1}^2 + \tilde{x}_{k2}\tilde{y}_{k2}^2) \end{pmatrix}\,.$$

It follows that

$$\hat{\alpha} = \frac{1}{2n}\sum_{k=1}^n(\tilde{y}_{k1} + \tilde{y}_{k2}) = \bar{y}\,.$$

d. (5) Compute the standard error of $\hat{\alpha}$.

Solution: we have

$$\begin{aligned} \mathrm{var}(\hat{\alpha}) &= \frac{n}{4n^2}(2\sigma^2 + 2\sigma^2 + 2\sigma^2) \\ &= \frac{3\sigma^2}{2n}\,. \end{aligned}$$