

SCHOOL OF ECONOMICS
DOCTORAL PROGRAMME IN ECONOMICS
PROBABILITY AND STATISTICS
WRITTEN EXAMINATION
JANUARY 27th, 2022

NAME AND SURNAME: _____ ID:

--	--	--	--	--	--	--	--

INSTRUCTIONS

Read the problems carefully before starting 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) Products are delivered in batches of size M . For quality control, n batches are selected by simple random sampling out of N batches delivered. In each selected batch a simple random sample of size m is selected. The percentage of defective items is to be estimated. The sampling procedures in selected batches are independent and independent of the selection procedures of batches.

- a. (10) Is the sample percentage of defective items an unbiased estimator of the population percentage of defective items. Explain.

Solution: define

$$I_k = \begin{cases} 1 & \text{if batch } k \text{ is selected} \\ 0 & \text{else} \end{cases}$$

for $k = 1, 2, \dots, N$. Let \bar{Y}_k be the sample proportion estimator based on a simple random sample of size m for $k = 1, 2, \dots, N$. Let

$$\bar{Y} = c \sum_{k=1}^N \bar{Y}_k I_k.$$

Computing expectations we get

$$E(\bar{Y}) = c \sum_{k=1}^N E(\bar{Y}_k) E(I_k) = c \sum_{k=1}^N p_k \cdot \frac{n}{N}$$

where p_k is the proportion of defective items in batch k . On the other hand we have

$$p = \frac{1}{N} \sum_{k=1}^N p_k.$$

Letting $c = 1/n$ makes \bar{Y} an unbiased estimate of the overall proportion p .

- b. (15) Denote the proportion of defective items in the k -th batch by p_k for $k = 1, 2, \dots, N$. Express the standard error of the sample percentage with these quantities.

Solution: compute

$$\begin{aligned} \text{var}(\bar{Y}) &= \text{var} \left(\frac{1}{n} \sum_{k=1}^N \bar{Y}_k I_k \right) \\ &= \frac{1}{n^2} \left(\sum_{k=1}^N \text{var}(\bar{Y}_k I_k) + 2 \sum_{k < l} \text{cov}(\bar{Y}_k I_k, \bar{Y}_l I_l) \right). \end{aligned}$$

From the text it follows that all the \bar{Y}_k are independent of I_1, \dots, I_N . We have

$$\text{var}(\bar{Y}_k I_k) = E(\bar{Y}_k^2 I_k^2) - E(\bar{Y}_k I_k)^2.$$

By independence it follows

$$E(\bar{Y}_k^2 I_k^2) = E(\bar{Y}_k^2)E(I_k).$$

From the known formula

$$\text{var}(\bar{Y}_k) = \frac{p_k(1-p_k)}{n} \cdot \frac{M-m}{M-1}$$

we have

$$E(\bar{Y}_k^2) = \frac{p_k(1-p_k)}{n} \cdot \frac{M-m}{M-1} + p_k^2.$$

Note that $E(I_k) = E(I_k^2) = n/N$. Furthermore, we have for $k < l$

$$\text{cov}(\bar{Y}_k I_k, \bar{Y}_l I_l) = E(\bar{Y}_k \bar{Y}_l I_k I_l) - E(\bar{Y}_k I_k)E(\bar{Y}_l I_l),$$

and by independence

$$\text{cov}(\bar{Y}_k I_k, \bar{Y}_l I_l) = p_k p_l E(I_k I_l) - p_k p_l \cdot \frac{n^2}{N^2}.$$

From simple random sampling we know

$$E(I_k I_l) = -\frac{n(N-n)}{N^2(N-1)}.$$

The formula for covariance simplifies to

$$\text{cov}(\bar{Y}_k I_k, \bar{Y}_l I_l) = -p_k p_l \cdot \frac{n(N-n)}{N^2(N-1)}.$$

Assembling all the quantities gives the variance.

2. (20) The observed values are pairs (x_i, y_i) for $i = 1, 2, \dots, n$. Assume that the pairs are an i.i.d. sample from the distribution given by the density

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{(1+\beta^2)x^2 - 2\beta xy + \alpha y^2}{2}}.$$

for $\alpha, \beta > 0$.

a. (10) Find the maximum likelihood estimates for the parameters α and β .

Solution: the log-likelihood function is

$$\ell(\alpha, \beta | \mathbf{x}, \mathbf{y}) = n \log(2\pi) - \frac{1}{2} \left(\frac{(1 + \beta^2)}{\alpha} \sum_{i=1}^n x_i^2 - 2\beta \sum_{i=1}^n x_i y_i + \alpha \sum_{i=1}^n y_i^2 \right).$$

Denote

$$m_{xx} = \frac{1}{n} \sum_{k=1}^n x_k^2, \quad m_{xy} = \frac{1}{n} \sum_{k=1}^n x_k y_k, \quad m_{yy} = \frac{1}{n} \sum_{k=1}^n y_k^2,$$

and rewrite

$$\ell(\alpha, \beta | \mathbf{x}, \mathbf{y}) = n \left(-\log(2\pi) - \frac{(1 + \beta^2) m_{xx}}{2\alpha} + \beta m_{xy} - \frac{\alpha m_{yy}}{2} \right).$$

Taking derivatives we get

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{2} \left(\frac{(1 + \beta^2) m_{xx}}{\alpha^2} - m_{yy} \right), \quad \frac{\partial \ell}{\partial \beta} = n \left(-\frac{\beta m_{xx}}{\alpha} + m_{xy} \right).$$

Equating the derivatives to zero, we get

$$\hat{\alpha} = \frac{m_{xx}}{\sqrt{m_{xx} m_{yy} - m_{xy}^2}}, \quad \hat{\beta} = \frac{m_{xy}}{\sqrt{m_{xx} m_{yy} - m_{xy}^2}}.$$

b. (5) Find the density of X .

Hint: note that you are integrating one of the normal densities.

Solution: we need to compute the marginal density. Integrating we get

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{(1+\beta^2)x^2 - 2\beta xy + \alpha y^2}{2}} dy \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{\alpha(y - \frac{\beta}{\alpha}x)^2 + \frac{x^2}{\alpha}}{2}} dy \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\alpha}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{\alpha}{2}(y - \frac{\beta}{\alpha}x)^2} dy \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\alpha}} e^{-\frac{x^2}{2\alpha}}. \end{aligned}$$

It follows that $X \sim N(0, \alpha)$ and so $E(X^2) = \alpha$.

- c. (10) Find the approximate standard errors for the maximum likelihood estimators of α and β .

Solution: let $n = 1$. In this case, the second derivatives are

$$\frac{\partial^2 \ell}{\partial \alpha^2} = -\frac{(1 + \beta^2)x^2}{\alpha^3}, \quad \frac{\partial^2 \ell}{\partial \alpha \partial \beta} = \frac{\beta x^2}{\alpha^2}, \quad \frac{\partial^2 \ell}{\partial \beta^2} = -\frac{x^2}{\alpha}.$$

Replace x by X and take expectations. Here we need $E(X^2) = \alpha$. The Fisher information matrix is

$$I(\alpha, \beta) = \begin{pmatrix} \frac{1+\beta^2}{\alpha^2} & -\frac{\beta}{\alpha} \\ -\frac{\beta}{\alpha} & 1 \end{pmatrix}.$$

Inverting we get

$$I^{-1}(\alpha, \beta) = \begin{pmatrix} \alpha^2 & \alpha\beta \\ \alpha\beta & 1 + \beta^2 \end{pmatrix}.$$

The approximate standard errors are

$$\text{se}(\hat{\alpha}) = \frac{\alpha}{\sqrt{n}} \quad \text{and} \quad \text{se}(\hat{\beta}) = \frac{\sqrt{1 + \beta^2}}{\sqrt{n}}.$$

3. (25) Assume that your observations are pairs $(x_1, y_1), \dots, (x_n, y_n)$. Assume the pairs are an i.i.d. sample from the density

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

for $\sigma > 0$, $x > 0$, $-\infty < y < \infty$. We would like to test the hypothesis

$$H_0: \theta = 0 \quad \text{versus} \quad H_1: \theta \neq 0.$$

a. (10) Find the maximum likelihood estimates for θ and σ .

Solution: the log-likelihood function is

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

Take partial derivatives to get

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

Set the partial derivatives to 0. From the first equation we have

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

and from the second

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

b. (15) Find the likelihood ratio statistic for testing the above hypothesis. What is the approximate distribution of the test statistic under H_0 ?

Solution: if $\theta = 0$ the log-likelihood functions attains its maximum for

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

It follows that

$$\lambda = -n \log \left(1 - \frac{(\sum_{k=1}^n y_k)^2}{\sum_{k=1}^n x_k \sum_{k=1}^n \frac{y_k^2}{x_k}} \right).$$

The approximate distribution of λ is $\chi^2(1)$.

4. (25) Assume the regression model

$$Y_k = \beta x_k + \epsilon_k$$

for $k = 1, 2, \dots, n$ where $\epsilon_1, \dots, \epsilon_n$ are uncorrelated, $E(\epsilon_k) = 0$ and $\text{var}(\epsilon_k) = \sigma^2$ for $k = 1, 2, \dots, n$. Assume that $x_k > 0$ for all $k = 1, 2, \dots, n$. Consider the following linear estimators of β :

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{k=1}^n x_k Y_k}{\sum_{k=1}^n x_k^2} \\ \hat{\beta}_2 &= \frac{1}{n} \sum_{k=1}^n \frac{Y_k}{x_k} \\ \hat{\beta}_3 &= \frac{\sum_{k=1}^n Y_k}{\sum_{k=1}^n x_k}\end{aligned}$$

a. (5) Are all estimators unbiased?

Solution: since $E(Y_k) = \beta x_k$ for all $k = 1, 2, \dots, n$ all the estimators are unbiased.

b. (10) Which of the estimators has the smallest standard error? Justify your answer.

Solution: all the estimators are unbiased. Gauss-Markov tells us that the best estimator is the one given by least squares and that is $\hat{\beta}_1$.

c. (10) Write down the standard errors for all three estimators.

Solution: we first compute the theoretical variances. Since Y_1, \dots, Y_n are uncorrelated we have

$$\begin{aligned}\text{var}(\hat{\beta}_1) &= \frac{\sigma^2}{\sum_{k=1}^n x_k^2} \\ \text{var}(\hat{\beta}_2) &= \frac{\sigma^2 \sum_{k=1}^n x_k^{-2}}{n^2} \\ \text{var}(\hat{\beta}_3) &= \frac{n\sigma^2}{(\sum_{k=1}^n x_k)^2}.\end{aligned}$$

We need an unbiased estimate of σ^2 . Theoretically, we have that

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{k=1}^n (Y_k - \hat{\beta} x_k)^2$$

is an unbiased estimator σ^2 . This gives us an unbiased estimator of σ^2 .