UNIVERSITY OF LJUBLJANA, SCHOOL OF ECONOMICS

QUANTITATIVE FINANCE AND ACTUARIAL SCIENCE

PROBABILITY AND STATISTICS

WRITTEN EXAMINATION

SEPTEMBER 2nd, 2024

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.	Total
1.				•	
2.				•	
3.			•	•	
4.				•	
Total					

1. (25) Suppose the population is stratified into K strata of sizes $N_1, \ldots, N_K$. Denote by μ_k the population mean in stratum k and by σ_k^2 the population variance in stratum k for $k = 1, 2, \ldots, K$. Let μ be the population mean for the whole population and σ^2 the population variance for the whole population. Suppose a stratified sample is taken with sample sizes in each stratum equal to $n_1, n_2, \ldots, n_K$. Let $\bar{X}_k$ be the sample mean in stratum k and let

$$\bar{X} = \sum_{k=1}^{K} \frac{N_k}{N} \bar{X}_k = \sum_{k=1}^{K} w_k \bar{X}_k .$$

a. (5) Compute $E\left[\left(\bar{X}_k - \bar{X}\right)^2\right]$.

Solution: we compute

$$\begin{aligned} E\left[\left(\bar{X}_k - \bar{X}\right)^2\right] &= \mathrm{var}\left(\bar{X}_k - \bar{X}\right) + \left(E\left(\bar{X}_k - \bar{X}\right)\right)^2 \\ &= \mathrm{var}(\bar{X}_k) + \mathrm{var}(\bar{X}) - 2\mathrm{cov}(\bar{X}_k, \bar{X}) + (\mu_k - \mu)^2 \\ &= \frac{\sigma_k^2}{n_k} \cdot \frac{N_k - n_k}{N_k - 1} + \sum_{i=1}^{K} w_i^2 \cdot \frac{\sigma_i^2}{n_i} \cdot \frac{N_i - n_i}{N_i - 1} \\ &\qquad -2w_k \cdot \frac{\sigma_k^2}{n_k} \cdot \frac{N_k - n_k}{N_k - 1} + (\mu_k - \mu)^2 . \end{aligned}$$

b. (10) Suggest an unbiased estimator for the quantity

$$\gamma^2 = \sum_{k=1}^{K} w_k (\mu_k - \mu)^2 .$$

Explain why the suggested estimator is unbiased.

Solution: since we have unbiased estimators for σ_k^2 the quantity

$$\hat{\gamma}_k^2 = \left(\bar{X}_k - \bar{X}\right)^2 - \frac{\hat{\sigma}_k^2}{n_k} \cdot \frac{N_k - n_k}{N_k - 1} - \sum_{i=1}^{K} w_i^2 \cdot \frac{\hat{\sigma}_i^2}{n_i} \cdot \frac{N_i - n_i}{N_i - 1} + 2w_k \cdot \frac{\hat{\sigma}_k^2}{n_k} \cdot \frac{N_k - n_k}{N_k - 1}$$

is an unbiased estimator of $(\mu_k - \mu)^2$. Multiplying γ_k^2 by w_k and summing over k we get an unbiased estimator of γ^2.

c. (10) Suggest an unbiased estimator of the population variance σ^2. Explain why your estimator is unbiased.

Hint: check that

$$\sigma^2 = \sum_{k=1}^{K} w_k \sigma_k^2 + \sum_{k=1}^{K} w_k (\mu_k - \mu)^2 .$$

Solution: we write

$$\sigma^2 = \sum_{k=1}^{K} w_k \sigma_k^2 + \gamma^2 \,.$$

Since both terms on the right can be estimated in an unbiased way we have that

$$\hat{\sigma}^2 = \sum_{k=1}^{K} w_k \hat{\sigma}_k^2 + \hat{\gamma}^2$$

is an unbiased estimator of $\hat{\sigma}^2$.

2. (25) Assume the sample values $x_1, x_2, \ldots, x_n$ are in independent identically distributed sample from the gamma distribution with parameters $a = 2$ and λ. The density of the distribution is

$$f(x) = \lambda^2 x e^{-\lambda x}$$

for $x > 0$.

a. (5) Find explicitly the maximum likelihood estimator for the parameter λ.

Solution: the log-likelihood function is

$$\ell(\lambda|\mathbf{x}) = 2n \log \lambda + \sum_{k=1}^{n} \log x_k - \lambda \sum_{k=1}^{n} x_k \,.$$

Taking derivatives and equation to 0 *we get*

$$\hat{\lambda} = \frac{2n}{\sum_{k=1}^{n} x_k} \,.$$

b. (10) Fix the maximum likelihood estimator so that it will be unbiased.

Hint: if U and V are independent with $U \sim \Gamma(a, \lambda)$ and $V \sim \Gamma(b, \lambda)$ then $U + V \sim \Gamma(a + b, \lambda)$.

Solution: following the hint, we have$\sum_{k=1}^{n} X_k \sim \Gamma(2n, \lambda)$*. We compute*

$$\begin{aligned} E(\hat{\lambda}) &= E\left(\frac{2n}{\sum_{k=1}^{n} X_k}\right) \\ &= 2n \frac{\lambda^{2n}}{\Gamma(2n)} \int_0^\infty \frac{1}{x} \cdot x^{2n-1} e^{-\lambda x} \, \mathrm{d}x \\ &= 2n \frac{\lambda^{2n}}{\Gamma(2n)} \cdot \frac{\Gamma(2n-1)}{\lambda^{2n-1}} \\ &= \frac{2n\lambda}{2n-1} \\ &= \frac{2n}{2n-1} \lambda \,. \end{aligned}$$

An unbiased estimator is

$$\tilde{\lambda} = \frac{2n-1}{2n} \hat{\lambda} = \frac{2n-1}{\sum_{k=1}^{n} X_k} \,.$$

c. (10) Using Fisher information find the approximate standard error for the maximum likelihood estimator.

Solution: compute for $n = 1$*.*

$$\ell'' = -\frac{2}{\lambda^2}\,.$$

The approximate standard error is

$$\mathrm{se}(\hat{\lambda}) = \frac{\lambda}{\sqrt{2n}}\,.$$

3. (25) Assume that your observations are pairs $(x_1, y_1), \ldots, (x_n, y_n)$. 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

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

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

Solution: the log-likelihood function is

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

Take partial derivatives to get

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

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

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

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

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

4. (25) Suppose that we have the regression model

$$\begin{aligned} Y_{i1} &= \alpha + \beta x_{i1} + \epsilon_i \\ Y_{i2} &= \alpha + \beta x_{i2} + \eta_i \end{aligned}$$

where $i = 1, 2, \ldots, n$ and we have $E(\epsilon_i) = E(\eta_i) = 0$, $\text{var}(\epsilon_i) = \text{var}(\eta_i) = \sigma^2$ and $\text{cov}(\epsilon_i, \eta_i) = \rho\sigma^2$ for some correlation coefficient $\rho \in (-1, 1)$. Further assume that the pairs $(\epsilon_1, \eta_1), (\epsilon_2, \eta_2), \ldots, (\epsilon_n, \eta_n)$ are independent.

a. (5) Denote

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{12} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n1} \\ 1 & x_{n2} \end{pmatrix} \quad \text{and} \quad \mathbf{Y} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ \vdots \\ Y_{n1} \\ Y_{n2} \end{pmatrix} .$$

Is

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

an unbiased estimator of the two regression parameters? Explain.

Solution: by assumptions

$$E(\mathbf{Y}) = \mathbf{X} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} .$$

Using this and the rules for expectations it follows that the estimate is unbiased.

b. (10) Suggest an unbiased estimator of σ^2.

Solution: one possibility is to use only every second observation and use the usual unbiased estimator for σ^2.

c. (10) Suppose that ρ is known and define new pairs

$$\begin{aligned} \tilde{Y}_{i1} &= (\sqrt{1-\rho} + \sqrt{1+\rho})Y_{i1} + (\sqrt{1-\rho} - \sqrt{1+\rho})Y_{i2} \\ \tilde{Y}_{i2} &= (\sqrt{1-\rho} - \sqrt{1+\rho})Y_{i1} + (\sqrt{1-\rho} + \sqrt{1+\rho})Y_{i2} \end{aligned}$$

$$\begin{aligned} \tilde{x}_{i1} &= (\sqrt{1-\rho} + \sqrt{1+\rho})x_{i1} + (\sqrt{1-\rho} - \sqrt{1+\rho})x_{i2} \\ \tilde{x}_{i2} &= (\sqrt{1-\rho} - \sqrt{1+\rho})x_{i1} + (\sqrt{1-\rho} + \sqrt{1+\rho})x_{i2} \end{aligned}$$

and

$$\begin{aligned} \tilde{\epsilon}_i &= (\sqrt{1-\rho} + \sqrt{1+\rho})\epsilon_i + (\sqrt{1-\rho} - \sqrt{1+\rho})\eta_i \\ \tilde{\eta}_i &= (\sqrt{1-\rho} - \sqrt{1+\rho})\epsilon_i + (\sqrt{1-\rho} + \sqrt{1+\rho})\eta_i \end{aligned} .$$

Define $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{X}}$ accordingly. The new pairs satisfy the equations

$$\begin{aligned} \tilde{Y}_{i1} &= \alpha_1 + \beta\tilde{x}_{i1} + \tilde{\epsilon}_i \\ \tilde{Y}_{i2} &= \alpha_1 + \beta\tilde{x}_{i2} + \tilde{\eta}_i \end{aligned}$$

where $\alpha_1 = 2\sqrt{1-\rho}\,\alpha$. Argue that this new model satisfies the usual conditions for the regression models. What is then the best linear unbiased estimator of the regression parameters α and β. Explain.

Solution: we need to prove $E(\tilde{\epsilon}_i) = E(\tilde{\eta}_i) = 0$ *which follows easily. By a computation we prove that* $\mathrm{var}(\epsilon_i) = \mathrm{var}(\eta_i) = 4(1-\rho^2)\sigma^2$ *and* $\mathrm{cov}(\tilde{\epsilon}_i, \tilde{\eta}_i) = 0$*. The best linear unbiased estimator for* α_1 *and* β *is given by the Gauss-Markov theorem. But because* α *and* α_1 *differ by a known constant it follows that* $\alpha/(2\sqrt{1-\rho})$ *is the best unbiased estimate for* α*.*