University of Ljubljana, School of Economics

Quantitative finance and actuarial science

Probability and statistics

Written examination

February 16th, 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

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

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 that the data $x_1, x_2, \ldots, x_n$ are an i.i.d.sample from the discrete distribution given by

$$P(X_1 = x) = (x-1)p^2(1-p)^{x-2}$$

for $x = 2, 3, \ldots$.

a. (10) Find the MLE estimate for the parameter p.

Solution: the log-likelihood function is

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

Taking derivatives we get

$$\frac{\partial \ell}{\partial p} = \frac{2n}{p} - \frac{\sum_{k=1}^{n}(x_k - 2)}{1-p}\,.$$

Setting the derivative to 0 *we get*

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

b. (10) Compute the Fisher information matrix $I(p)$.

Solution: we compute the second derivatives of the log-likelihood function for $n = 1$. *We get*

$$\frac{\partial^2 \ell}{\partial p^2} = -\frac{2}{p^2} - \frac{x_1 - 2}{(1-p)^2}\,.$$

Replace x_1 *by* X_1 *and compute the expectation. We get*

$$I(p) = \frac{2}{p^2} - \frac{E(X_1) - 2}{(1-p)^2}\,.$$

To compute $E(X_1)$ *we either notice that* $X_1 \sim NegBin(2, p)$ *and hence* $E(X_1) = 2/p$,

$$E(X_1) = \sum_{k=2}^{\infty} k(k-1)p^2(1-p)^{k-2}$$

or we notice that

$$\sum_{k=2}^{\infty} k(k-1)x^{k-2} = \frac{d^2}{dx^2}\left(\sum_{k=0}^{\infty} x^k\right) = \frac{2}{(1-x)^3}$$

which gives us the same result. Finally,

$$I(p) = \frac{2}{p^2(1-p)}\,.$$

c. (5) Write the 95%-confidence interval for the parameter p based on $x_1, x_2, \ldots, x_n$.

Solution: the interval is

$$\hat{p} \pm 1.96 \cdot \frac{\hat{p}\sqrt{1-\hat{p}}}{\sqrt{2n}}\,.$$

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

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

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

4. (25) Assume the following regression model

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

for $i = 1, 2, \ldots, n$. Assume that the pairs $(\epsilon_1, \eta_1), \ldots, (\epsilon_n, \eta_n)$ are independent and identically distributed with $E(\epsilon_i) = E(\eta_i) = 0$, $\mathrm{var}(\epsilon_i) = \mathrm{var}(\eta_i) = \sigma^2$ and $\mathrm{corr}(\epsilon_i, \eta_i) = \rho$. Assume that ρ is known.

a. (5) Let

$$\hat{\beta} = \frac{\sum_{i=1}^n (Y_{i1}x_{i1} + Y_{i2}x_{i2})}{\sum_{i=1}^n (x_{i1}^2 + x_{i2}^2)} .$$

Is this estimator unbiased? Compute its standard error.

Solution: all the estimators in the sequel are of the form

$$\hat{\beta} = \sum_{i=1}^n (a_i Y_{i1} + b_i Y_{i2})$$

for suitable a_i and b_i. We have

$$E(\hat{\beta}) = \beta \sum_{i=1}^n (a_i x_{i1} + b_i x_{i2})$$

and

$$\mathrm{var}(\hat{\beta}) = \sum_{i=1}^n \mathrm{var}(a_i Y_{i1} + b_i Y_{i2}) = \sigma^2 \sum_{i=1}^n (a_i^2 + b_i^2 + 2\rho a_i b_i) .$$

Plugging in the respective a_i and b_i we find that all the estimators are unbiased and we derive the formulae for standard errors.

b. (5) Adding we get

$$Y_{i1} + Y_{i2} = \beta(x_{i1} + x_{i2}) + \xi_i ,$$

where $\xi_i = \epsilon_i + \eta_i$. The terms $\xi_1, \ldots, \xi_n$ are uncorrelated with $E(\xi_i) = 0$ and $\mathrm{var}(\xi_i) = \sigma^2(2 + \rho)$. The parameter β can be estimated as

$$\hat{\beta} = \frac{\sum_{i=1}^n (Y_{i1} + Y_{i2})(x_{i1} + x_{i2})}{\sum_{i=1}^n (x_{i1} + x_{i2})^2} .$$

Is this estimator unbiased? Compute ist standard error.

Solution: see a.

c. (5) Replace for each $i = 1, 2, \ldots, n$ the second equation by

$$\frac{Y_{i2} - \rho Y_{i1}}{1 - \rho^2} = \beta \left(\frac{x_{i2} - \rho x_{i1}}{1 - \rho^2} \right) + \left(\frac{\eta_i - \rho\epsilon_i}{1 - \rho^2} \right) .$$

Denote

$$\tilde{Y}_{i2} = \frac{Y_{i2} - \rho Y_{i1}}{1-\rho^2} \qquad \text{in} \quad \tilde{x}_{i2} = \frac{x_{i2} - \rho x_{i1}}{1-\rho^2}\,.$$

Estimate β by

$$\hat{\beta} = \frac{\sum_{i=1}^n \left(Y_{i1}x_{i1} + \tilde{Y}_{i2}\tilde{x}_{i2}\right)}{\sum_{i=1}^n (x_{i1}^2 + \tilde{x}_{i2}^2)}\,.$$

Is this estimate unbiased? Compute its standard error.

Solution: See a.

d. (10) Which of the above estimators has the smallest standard error? Explain.

Solution: let

$$\tilde{\eta}_i = \frac{\eta_i - \rho\epsilon_i}{1-\rho^2}\,.$$

This random variable is uncorrelated with ϵ_i and $E(\tilde{\eta}_i) = 0$ and $\mathrm{var}(\tilde{\eta}_i) = \sigma^2$. The model in c. satisfies all the assumptions of the Gauss-Markov theorem which means that the estimator in c. is the best linear unbiased estimator of the parameters.