University of Ljubljana, School of Economics

Quantitative finance and actuarial science

Probability and statistics

Written examination

August 31st, 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) Suppose a stratified sample is taken from a population of size N. The strata are of size $N_1, N_2, \ldots, N_K$, and the simple random samples are of size $n_1, n_2, \ldots, n_K$. Denote by μ the population mean and by σ^2 the population variance for the entire population, and by μ_k and σ_k^2 the population means and the population variances for the strata.

a. (5) Show that

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

where $w_k = \frac{N_k}{N}$ for $k = 1, 2, \ldots, K$.

Solution: by definition we have

$$\sigma^2 = \frac{1}{N} \left(\sum_{k=1}^{K} \sum_{i=1}^{N_k} (y_{ki} - \mu)^2 \right)$$

where y_{ki} is the value for the i-th unit in the k-th stratum. Note that

$$\begin{aligned}
&\sum_{i=1}^{N_k} (y_{ki} - \mu)^2 = \\
&= \sum_{i=1}^{N_k} (y_{ki} - \mu_k + \mu_k - \mu)^2 \\
&= \sum_{i=1}^{N_k} (y_{ki} - \mu_k)^2 + \sum_{i=1}^{N_k} (\mu_k - \mu)^2 + 2(\mu_k - \mu) \sum_{i=1}^{N_k} (y_{ki} - \mu) \\
&= \sum_{i=1}^{N_k} (y_{ki} - \mu_k)^2 + \sum_{i=1}^{N_k} (\mu_k - \mu)^2 \\
&= N_k \sigma_k^2 + N_k (\mu_k - \mu)^2 \,.
\end{aligned}$$

Using this in the above summation gives the result.

b. (10) Let $\bar{Y}_k$ be the sample average in the k-th stratum for $k = 1, 2, \ldots, K$ and $\bar{Y} = \sum_{k=1}^{K} w_k \bar{Y}_k$ the unbiased estomator of the population mean. The estimators $\bar{Y}_1, \ldots, \bar{Y}_n$ are assumed to be independent. To estimate σ^2, we need to estimate the quantity

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

The estimator

$$\hat{\sigma}_b^2 = \sum_{k=1}^{K} w_k \bar{Y}_k^2 - \bar{Y}^2$$

is suggested. Show that

$$E(\hat{\sigma}_b^2) = \sum_{k=1}^{K} w_k(1-w_k)\mathrm{var}(\bar{Y}_k) + \sum_{k=1}^{K} w_k \mu_k^2 - \mu^2 \,.$$

Solution: we know that

$$E(\bar{Y}_k^2) = \mathrm{var}(\bar{Y}_k^2) + \mu_k^2$$

and

$$E(\bar{Y}^2) = \mathrm{var}(\bar{Y}) + \mu^2 \,.$$

We have

$$E(\hat{\sigma}_b^2) = \sum_{k=1}^{K} w_k \left(\mathrm{var}(\bar{Y}_k^2) + \mu_k^2\right) - \mathrm{var}(\bar{Y}) - \mu^2 \,.$$

Taking into account that

$$\mathrm{var}(\bar{Y}) = \sum_{k=1}^{K} w_k^2 \mathrm{var}(\bar{Y}_k)$$

the result follows.

c. (10) Is there an unbiased estimator of σ^2? Explain your answer.

Solution: we know that

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

We have unbiased estimators for σ_k^2. The second term can be estimated by

$$\sum_{k=1}^{K} w_k \bar{Y}_k^2 - \bar{Y}^2 - \sum_{k=1}^{K} w_k(1-w_k)\frac{\hat{\sigma}_k^2}{n_k} \cdot \frac{N_k - n_k}{N_k - 1} \,.$$

This last term is an unbiased estimator of the second term.

2. (25) Let the observed values $x_1, x_2, \ldots, x_n$ be generated as independent, identically distributed random variables $X_1, X_2, \ldots, X_n$ with distribution

$$P(X_1 = x) = \frac{(\theta-1)^{x-1}}{\theta^x}$$

for $x = 1, 2, 3, \ldots$ and $\theta > 1$.

a. (10) Find the MLE estimate of θ based on the observations.

Solution: we find

$$\ell(\theta, \mathbf{x}) = \left(\sum_{k=1}^n x_k - n\right) \log(\theta - 1) - \left(\sum_{k=1}^n x_k\right) \log\theta\,.$$

Taking the derivative we have

$$\ell'(\theta, \mathbf{x}) = \frac{\sum_{k=1}^n x_k - n}{\theta - 1} - \frac{\sum_{k=1}^n x_k}{\theta} = 0\,.$$

It follows that

$$\hat{\theta} = \frac{1}{n}\sum_{k=1}^n x_k = \bar{x}\,.$$

b. (15) Write an approximate 99%-confidence interval for θ based on the observations. Assume as known that

$$\sum_{x=1}^\infty x a^{x-1} = \frac{1}{(1-a)^2}$$

for $|a| < 1$.

Solution: we have

$$\ell''(\theta, x) = -\frac{x-1}{(\theta-1)^2} + \frac{x}{\theta^2}\,.$$

To find the Fisher information we need

$$E(X_1) = \sum_{x=1}^\infty x \frac{(\theta-1)^{x-1}}{\theta^x}\,.$$

Using the hint we get

$$E(X_1) = \frac{1}{\theta} \cdot \left(1 - \frac{\theta-1}{\theta}\right)^{-2} = \theta\,.$$

We have

$$I(\theta) = \frac{1}{\theta(\theta-1)}\,.$$

An approximate 99%-confidence interval is

$$\hat{\theta} \pm 2.56 \cdot \sqrt{\frac{\hat{\theta}(\hat{\theta}-1)}{n}}\,.$$

3. (25) Assume the observations $x_1, \ldots, x_n$ are an i.i.d.sample from the $\Gamma(2, \theta)$ distribution with density

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

for $x > 0$ and $\theta > 0$.

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

   Solution: the log-likelihood function is

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

   Equating the derivative to 0 *we get*

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

b. (10) For the testing problem $H_0\colon \theta = 1$ versus $H_1\colon \theta \neq 1$ find the Wilks's test statistic λ. Describe when you would reject H_0 given that the size of the test is $1 - \alpha$ with $\alpha \in (0, 1)$.

   Solution: by definition

$$\lambda = 2\ell(\hat{\theta}) - 2\ell(1) .$$

   Using the maximum likelihood estimator $\hat{\theta}$ *we get*

$$\lambda = -4n \log\left(\frac{\bar{x}}{2}\right) + 2n\left(\bar{x} - 2\right) .$$

   By Wilks's theorem under H_0 *the distribution of the test statistic* λ *is approximately* $\chi^2(1)$. *The null-hypothesis is rejected when* $\lambda > c_\alpha$ *where* c_α *is such that* $P(\chi^2(1) \geq c_\alpha) = \alpha$.

c. (10) The function

$$f(y) = -4n \log\left(\frac{y}{2}\right) + 2n(y - 2)$$

   is strictly decreasing on $(0, 2)$ and strictly increasing on $(2, \infty)$. Assume for all $c > \min_{y>0} f(y)$ you can find the two solutions of the equation $f(y) = c$. Can you use this information to give an exact test given $\alpha \in (0, 1)$? Describe the procedure. No calculations are required.

   Hint: by properties of the gamma distribution $\bar{X} \sim \Gamma(2n, \theta/n)$.

   Solution: given the assumptions we can find such a c_α *that under* H_0 *we have*

$$P_{H_0}\left(f(\bar{X}) \geq c_\alpha\right) = \alpha .$$

   Let $x_1 < x_2$ *be the solutions of the equation* $f(x) = c_\alpha$. *The test that rejects* H_0 *when either* $\bar{X} < x_1$ *or* $\bar{X} > x_2$ *is exact.*

4. (25) Assume the regression equations are

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

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

$$E(\epsilon_k) = 0 \quad \text{and} \quad \operatorname{var}(\epsilon_k) = \sigma^2(1+\tau^2)$$

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

$$\operatorname{cov}(\epsilon_k, \epsilon_l) = \sigma^2\tau^2$$

for $k \neq l$ where τ^2 is assumed to be a known constant. Assume that $\sum_{k=1}^n x_k = 0$.

a. (10) Denote $\bar{Y} = \frac{1}{n}\sum_{k=1}^n Y_k$. Compute

$$\operatorname{cov}\left(Y_k - c\bar{Y}, Y_l - c\bar{Y}\right)$$

for $k \neq l$. Here c is an arbitrary constant.

Solution: from the assumptions we have

$$\operatorname{cov}\left(Y_k, \bar{Y}\right) = \frac{\sigma^2}{n}\left(1 + n\tau^2\right)$$

and

$$\operatorname{cov}\left(\bar{Y}, \bar{Y}\right) = \frac{\sigma^2}{n}\left(1 + n\tau^2\right) .$$

We have

$$\begin{aligned} &\operatorname{cov}\left(Y_k - c\bar{Y}, Y_l - c\bar{Y}\right) \\ &= \operatorname{cov}(Y_k, Y_l) - 2c \cdot \operatorname{cov}\left(Y_k, \bar{Y}\right) + c^2 \cdot \operatorname{cov}\left(\bar{Y}, \bar{Y}\right) \\ &= \sigma^2\left(\tau^2 - \frac{2c}{n}\left(1 + n\tau^2\right) + \frac{c^2}{n}\left(1 + n\tau^2\right)\right) . \end{aligned}$$

b. (10) Find an explicit formula for the best linear unbiased estimator of β.

Hint: choose

$$c = 1 - \sqrt{\frac{1}{1 + n\tau^2}} .$$

Solution: with the above choice of c we have that $c \in (0,1)$ and

$$\operatorname{cov}\left(Y_k - c\bar{Y}, Y_l - c\bar{Y}\right) = 0$$

for $k \neq l$. Define

$$\tilde{Y}_k = Y_k - c\bar{Y} ,$$

$$\tilde{\epsilon}_k = \epsilon_k - c\bar{\epsilon}$$

and

$$\tilde{\mathbf{X}} = \begin{pmatrix} 1-c & x_1 \\ 1-c & x_2 \\ \vdots & \vdots \\ 1-c & x_n \end{pmatrix} .$$

We have

$$\tilde{Y}_k = \alpha(1-c) + \beta x_k + \tilde{\epsilon}_k$$

for $k = 1, 2, \ldots, n$. The new regression equations satisfy the usual assumptions of the Gauss-Markov theorem. The best linear estimators of the regression parameters are

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} n(1-c)^2 & 0 \\ 0 & \sum_{k=1}^n x_k^2 \end{pmatrix}^{-1} \begin{pmatrix} (1-c)\sum_{k=1}^n Y_k \\ \sum_{k=1}^n x_k Y_k \end{pmatrix} .$$

We get

$$\hat{\beta} = \frac{\sum_{k=1}^n x_k Y_k}{\sum_{k=1}^n x_k^2} .$$

c. (5) Compute the variance of the best linear unbiased estimator $\hat{\beta}$.

Solution: we compute directly

$$\begin{aligned} \mathrm{var}(\hat{\beta}) &= \mathrm{var}\left(\frac{\sum_{k=1}^n x_k Y_k}{\sum_{k=1}^n x_k^2}\right) \\ &= \frac{\sigma^2}{\left(\sum_{k=1}^n x_k^2\right)^2}\left(\sum_{k=1}^n x_k^2(1+\tau^2) + \sum_{\substack{k,l \\ k\neq l}} x_k x_l \tau^2\right) \\ &= \frac{\sigma^2}{\left(\sum_{k=1}^n x_k^2\right)^2}\sum_{k=1}^n x_k^2(1+\tau^2) \\ &= \frac{\sigma^2(1+\tau^2)}{\sum_{k=1}^n x_k^2} \end{aligned}$$