School of Economics

Doctoral Programme in Economics

Probability and statistics

Written examination

February 14$^{\text{th}}$, 2025

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.	Total
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

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

   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) The Gumbel distribution of type 1 for extreme values is given by the density

$$f_X(x) = \frac{1}{\beta} e^{\frac{x-\mu}{\beta}} e^{-e^{\frac{x-\mu}{\beta}}} .$$

for $\beta > 0$, $-\infty < \mu < \infty$ and $-\infty < x < \infty$. Assume as known the following integrals

$$\begin{aligned}
&\int_{-\infty}^{\infty} e^{2u} e^{-e^u}\, du = 1 \\
&\int_{-\infty}^{\infty} u e^{u} e^{-e^u}\, du = -\gamma\,, \\
&\int_{-\infty}^{\infty} u e^{2u} e^{-e^u}\, du = 1-\gamma\,, \\
&\int_{-\infty}^{\infty} u^2 e^{2u} e^{-e^u}\, du = -2\gamma + \gamma^2 + \tfrac{\pi^2}{6}\,,
\end{aligned}$$

where $\gamma = 0,577216\ldots$ is the Euler constant. Assume that the observed values are an i.i.d. sample from the Gumbel distribution.

a. (15) Compute the Fisher information matrix $I(\beta, \mu)$.

Solution: we have

$$\ell = -\log\beta + \frac{x-\mu}{\beta} - e^{-\frac{x-\mu}{\beta}} .$$

Taking partial derivatives we get

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \beta^2} &= -\frac{(x-\mu)^2 e^{\frac{x-\mu}{\beta}}}{\beta^4} - \frac{2(x-\mu)e^{\frac{x-\mu}{\beta}}}{\beta^3} + \frac{2(x-\mu)}{\beta^3} + \frac{1}{\beta^2} \\
\frac{\partial^2 \ell}{\partial \beta\,\partial \mu} &= -\frac{(x-\mu)e^{\frac{x-\mu}{\beta}}}{\beta^3} - \frac{e^{\frac{x-\mu}{\beta}}}{\beta^2} + \frac{1}{\beta^2} \\
\frac{\partial^2 \ell}{\partial \mu^2} &= -\frac{e^{\frac{x-\mu}{\beta}}}{\beta^2} .
\end{aligned}$$

Introduce the new variable $u = (x-\mu)/\beta$ in all integrals. We get

$$\begin{aligned}
-E\left(\frac{\partial^2 \ell}{\partial \beta^2}\right) &= \frac{1}{\beta^2}\int_{-\infty}^{\infty} \left(u^2 e^u + 2ue^u - 2u - 1\right) e^u e^{-e^u}\, du \\
-E\left(\frac{\partial^2 \ell}{\partial \beta\,\partial \mu}\right) &= \frac{1}{\beta^2}\int_{-\infty}^{\infty} \left(ue^u + e^u - 1\right) e^u e^{-e^u}\, du \\
-E\left(\frac{\partial^2 \ell}{\partial \mu^2}\right) &= \frac{1}{\beta^2}\int_{-\infty}^{\infty} e^u e^u e^{-e^u}\, du
\end{aligned}$$

It follows

$$I(\beta,\mu) = \begin{pmatrix} \frac{1}{\beta^2}\left((1-\gamma)^2 + \pi^2/6\right) & \frac{1-\gamma}{\beta^2} \\ \frac{1-\gamma}{\beta^2} & \frac{1}{\beta^2} \end{pmatrix} .$$

b. (10) Assume $\hat{\mu}$ and $\hat{\beta}$ are maximum likelihood estimates based on observed values $x_1, x_2, \ldots, x_n$. Give an approximate 95%-confidence interval for both parameters based on the observed values.

Solution: we compute

$$I^{-1}(\beta,\mu) = \frac{6\beta^2}{\pi^2}\begin{pmatrix} 1 & -1+\gamma \\ -1+\gamma & (1-\gamma)^2 + \pi^2/6 \end{pmatrix} .$$

We get

$$\mathrm{se}(\hat{\beta}) = \frac{\sqrt{6}\hat{\beta}}{\pi\sqrt{n}} \qquad \textit{in} \qquad \mathrm{se}(\hat{\mu}) = \frac{\sqrt{6((1-\gamma)^2 + \pi^2/6)}\hat{\beta}}{\pi\sqrt{n}}\,.$$

The confidence intervals follow.

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 bivariate normal density

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2}{2(1-\rho^2)}} .$$

Assume that $\rho \in (-1, 1)$ is known. We would like to test the hypothesis

$$H_0 \colon \mu = \nu \quad \text{versus} \quad H_1 \colon \mu \neq \nu \, .$$

a. (10) Find the maximum likelihood estimates for μ and ν.

Solution: the log-likelihood function is

$$-n \log 2\pi - \frac{n}{2} \log(1-\rho^2)$$

$$-\frac{1}{2(1-\rho^2)} \sum_{k=1}^{n} \left[(x_k-\mu)^2 - 2\rho(x_k-\mu)(y_k-\nu) + (y_k-\nu)^2 \right] .$$

Taking partial derivatives we get

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{1-\rho^2} \sum_{k=1}^{n} \left[(x_k-\mu) - \rho(y_k-\nu) \right]$$

and

$$\frac{\partial \ell}{\partial \nu} = \frac{1}{1-\rho^2} \sum_{k=1}^{n} \left[-\rho(x_k-\mu) + (y_k-\nu) \right] .$$

The partial derivatives are set to 0 *and we get the system of linear equations equal to*

$$\begin{array}{rcrcl} n\mu & - & \rho n\nu & = & \sum_{k=1}^n x_k - \rho \sum_{k=1}^n y_k \\ -\rho n\mu & + & n\nu & = & -\rho \sum_{k=1}^n x_k + \sum_{k=1}^n y_k \end{array}$$

Dividing by n and solving for μ and ν we get

$$\hat{\mu} = \bar{x} \quad \textit{in} \quad \hat{\nu} = \bar{y} \, .$$

b. (10) Find the likelihood ratio statistic for testing the above hypothesis. What is the approximate distribution of the test statistic under H_0?

Solution: if $\mu = \nu$, the likelihood function reduces to

$$-n \log 2\pi - \frac{n}{2} \log(1-\rho^2)$$

$$-\frac{1}{2(1-\rho^2)} \sum_{k=1}^{n} \left[(x_k-\mu)^2 - 2\rho(x_k-\mu)(y_k-\mu) + (y_k-\mu)^2 \right] .$$

We have

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{1-\rho^2} \sum_{k=1}^{n} \left[(x_k-\mu) - \rho(y_k-\mu) - \rho(x_k-\mu) + (y_k-\mu)\right] .$$

The constrained maxumum likelihood estimate is

$$\tilde{\mu} = \tilde{\nu} = \frac{\bar{x}+\bar{y}}{2} .$$

By definition

$$\lambda = 2\log(\hat{\mu},\hat{\nu}|\mathbf{x},\mathbf{y}) - 2\log(\tilde{\mu},\tilde{\nu}|\mathbf{x},\mathbf{y}) .$$

In more compact form we get

$$\begin{aligned} &\log(\hat{\mu},\hat{\nu}|\mathbf{x},\mathbf{y}) \\ &= -n\log 2\pi - \frac{n}{2}\log(1-\rho^2) \\ &\quad -\frac{1}{2(1-\rho^2)}\left[\sum_{k=1}^{n} x_k^2 - n\bar{x}^2 - 2\rho\left(\sum_{k=1}^{n} x_k y_k - n\bar{x}\bar{y}\right) + \sum_{k=1}^{n} y_k^2 - n\bar{y}^2\right] \end{aligned}$$

and

$$\begin{aligned} &\log(\tilde{\mu},\tilde{\nu}|\mathbf{x},\mathbf{y}) \\ &= -n\log 2\pi - \frac{n}{2}\log(1-\rho^2) \\ &\quad -\frac{1}{2(1-\rho^2)}\Bigg[\sum_{k=1}^{n} x_k^2 - n\bar{x}(\bar{x}+\bar{y}) + \frac{n}{4}(\bar{x}+\bar{y})^2 \\ &\qquad -2\rho\left(\sum_{k=1}^{n} x_k y_k - \frac{n\bar{x}}{2}(\bar{x}+\bar{y}) - \frac{n\bar{y}}{2}(\bar{x}+\bar{y}) + \frac{n}{4}(\bar{x}+\bar{y})^2\right) \\ &\qquad + \sum_{k=1}^{n} y_k^2 - n\bar{y}(\bar{x}+\bar{y}) + \frac{n}{4}(\bar{x}+\bar{y})^2\Bigg] \\ &= -\frac{1}{2(1-\rho^2)}\left[\sum_{k=1}^{n}(x_k^2 + y_k^2 - 2\rho x_k y_k) - n(1-\rho)(\bar{x}+\bar{y})^2\right] . \end{aligned}$$

We have

$$\lambda = \frac{n}{1-\rho^2}\left[\bar{x}^2 - 2\rho\bar{x}\bar{y} + \bar{y}^2 - \frac{1}{2}(1-\rho)(\bar{x}+\bar{y})^2\right]$$

or alternatively

$$\lambda = \frac{n}{2-2\rho}\left(\bar{x}-\bar{y}\right)^2 .$$

The approximate distribution of λ is $\chi^2(1)$ by Wilks's theorem. In fact, the $\chi^2(1)$ distribution is the exact distribution of λ.

c. (5) What is the distribution of $X - Y$ if H_0 holds? Can you use the result to give an alternative test statistic to test the above hypothesis? What is the distribution of your test statistic under H_0?

Solution: The distribution of the difference if H_0 holds is $\mathrm{N}(0, 2 - 2\rho)$. We can apply the usual z-test to the observed values $z_1 = x_1 - y_1, \ldots, x_n - y_n$ since the variance is known. We get exactly the test from the previous part.

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

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

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

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

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

a. (5) Let

$$\hat{\alpha} = \frac{1}{n}\sum_{k=1}^n Y_k \qquad \text{and} \qquad \hat{\beta} = \frac{\sum_{k=1}^n x_k Y_k}{\sum_{k=1}^n x_k^2}$$

be the ordinary least squares estimators of the two regression parameters. Show that the estimators are unbiased.

Solution: from the assumptions we have

$$E\left(\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix}\right) = \begin{pmatrix}\frac{1}{n}\sum_{k=1}^n(\alpha+\beta+E(\epsilon_k))\\ \frac{\sum_{k=1}^n x_k(\alpha+\beta x_k+E(\epsilon_k)}{\sum_{k=1}^n x_k^2}\end{pmatrix} = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}.$$

b. (5) Let

$$\begin{pmatrix}\tilde{\alpha}\\ \tilde{\beta}\end{pmatrix} = \begin{pmatrix}\sum_{k=1}^n a_k Y_k\\ \sum_{k=1}^n b_k Y_k\end{pmatrix}$$

be a linear unbiased estimator of the regression parameters. Compute

$$\mathrm{cov}(\tilde{\alpha} - \hat{\alpha}, \hat{\alpha}).$$

Solution: by assumption

$$E\left(\sum_{k=1}^n a_k E(Y_k)\right) = \sum_{k=1}^n a_k(\alpha + \beta x_k) = \alpha$$

which means $\sum_{k=1}^n a_k = 1$. *Compute*

$$\begin{aligned}
&\mathrm{cov}(\tilde{\alpha} - \hat{\alpha}, \hat{\alpha})\\
&= \frac{1}{n}\mathrm{cov}\left(\sum_{k=1}^n (a_k - \frac{1}{n})Y_k, \sum_{l=1}^n Y_l\right)\\
&= \frac{1}{n}\sum_{k=1}^n\left(a_k - \frac{1}{n}\right)\sigma^2(1+\tau^2)\\
&\qquad + \frac{1}{n}\sum_{k\neq l}\frac{1}{n}\left(a_k - \frac{1}{n}\right)\sigma^2\tau^2\\
&= 0.
\end{aligned}$$

c. (5) Find the best unbiased linear estimator of α.

Solution: compute

$$\mathrm{var}(\tilde{\alpha}) = \mathrm{var}(\tilde{\alpha} - \hat{\alpha} + \hat{\alpha}) + \mathrm{var}(\hat{\alpha}) + \mathrm{var}(\tilde{\alpha} - \hat{\alpha}) \geq \mathrm{var}(\hat{\alpha})\,.$$

The assertion follows.

d. (10) Find the best unbiased linear estimator of β.

Solution: by assumption

$$E\left(\tilde{\beta}\right) = \sum_{k=1}^{n} b_k(\alpha + \beta x_k)$$

which implies $\sum_{k=1}^{n} b_k = 0$ *and* $\sum_{k=1}^{n} b_k x_k = 1$. *Denote* $\sum_{k=1}^{n} x_k^2 = s$ *and compute*

$$\begin{aligned}
&\mathrm{cov}(\tilde{\beta} - \hat{\beta}, \hat{\beta}) \\
&= \mathrm{cov}\left(\sum_{k=1}^{n}\left(b_k - \frac{x_k}{s}\right)Y_k, \sum_{l=1}^{n}\frac{x_l}{s}Y_l\right) \\
&= \sum_{k=1}^{n}\left(b_k - \frac{x_k}{s}\right)\left(\frac{x_k}{s}\right)\sigma^2(1+\tau^2) \\
&\qquad\qquad \sum_{k\neq l}\left(b_k - \frac{x_k}{s}\right)\frac{x_l}{s}\sigma^2\tau^2 \\
&= 0\,.
\end{aligned}$$

The assertion about the best unbiased linear estimator follows the same way as in c.