University of Ljubljana, Faculty of Economics

Quantitative finance and actuarial science

Probability and statistics

Written examination

February 12th, 2021

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 we have a population with N units. The values of the statistical variable are $x_1, x_2, \ldots, x_N$. Denote by μ the population mean and by σ^2 the population variance.

a. (5) Suppose you chose a simple random sample of size n. Denote

$$\gamma = \frac{1}{N} \sum_{k=1}^{N} x_k^2 .$$

Suggest an unbiased estimate for γ. Explain why it is unbiased.

Solution: An unbiased estimate of γ is the sample average of the squares of sample values. We also have

$$\sigma^2 = \frac{1}{N} \sum_{k=1}^{N} x_k^2 - \mu^2 = \gamma - \mu^2 .$$

b. (5) Suppose you chose a simple random sample of size n. Suggest an unbiased estimate for μ^2.

Hint: Note that $\sigma^2 = \gamma - \mu^2$.

Solution: We know that

$$\hat{\sigma}^2 = \frac{N-1}{N(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

is an unbiased estimate of σ^2. We have denoted the sample values by $X_1, \ldots, X_n$. Since in the above equation in the hint we have unbiased estimates for two of the three quantities and the relationship is linear, we can estimate the third, i.e. μ^2, in an unbiased way.

c. (5) Assume now that the population is divided into K equally sized groups of size M so that $N = KM$. A sample is chosen in such a way that k groups are chosen from all the K groups by simple random sampling. Then all the units from the chosen groups are included into the sample. For the estimator we chose the average of all the kM sample values. Denote by μ_k the population average for the k-the group and by σ_k^2 the population variance for the k-th group. Find the standard error of the suggested estimator using the quantity

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

Solution: Since all the groups are of equal size we have $\mu = \frac{1}{K}\sum_{k=1}^{K}\mu_k$*. We can think that we are choosing a simple random sample form a population of groups. The estimator is therefore unbiased and its variance is given by*

$$\operatorname{var}(\bar{X}) = \frac{\tau^2}{k} \cdot \frac{K-k}{K-1}\,,$$

where

$$\tau^2 = \frac{1}{K}\sum_{r=1}^{K}(\mu_r - \mu)^2\,.$$

d. (10) Assume that the sample is as in c. We would like to estimate the population variance σ^2 on the basis of the sample. Suggest and unbiased estimate. Explain why it is unbiased.

Hint: look at a.

Solution: We think of groups as our primary sampling units. From a. we know that μ^2 *can be estimated in an unbiased way. Returning to our sampling procedure we see that we have an unbiased estimator of*

$$\frac{1}{N}\sum_{k=1}^{N} x_k^2\,.$$

Since

$$\sigma^2 = \frac{1}{N}\sum_{k=1}^{N} x_k^2 - \mu^2$$

and we know how to estimate both quantities on the right we can estimate σ^2 *in an unbiased way.*

To express the estimator explicitly denote by X_{ij} *the value of the variable for the* j*th unit in the* i*th group selected and let* A_i *be the average in this group, and by* $\bar{A}$ *the average of all the group averages which is our estimator. We have*

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{kM}\sum_{i=1}^{k}\sum_{j=1}^{M} X_{ij}^2 - \frac{1}{k}\sum_{i=1}^{k} A_i^2 + \frac{K-1}{K(k-1)}\sum_{i=1}^{k}(A_i - \bar{A})^2 \\ &= \frac{1}{kM}\sum_{i=1}^{k}\sum_{j=1}^{M}(X_{ij} - A_i)^2 + \frac{K-1}{K(k-1)}\sum_{i=1}^{k}(A_i - \bar{A})^2\,.\end{aligned}$$

2. (25) Assume the data $x_1, x_2, \ldots, x_n$ are an i.i.d. sample from the distribution given by

$$P(X_1 = x) = \binom{2x}{x} \frac{\beta^x}{4^x (1+\beta)^{x+\frac{1}{2}}}$$

for $x = 0, 1, \ldots$ and $\beta > 0$.

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

Solution: The log-likelihood function is given by

$$\ell(\beta|\mathbf{x}) = \sum_{k=1}^{n} \log \binom{2x_k}{x_k} + \log\beta \sum_{k=1}^{n} x_k - \log 4 \sum_{k=1}^{n} - \log(1+\beta) \sum_{k=1}^{n} \left(x_k + \frac{1}{2}\right).$$

Taking derivatives and equating with 0 *we get the equation*

$$\frac{1}{\beta}\sum_{k=1}^{n} x_k - \frac{1}{1+\beta}\sum_{k=1}^{n}\left(x_k + \frac{1}{2}\right) = 0\,.$$

Hence

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

b. (5) Convince yourself that

$$\begin{aligned} E(X_1) &= \sum_{k=0}^{\infty} kP(X_1 = k) \\ &= \frac{2\beta}{4(1+\beta)} \sum_{k=1}^{\infty} [2(k-1)+1] P(X_1 = k-1) \\ &= \frac{2\beta}{4(1+\beta)} 2E(X_1) + \frac{2\beta}{4(1+\beta)}\,. \end{aligned}$$

Use this to show that the maximum likelihood estimator is unbiased.

Solution: The equality can be checked by a straightforward computation. The equality transforms into

$$E(X_1) = \frac{\beta}{1+\beta} E(X_1) + \frac{\beta}{2(1+\beta)}$$

or

$$E(X_1) = \frac{\beta}{2}\,.$$

We have

$$E\left(\hat{\beta}\right) = E\left(\frac{2\sum_{k=1}^{n} X_k}{n}\right) = \beta\,,$$

hence the estimator is unbiased.

c. (5) Use the Fisher information to give an approximate standard error for the maximum likelihood estimator.

Solution: Compute for $n = 1$:

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

hence

$$E(-\ell'') = \frac{1}{2\beta} + \frac{\frac{\beta}{2} + \frac{1}{2}}{(1+\beta)^2} = \frac{1}{2} \cdot \frac{1}{\beta(\beta+1)} .$$

It follows that

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

d. (10) Convince yourself that

$$\begin{aligned} E(X_1^2) &= \sum_{k=0}^{\infty} k^2 P(X_1 = k) \\ &= \frac{\beta}{4(1+\beta)} \sum_{k=1}^{\infty} \left[4(k-1)^2 + 6(k-1) + 2\right] P(X_1 = k-1) \\ &= \frac{\beta}{4(1+\beta)} \left(4E(X_1^2) + 6E(X_1) + 2\right) . \end{aligned}$$

Compute the exact standard error of the maximum likelihood estimator.

Solution: The equality is checked by a straightforward caclulation. We get the equation

$$E(X_1^2)(1+\beta) = \beta E(X_1^2) + \frac{3\beta}{2} E(X_1) + \frac{\beta}{2}$$

or

$$E(X_1^2) = \frac{\beta(2+3\beta)}{4}$$

and as a consequence

$$\text{var}(X_1) = \frac{\beta(1+\beta)}{2} .$$

the exact variance of the estimator $\hat{\beta}$ is

$$\text{var}(\hat{\beta}) = \frac{2\beta(1+\beta)}{n^2}$$

3. (25) Gauss' gamma distribution is given by the density

$$f(x,y) = \sqrt{\frac{2\lambda}{\pi}}\, y\, e^{-y}\, e^{-\frac{\lambda y (x-\mu)^2}{2}}\,.$$

for $-\infty < x < \infty$ and $y > 0$ and $(\mu, \lambda) \in \mathbb{R} \times (0, \infty)$. Assume that the observations are pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ generated as independent random pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ with density $f(x, y)$. We would like to test

$$H_0\colon \mu = 0 \qquad \text{versus} \qquad H_1\colon \mu \neq 0\,.$$

a. (15) Compute the maximum likelihood estimates of the parameters. Compute the maximum likelihood estimate of λ when $\mu = 0$.

Solution: The log-likelihood function is

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

Equate the partial derivatives with 0 to get

$$\frac{n}{2\lambda} - \frac{1}{2}\sum_{k=1}^{n} y_k (x_k - \mu)^2 = 0$$

and

$$\lambda \sum_{k=1}^{n} y_k (x_k - \mu) = 0\,.$$

From the second equation we get

$$\hat{\mu} = \frac{\sum_{k=1}^{n} x_k y_k}{\sum_{k=1}^{n} y_k}\,.$$

Insert this into the first equation to get

$$\hat{\lambda} = \frac{n}{\sum_{k=1}^{n} y_k (x_k - \hat{\mu})^2}\,.$$

When $\mu = 0$ the first equation determines λ. We get

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

b. (10) Find the likelihood ratio statistics for the above testing problem. What is its approximate distribution under H_0?

Solution: Ths test statistic is

$$\begin{aligned}\lambda &= 2\Big[\ell(\hat\lambda, \hat\mu|\mathbf{x}, \mathbf{y}) - \ell(\tilde\lambda, 0|\mathbf{x}, \mathbf{y})\Big] \\ &= n(\log\hat\lambda - \log\tilde\lambda) - \hat\lambda\sum_{k=1}^n y_k(x_k - \hat\mu)^2 + \tilde\lambda\sum_{k=1}^n x_k^2 y_k\,.\end{aligned}$$

However, from the equations for estimates we get that

$$\hat\lambda\sum_{k=1}^n y_k(x_k - \hat\mu)^2 = \tilde\lambda\sum_{k=1}^n x_k^2 y_k = n\,,$$

which implies

$$\lambda = n\log\frac{\hat\lambda}{\tilde\lambda}\,.$$

By Wilks's theorem, under H_0 the distribution of the test statistic is approximately $\chi^2(1)$.

4. (25) Assume the regression model

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

for $k = 1, 2, \ldots, n$ where $\epsilon_1, \ldots, \epsilon_n$ are uncorrelated, $E(\epsilon_k) = 0$ and $\text{var}(\epsilon_k) = \sigma^2$ for $k = 1, 2, \ldots, n$. Assume that $x_k > 0$ for all $k = 1, 2, \ldots, n$. 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 $k = 1, 2, \ldots, n$ 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. Guass-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 $Y_1, \ldots, Y_n$ are uncorrelated we have

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

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.