

UNIVERSITY OF OSLO
DEPARTMENT OF ECONOMICS

Exam: **ECON4130 – Statistics 2**

Date of exam: Friday, December 8, 2017 **Grades are given:** January 4, 2018

Time for exam: 09.00 a.m. – 12.00 noon

The problem set covers 5 pages

Resources allowed:

- Open book exam, where all written and printed resources – as well as calculator - is allowed

The grades given: A-F, with A as the best and E as the weakest passing grade. F is fail.

Problem 1

- A.** Let (X,Y) be two random variables (rv's) such that X is normally distributed with expectation, $E(X) = 0$, and variance, $\text{var}(X) = 1$ (written short, $X \sim N(0, 1)$), and the conditional distribution of Y given that $X = x$ is fixed, is normal with conditional expectation, $E(Y | x) = x$, and conditional variance, $\text{var}(Y | x) = 1$ (written short $Y | x \sim N(x, 1)$).

[**A reminder you may need below:** The general normal probability density function (pdf) for $V \sim N(\mu, \sigma^2)$ is given by

$$f_V(v; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(v-\mu)^2}, \quad -\infty < v < \infty$$

and the moment generating function (mgf) by

$$M_V(t) = e^{i\mu + \frac{t^2}{2}\sigma^2}, \quad -\infty < t < \infty]$$

- i.** Write up the formula for the pdf, $f(y | x)$, of Y given that $X = x$.

- ii.** Show that $E(Y) = 0$ and $\text{var}(Y) = 2$.

[**Hint.** Use, e.g., the law of total expectation.]

- iii.** Show that the correlation coefficient between X and Y is $\rho = 1/\sqrt{2}$.

[**Hint.** Remember that $E(XY) = E[X \cdot E(Y | X)]$.]

B. Let X_1, X_2, \dots, X_n be independent and identically distributed (*iid*) where $X_i \sim N(0, \theta)$, $i = 1, 2, \dots, n$ (and, hence, with $E(X_i) = 0$, $\text{var}(X_i) = \theta > 0$, and standard deviation, $\text{SD}(X_i) = \sqrt{\theta}$).

i. Show that the maximum likelihood estimator (mle) of θ is $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i^2$.

ii. Show that $\hat{\theta}$ is also the moment method estimator (mme) of θ .

C. i. Let X_i be as defined in section **B**. Explain why $\frac{X_i^2}{\theta}$ is chi-square distributed with 1 degree of freedom (in short: $\frac{X_i^2}{\theta} \sim \chi_1^2$), $i = 1, 2, \dots, n$.

[Hint. You may take as a fact (that you do not need to prove), that if $Z \sim N(0, 1)$, then Z^2 is chi-square distributed with 1 degree of freedom ($Z^2 \sim \chi_1^2$).]

ii. Show that the mle, $\hat{\theta}$, is unbiased and has variance, $\text{var}(\hat{\theta}) = \frac{2\theta^2}{n}$.

iii. Show that the mle, $\hat{\theta}$ is efficient in the sense that it has smallest possible variance among all unbiased estimators for θ that are based on X_1, X_2, \dots, X_n .

D. i. Explain why $\frac{n\hat{\theta}}{\theta}$ is chi-square distributed with n degrees of freedom

(short: $\frac{n\hat{\theta}}{\theta} \sim \chi_n^2$).

[Hint. You may benefit from the hint in **C(i)** and other relevant established properties of the chi-square distribution.]

ii. Data: We have $n = 15$ observations, x_1, x_2, \dots, x_n , where $\sum_{i=1}^{15} x_i^2 = 91.28$. Assuming the model in section **B.**, develop and calculate an (exact) 95% confidence interval (ci), based on $\hat{\theta}$, for the standard deviation of X_i , i.e., for $\sqrt{\theta}$.

[Hint. Find first a 95% ci for θ derived from the probability statement, $P\left(c_1 < \frac{n\hat{\theta}}{\theta} < c_2\right) = 0.95$, where c_1, c_2 are suitable percentiles in the relevant chi-square distribution. Then transform this interval to a 95% ci for $\sqrt{\theta}$.]

iii. Why is $\sqrt{\hat{\theta}}$ a consistent estimator of $\sqrt{\theta}$? Is $\sqrt{\hat{\theta}}$ unbiased for $\sqrt{\theta}$? Give reasons for your answers.

- E.** Return to the situation in section **A**. Show that the marginal distribution of Y is normal ($N(0, 2)$).

[Hint. There are several ways to show this. If you do not have a way of your own, you may try one of the following two approaches:

Approach 1: Write up the mgf, i.e., $E[e^{tY} | x]$ of the conditional distribution of Y given $X = x$ and find the (marginal) mgf of Y using the law of total expectation.

Approach 2: Show that (X, Y) is jointly bivariate normally distributed. One way to do this is first to write up the general formula for the bivariate normal pdf, then substitute the five parameter values you found in section **A**, and show that the resulting expression reduces to the actual joint pdf for (X, Y) that you find from the information in section **A**.]

Problem 2

- A.** Let X be Pareto distributed with parameters (b, θ) and cumulative distribution function (cdf)

$$F(x) = \begin{cases} 1 - (b/x)^\theta & \text{for } x > b \\ 0 & \text{for } x \leq b \end{cases}$$

where b is a known positive number, and $\theta > 0$ a parameter.

- i.** Show that $V = \theta \cdot \ln\left(\frac{X}{b}\right)$ is exponentially distributed with parameter 1 (in short: $V \sim \exp(1)$).
- ii.** Let $Y = \ln\left(\frac{X}{b}\right) = \frac{V}{\theta}$. Explain why $Y \sim \exp(\theta)$.

- B.** Suppose Y is gamma distributed with shape parameter $\alpha > 0$ and scale parameter $\lambda > 0$ (in short: $Y \sim \Gamma(\alpha, \lambda)$).

- i.** Use the moment generating function (mgf) of Y , $M_Y(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha$ for $t < \lambda$, to show that the expected value is $E(Y) = \frac{\alpha}{\lambda}$.
- ii.** Use the mgf of Y to show that $\frac{Y}{c} \sim \Gamma(\alpha, c\lambda)$, where c is an arbitrary positive constant.

- C. Introduction.** Let X be the income of a randomly chosen woman in Norway with income over 250 000 kr. On several occasions in the course we have assumed that X is

Pareto distributed (b, θ) , where $b = 250\,000$, and $\theta > 0$ a parameter. We want to test this Pareto assumption here in this section. The data is a random sample of $n = 2361$ women, all with income over 250 000 kr., drawn from SSB's¹ database in 1998.

One way to test the Pareto assumption is to embed the Pareto model into a larger model and test the Pareto assumption as a sub-model: According to section A, assuming

$X \sim \text{Pareto}(b, \theta)$ is equivalent with assuming $Y = \ln\left(\frac{X}{b}\right) \sim \exp(\theta)$. Since the $\exp(\theta)$

distribution is the same as the $\Gamma(1, \theta)$ distribution, a wider model would be to assume

$Y = \ln\left(\frac{X}{b}\right) \sim \Gamma(\alpha, \lambda)$, where $\alpha > 0$ and $\lambda > 0$ are parameters. To test the Pareto

assumption would then be equivalent to test $H_0 : \alpha = 1$ in the wider model.

Question.

- (1) **Model:** Let Y_1, Y_2, \dots, Y_n be iid with $Y_i \sim \Gamma(\alpha, \lambda)$ for $i = 1, 2, \dots, n$,

where $Y_i = \ln\left(\frac{X_i}{b}\right)$, $b = 250\,000$, and $n = 2361$.

In the appendix you will find (part of) the Stata output from a maximum likelihood estimation (mle) of model (1).

Use the output to test $H_0 : \alpha = 1$ versus $H_1 : \alpha \neq 1$ based on the mle $\hat{\alpha}$ and its standard error. Calculate the p-value (approximately) and comment on the result.

[**Hint.** You may consider the standard errors for the coefficients estimated in the output as consistent estimates of the theoretical standard errors.]

- D.** Alternatively, we may test the Pareto model (i.e., the exponential model for Y) versus the gamma model for Y , as in section C, using a likelihood ratio test (LR):

- i.** The Stata output in the appendix contains the maximum log likelihood for the gamma model (1) under the heading "Log likelihood", based on the available data for $Y = \ln(X/b)$. Using the same data and the maximum likelihood method to estimate the exponential model for Y (that corresponds to the Pareto model for X), gives, according to Stata, a maximum log likelihood equal to 799.25246.

Use this information to perform the LR test for testing the null hypothesis of an exponential distribution for Y versus the gamma model in (1). Use the level of significance 5% and formulate a conclusion.

- ii.** Suppose that V is chi-square distributed with 1 degree of freedom

¹ SSB is *Statistics Norway*.

