UNIVERSITETET I OSLO

Matematisk Institutt

EXAM IN:	**STK 4180/9180 – Confidence Distributions**
	Part II of two parts
WITH:	**Nils Lid Hjort**
AUXILIA:	**Calculator, plus one single sheet of paper**
	with the candidate's own personal notes
TIME FOR EXAM:	**Part I: The Project, 1–13/vi/2016;**
	Part II: Thursday 13/vi s.y., 14:30–18:30, written exam

This exam set contains four exercises and comprises three pages.

Exercise 1

Consider a simple regression model, with just a single covariate, of the form

$$Y_i = bx_i + \varepsilon_i \quad \text{for } i = 1, \ldots, n,$$

and with the ε_i being i.i.d. $\mathrm{N}(0, \sigma^2)$. For simplicity we take the error standard deviation σ to be known and equal to 1, so the regression slope coefficient b is the single unknown parameter. You do not need to use time here to show that the least squares estimator is

$$\widehat{b} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2} = (1/M) \sum_{i=1}^n x_i Y_i,$$

with $M = \sum_{i=1}^n x_i^2$. You may also find it convenient for some of the work below to use that the log-likelihood function may be written as

$$\ell(b) = -\tfrac{1}{2} Q_0 - \tfrac{1}{2} M(b - \widehat{b})^2 - \tfrac{1}{2} n \log(2\pi),$$

where $Q_0 = \sum_{i=1}^n (Y_i - \widehat{b} x_i)^2$.

(a) Show that $\widehat{b} \sim \mathrm{N}(b, 1/M)$, and use this to put up a confidence distribution for b.

(b) Suppose x_{new} is the covariate associated with a new and not yet observed Y_{new}, which by assumptions has the $\mathrm{N}(bx_{\mathrm{new}}, 1)$ distribution. Find a confidence distribution for $\mathrm{E}\, Y_{\mathrm{new}} = bx_{\mathrm{new}}$.

(c) Find also a predictive confidence distribution for Y_{new} itself.

(d) One is interested in when the regression line $y = bx$ crosses a certain threshold y_0 (a known non-zero number), i.e. in $x_0 = y_0/b$. Work out an expression for the profiled log-likelihood function $\ell_{\mathrm{prof}}(x_0)$, and then show that the deviance function may be expressed as

$$D(x_0) = M(y_0/x_0 - \widehat{b})^2.$$

What is the distribution of $D(x_0)$, under the true value of x_0?

(e) Construct a confidence curve $\mathrm{cc}(x_0)$ for x_0. Compute its value at x_0 being equal to infinity, and comment on this.

(f) Explain briefly which modifications ought to be carried out in the case where the error standard deviation σ is not known.

Exercise 2

Suppose one observes data points $(x_{\rm obs}, y_{\rm obs})$ from the model where $X \sim \mathrm{N}(\theta, 1)$ and $Y \sim \mathrm{N}(\theta + \delta, 1)$ are independent. The parameters θ and δ are unknown, and primary interest lies with the difference parameter δ.

(a) What is the distribution of $D = Y - X$? Put up a confidence distribution for δ based on D.

(b) Another construction is

$$C^*(\delta) = \mathrm{Pr}_\delta\{Y \ge y_{\rm obs} \,|\, Z = z_{\rm obs}\},$$

where $Z = X + Y$ and $z_{\rm obs} = x_{\rm obs} + y_{\rm obs}$. Derive a clear formula for this $C^*(\delta)$, and show that it indeed is a confidence distribution.

(c) Use theory from the Schweder and Hjort book to find the power optimal confidence distribution for δ, and comment on what you find.

Exercise 3

Consider a setup with independent observations $Y_1, \ldots, Y_n$ from the exponential distribution with density $\theta \exp(-\theta y)$ for $y > 0$. This can be represented as $Y_i = V_i/\theta$ for $i = 1, \ldots, n$, where the V_i are independent and standard exponentially distributed. I also note, in case this might be useful for some of your calculations below, that $2V_i \sim \chi^2_2$, from which follows, for example, that

$$\bar V_n = \frac{\sum_{i=1}^n 2V_i}{2n} = \frac{\chi^2_{2n}}{2n}.$$

The density for the χ^2_m, should you happen to need it today, is

$$\gamma_m(x) = \frac{1}{2^{m/2}\Gamma(m/2)} x^{m/2-1} \exp(-\tfrac12 x) \quad \text{for } x > 0.$$

(a) Show that the maximum likelihood estimator is $\widehat\theta = 1/\bar Y$, where $\bar Y = n^{-1}\sum_{i=1}^n Y_i$, and that its distribution may be represented as that of $\theta/\bar V$.

(b) Construct a natural confidence distribution for θ. Also write down a formula for the resulting confidence density, and find the point of maximum confidence.

(c) Find a predictive confidence distribution for the next observation in the sequence, i.e. for data point no. $n + 1$, given that we have observed the n first values.

Exercise 4

Suppose one observes independent pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ from the bivariate normal distribution with correlation coefficient ρ. The usual estimator is of course

$$\widehat{\rho} = \frac{1}{n}\sum_{i=1}^{n} \frac{X_i - \bar{X}}{\widehat{\sigma}_x}\frac{Y_i - \bar{Y}}{\widehat{\sigma}_y},$$

in terms of averages $\bar{X}$ and $\bar{Y}$ and the usual empirical standard deviation estimates $\widehat{\sigma}_x$ and $\widehat{\sigma}_y$. This exercise concerns constructing a confidence distributions for ρ.

(a) It is known from standard multivariate normal theory that $\widehat{\rho}$ is *approximately* normal, with standard deviation *approximately* $(1-\rho^2)/\sqrt{n}$, i.e.

$$\widehat{\rho} \approx \mathrm{N}(\rho, (1/n)(1-\rho^2)^2).$$

Construct an approximate confidence distribution for ρ based on this.

(b) Explain one or two ways in which a more accurate confidence distribution can be constructed.

(c) Indicate briefly how one may go about constructing a confidence distribution for the correlation coefficient without assuming a binormal distribution for the pairs.