UNIVERSITETET I OSLO

Matematisk Institutt

EXAM IN:	**STK 4011/9011 – Statistical Inference Theory**
	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, 2–18/xii/2014;**
	Part II: 8/xii s.y., 9:00–13:00, written exam

This exam set contains four exercises and comprises three pages.

Exercise 1

We shall work with the so-called geometric distribution, defined by

$$f(x,\theta) = (1-\theta)^{x-1}\theta \quad \text{for } x = 1, 2, 3, \ldots,$$

with θ an unknown parameter in $(0, 1)$. This is the distribution of how many independent experiments one needs to carry out until a certain event occurs, with θ denoting the probability of this event. Its mean and variance are $1/\theta$ and $(1-\theta)/\theta^2$ (which you do not need to prove here). Assume in what follows that $X_1, \ldots, X_n$ are independent observations stemming from this distribution.

(a) Letting $\bar X_n = n^{-1}\sum_{i=1}^n X_i$ the the sample average, use the central limit theorem to identify the limit distribution of $\sqrt{n}(\bar X_n - 1/\theta)$.

(b) Then use the delta method to find the limit distribution of $\sqrt{n}(1/\bar X_n - \theta)$.

(c) If $\widetilde\theta$ is an unbiased estimator of θ, show via results from the curriculum that

$$\operatorname{Var}\widetilde\theta \ge \frac{\theta^2(1-\theta)}{n}.$$

Compare this with your result from (b), and comment.

(d) Now assume that uncertainty about the unknown parameter θ is quantified in terms of a prior density $\pi(\theta)$, and that this prior is the uniform on $(0, 1)$. Find the posterior distribution, $\pi(\theta \,|\, \text{data})$. Also show that the posterior mean becomes

$$\theta^* = \frac{n+1}{n\bar X_n + 2}.$$

(e) Going back to the ordinary non-Bayesian framework, with observations $X_1, X_2, \ldots$ seen as independent from the geometric distribution above with some underlying unknown but fixed θ, show that $\sqrt{n}(\theta^* - \theta)$ has the same limit distribution as that found under (b).

Exercise 2

Assume that independent observations $X_1, \ldots, X_n$ come from the exponential distribution with parameter θ, i.e. with density and cumulative function

$$f(x, \theta) = \theta e^{-\theta x} \quad \text{and} \quad F(x, \theta) = 1 - e^{-\theta x} \quad \text{for } x > 0,$$

where θ is an unknown positive parameter. The mean and variance of X_i is $1/\theta$ and $1/\theta^2$. In the following you may also use without having to prove it here the fact that a sum $T = X_1 + \cdots + X_n$ of n such variables has a Gamma distribution with parameters (n, θ), with density

$$g_n(t) = \frac{1}{\Gamma(n)} \theta^n s^{n-1} e^{-\theta t} \quad \text{for } t > 0.$$

(As usual $\Gamma(\cdot)$ is the gamma function, and for integer values, $\Gamma(n) = (n-1)!$.)

(a) Show that $T = \sum_{i=1}^n X_i$ is sufficient, and explain what this property means.

(b) Show that the maximum likelihood estimator for θ is

$$\widehat{\theta} = \frac{n}{T} = \frac{1}{\bar{X}},$$

with $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ as usual denoting the sample mean.

(c) Show that

$$\widehat{\lambda} = \frac{1}{\widehat{\theta}} = \frac{T}{n}$$

is the uniformly minimum variance unbiased estimator of the parameter $\lambda = 1/\theta$.

(d) What is the maximum likelihood estimator of $F(x, \theta)$, for a given x?

(e) Give an expression for the joint density $g(x_1, s)$ of the two variables (X_1, S), where $S = X_2 + \cdots + X_n$. Use this to work out the joint density $h(v, t)$ for the two variables (V, T), where T is as above and $V = X_1/T$. Show in particular from this that

$$P(V \le v) = 1 - (1 - v)^{n-1} \quad \text{for } 0 \le v \le 1.$$

(f) Show that the simple variable

$$\widetilde{F}(x) = I\{X_1 \le x\} = \begin{cases} 1 & \text{if } X_1 \le x, \\ 0 & \text{if } X_1 > x \end{cases}$$

is an unbiased estimator for $F(x, \theta)$. Then find a formula for the uniformly minimum variance unbiased estimator of $F(x, \theta)$.

(g) Show that

$$\mathrm{E}\{\tfrac{1}{2}(X_1 - X_2)^2 \,|\, \bar{X}\} = \frac{n}{n+1} \bar{X}^2.$$

Exercise 3

We continue with the situation given in the above exercise, with independent observations $X_1, \ldots, X_n$ from the exponential model.

(a) One can easily show that the variables $Z_i = 2\theta X_i$ have the χ^2_2 distribution (and you do not need to prove this here). Show from this that

$$\widehat{\theta} \quad \text{has the distribution of} \quad \theta \frac{2n}{\chi^2_{2n}}.$$

(b) Construct a 99% confidence distribution for θ based on the above.

(c) Set up the likelihood ratio test for testing $H_0\colon \theta = 1$ versus the alternative that $\theta \neq 1$, with significance level ('type I error') equal to 0.01. Work with the algebra to arrive at a clear testing recipe.

Exercise 4

As explained in the course, we say that a sequence of random variables $A_1, A_2, A_3, \ldots$ converges in probability to a constant a, and write $A_n \to_{\rm pr} a$ to indicate this, if $P(|A_n - a| \geq \varepsilon) \to 0$ for each positive ε. If the A_n in question is seen as an estimator of the parameter a, then we say that A_n is consistent for a.

(a) Show that if $A_n \to_{\rm pr} a$ and $B_n \to_{\rm pr} b$, then indeed $A_n + B_n \to_{\rm pr} a + b$.

(b) Suppose again that $A_n \to_{\rm pr} a$. Show that if h is a continuous function, defined in at least a neighbourhood around a, then $h(A_n) \to_{\rm pr} h(a)$.

(c) Assume $X_1, X_2, \ldots$ is a sequence of independent random variables with the same distribution, with finite mean ξ and variance σ^2. Show, e.g. from the Chebyshov inequality (неравенство Чебышёva), that the sample average $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ is consistent for ξ. (One may actually prove this also without the finite variance assumption, but then the proof becomes harder. The natural sufficient condition is simply that $\mathrm{E}\,|X_i|$ is finite.)

(d) Use the above to demonstrate that

$$\widehat{\sigma}^2_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \to_{\rm pr} \sigma^2.$$

Finally show that $\widehat{\sigma}_n$ is a consistent estimator of σ.