UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

Examination in:	STK9011 — Statistical Inference Theory.
Day of examination:	Monday December 16th 2013.
Examination hours:	09.00 – 13.00.

This problem set consists of 3 pages.

Appendices:	Selected definitions and theorems from Casella & Berger. Table of common distributions from Casella & Berger.
Permitted aids:	Approved calculator.

Please make sure that your copy of the problem set is complete before you attempt to answer anything.

Problem 1

Assume that $X_1, X_2, \ldots, X_n$ are iid and geometrically distributed with

$$f(x|p) = \mathrm{P}(X_i = x|p) = (1-p)^{x-1}p; x = 1, 2, \ldots$$

where $0 < p \leq 1$. Then $\mu = \mathrm{E}[X_i] = \frac{1}{p}$ and $\sigma^2 = \mathrm{Var}(X_i) = \frac{1-p}{p^2}$ (you are not asked to show this).

a) Demonstrate that the family of geometric distributions belongs to the exponential family of distributions.

   Show, based on $\mathbf{X} = (X_1, \ldots, X_n)$, that the statistic $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ is sufficient and complete for the parameter p by using characteristics of the exponential families of distributions.

b) Demonstrate that the maximum likelihood estimator (MLE) for p is given by $\hat{p} = \frac{n}{\sum_{i=1}^{n} X_i} = 1/\bar{X}$.

   Specify the MLE's of μ and σ^2 and denote these by $\hat{\mu}$ and $\hat{\sigma}^2$.

   Decide which of the MLE's $\hat{p}, \hat{\mu}$ and $\hat{\sigma}^2$ that are unbiased estimators? Justify your answers.

c) Demonstrate that the MLE $\hat{\mu}$ for μ is UMVUE for μ by using the Cramér-Rao inequality.

(Continued on page 2.)

d) Define an estimator for p based on only the first random variable X_1 given by

$$\tilde{p} = I(X_1 = 1) = \begin{cases} 1 & \text{if } X_1 = 1 \\ 0 & \text{otherwise} \end{cases}$$

Show that $\tilde{p}$ is unbiased.

Develop an expression for an unbiased estimator of p based on the full sample $\mathbf{X} = (X_1, \ldots, X_n)'$ using the Rao-Blackwell theorem. You can use without proof that $T = \sum_{i=1}^{n} X_i$ has a negative binomial distribution with $\mathrm{P}(T = t) = \binom{t-1}{n-1}(1-p)^{t-n}p^n$ for $t = n, n+1, \ldots$.

Is this unbiased estimator a uniform minimum variance unbiased estimator (UMVUE)? Justify your answer.

e) Specify the asymptotic distributions of (i) $\sqrt{n}(\hat{p} - p)$, (ii) $\sqrt{n}(\hat{\mu} - \mu)$ and (iii) $\sqrt{n}(\hat{\sigma}^2 - \sigma^2)$ when $n \to \infty$.

How can you estimate the large sample variances of $\hat{p}, \hat{\mu}$ and $\hat{\sigma}^2$?

f) Consider the hypothesis

$$\mathrm{H}_0 : p = p_0 \text{ versus } \mathrm{H}_1 : p \neq p_0$$

for a fixed value $0 < p_0 \leq 1$.

Specify the rejection region for a level α asymptotic likelihood ratio test (LRT). Justify the answer.

Specify the corresponding asymptotic $(1 - \alpha)$ confidence region for the parameter p. Justify the answer.

Problem 2

The Poisson distribution with unknown parameter $\lambda > 0$ for a random variable X is given by the probability mass function

$$f(x|\lambda) = \mathrm{P}(X = x|\lambda) = \frac{\lambda^x}{x!}\exp(-\lambda) \text{ for } x = 0, 1, 2, \ldots$$

Assume that a vector of random variables $\mathbf{X} = (X_1, X_2, X_3)'$ where the X_i are independent and

$$\begin{aligned} X_1 &\sim f(x|\lambda) \\ X_2 &\sim f(x|2\lambda) \\ X_3 &\sim f(x|3\lambda) \end{aligned}$$

Note that $\mathrm{E}[X_i] = i\lambda$ are different for $i = 1, 2, 3$.

a) Identify a sufficient statistic for λ based on $\mathbf{X}$ by using the factorization theorem.

Develop an expression for the maximum likelihood estimator (MLE) for λ, denote it by $\hat{\lambda}$.

(Continued on page 3.)

Show that the estimator $\hat{\lambda}$ is a uniform minimum variance unbiased estimator (UMVUE) for λ using the Cramér-Rao inequality.

We now assume the prior for λ given by the density $\pi(\lambda|\alpha,\beta) = \frac{\lambda^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha}\exp(-\frac{\lambda}{\beta})$, i.e. λ is gamma-distributed with parameters (α,β) where $\alpha > 0, \beta > 0$.

b) Given observed values $\mathbf{x}$ of the random variable $\mathbf{X}$ develop the Bayes estimate and its associated estimation variance

$$\begin{aligned} \tilde{\lambda} &= \mathrm{E}[\lambda|\mathbf{x}] \\ \tilde{\sigma}^2 &= \mathrm{Var}[\lambda|\mathbf{x}] \end{aligned}$$

Discuss the behaviour of $\tilde{\lambda}$ in comparison to the prior distribution when

$$\begin{array}{lll} (i) & \alpha \to 0, \beta \to \infty & \text{and } \alpha\beta = \mu \\ (ii) & \alpha \to \infty, \beta \to 0 & \text{and } \alpha\beta = \mu \end{array}$$

where $0 < \mu < \infty$ is a constant. Comment on your answer.

Problem 3

Assume that $X_i, i = 1, \ldots, n$ are iid and exponentially distributed with density $f(x;\lambda) = \lambda\exp(-\lambda x)$. You can use without proof that the log-likelihood of $\mathbf{X} = (X_1, \ldots, X_n)$ equals $l_X(\lambda) = n\log(\lambda) - \lambda\sum_{i=1}^n X_i$.

The X_i's are not observed directly. Rather one observes, for given z_i,

$$Y_i = I(X_i \leq z_i) = \begin{cases} 1 & \text{if } X_i \leq z_i \\ 0 & \text{if } X_i > z_i \end{cases}.$$

The Y_i are thus indicators of whether $X_i \leq z_i$ and can be considered as incompletely observed data from complete data $\mathbf{X}$. Such data are sometimes referred to as "current status data" when the X_i's are times to some event and one inspects at time z_i whether the event has occurred or not. The observed data are summarized as $(z_1, Y_1), \ldots, (z_n, Y_n)$.

a) Find an expression for the likelihood $L_Y(\lambda)$ for the observed current status data $\mathbf{Y} = \{(z_1, Y_1), \ldots, (z_n, Y_n)\}$.

Describe the EM-algorithm for obtaining the MLE of λ based on $\mathbf{Y}$ using the complete data log-likelihood $l_X(\lambda)$.

In particular find the conditional expectations $\mathrm{E}[X_i|X_i \leq z_i]$ and $\mathrm{E}[X_i|X_i > z_i]$.

END