Data provided:
"Statistics Tables" by H.R. Neave

MAS274

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2011-2012**

Statistical Reasoning **2 hours**

RESTRICTED OPEN BOOK EXAMINATION.
Candidates may bring to the examination lecture notes and associated lecture material (but no textbooks) plus a calculator that conforms to University regulations.
Attempt all questions.[Q1 18, Q2 32, Q3 23, Q4 17] Total marks 90.

1 Suppose $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ are independent Gaussian variables, with $\mathrm{E}[X_i] = a_i \mu$ and $\mathrm{Var}[X_i] = \frac{1}{2}$, where $\{a_1, \ldots, a_n\}$ are known constants.

(a) (i) Write down the likelihood for μ. ***(3 marks)***

(ii) Find a sufficient statistic and the MLE for μ. ***(5 marks)***

(b) Assume a priori $\pi(\mu) = \mathrm{N}\left(\mu \mid m, \frac{1}{v}\right)$, with $m \in \mathbb{R}$ and $v > 0$.

(i) Calculate the posterior distribution of the parameter. Give the posterior mean and variance. ***(7 marks)***

(ii) Provide a 90% highest posterior density interval for μ. ***(3 marks)***

2 A recent study shows that the concentration of a toxic substance in the soil, X, is distributed uniformly between zero and an upper bound, θ, which is related to the PH of the soil.

(a) Write down the density function of X, identifying clearly its support and the parameter space. Find $\mathrm{E}[X \mid \theta]$. ***(3 marks)***

(b) A surveyor is commissioned to provide measurements of the substance in a patch of land. A random sample $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ is recorded, with each X_i following the same uniform distribution on $[0, \theta]$.

(i) Write down the likelihood and prove that the MLE of θ is the maximum observed value in the sample. Is the MLE a sufficient statistic? ***(5 marks)***

(ii) Is the MLE unbiased? If not, propose an unbiased estimator based on the MLE. ***(6 marks)***

[Hint: If $\boldsymbol{X}$ is a random sample and $T(x) = \max\{X_1, \ldots, X_n\}$, then $F_T(t) = \prod_{i=1}^{n} \mathrm{P}[X_i \leq t]$].

(c) An advisor to the council has asked you to consider the Method of Moments estimator, $\tilde{\theta} = 2\bar{X}$, with $\bar{X}$ the sample mean, as an alternative to the MLE. Based on their MSE, which of these two estimators would you recommend to be used in the report to the council? ***(7 marks)***

(d) A soil expert is consulted so we can use his opinion as a prior in a Bayesian analysis. His description is encoded in the prior distribution

$$\pi(\theta) = b^2\theta^{-2} \; ; \qquad \theta \geq b, b > 0$$

(i) The expert also says that b is very close to zero, in particular it is smaller than any observed concentration of the toxic substance. Find the posterior distribution using the expert's opinion. ***(6 marks)***

(ii) The council advisor believes that under- and over-estimating the parameter have similar consequences and thus would like you to use a quadratic loss function to estimate the parameter. Is this estimator unbiased? ***(5 marks)***

3 Suppose X follows an exponential distribution with $\mathrm{E}[X \mid \lambda] = 1/\lambda$.

(a) (i) Show that the k-unit likelihood region for λ, given an observation x, can be written as

$$R_k = \{\lambda \,:\, \lambda x - \log \lambda \le c\,\}$$

for some value c. *(4 marks)*

(ii) If $x = 1$ and $k = 2$, find the value of c. Sketch the likelihood region. *(3 marks)*

(b) A random sample of size n is available.

(i) Write down the MLE for λ. *(3 marks)*

(ii) Prove that the uniformly most powerful test for $\mathcal{H}_0 \equiv \{\lambda = \lambda_0\}$ against $\mathcal{H}_1 \equiv \{\lambda > \lambda_0\}$ rejects $\mathcal{H}_0$ if the MLE is sufficiently large. *(4 marks)*

(iii) Write down the asymptotic distribution of the MLE of $\phi = \sqrt{\lambda}$. *(4 marks)*

(iv) Give an approximate 90% critical region for testing $\mathcal{H}_0 \equiv \{\phi = 1\}$ against $\mathcal{H}_1 \equiv \{\phi \le 1\}$. *(5 marks)*

4 The vector $(X, Y) \in \mathbb{R}^2$ has the joint density

$$f(x, y \mid \rho) = \frac{\sqrt{1-\rho^2}}{2\pi} \exp\left[-\frac{x^2 - 2\rho xy + y^2}{2}\right]$$

where the unknown parameter, $\rho \in (-1, 1)$ measures the correlation between X and Y. Based on a random sample $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$:

(a) Show that the log-likelihood can be written as

$$\ell(\rho; S) = \frac{n}{2}\log(1-\rho^2) + nS\rho + h(\boldsymbol{x}, \boldsymbol{y}) \ ,$$

where $h(\boldsymbol{x}, \boldsymbol{y})$ does not involve the parameter and with $S = 1/n \sum_{i=1}^{n} X_i Y_i$.

(3 marks)

(b) Show that the MLE, $\hat{\rho}$, satisfies

$$\hat{\rho}^2 + \frac{\hat{\rho}}{S} - 1 = 0 \ ;$$

and show that there is only one solution satisfying $\hat{\rho} \in (-1, 1)$.

(5 marks)

(c) Show that the asymptotic variance of the MLE is

$$\frac{1}{n}\frac{\left(1-\rho^2\right)^2}{1+\rho^2}$$

(4 marks)

(d) Suppose that $n = 300$ and $\sum x_i y_i = -60$. Give an approximate confidence interval of size 95% for ρ. *(5 marks)*

End of Question Paper