The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2012–2013**

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 32, Q2 17, Q3 23, Q4 28] Total marks 100.

1 Assume $\{X_1, \ldots, X_n\}$ is a random sample from a uniform distribution on $[\theta, 1]$, with $\theta \in (-\infty, 1)$.

(a) Prove that $\tilde{\theta} = 2\bar{X} - 1$ is an unbiased estimator of θ and calculate its MSE. ***(5 marks)***

(b) (i) Write down the likelihood function of θ based on the data, making explicit the region where this function is non-zero. ***(4 marks)***

(ii) Prove that the MLE, $\widehat{\theta} = X_{(1)} = \min\{X_1, \ldots, X_n\}$. ***(3 marks)***

(iii) Calculate the sampling distribution of the MLE. ***(7 marks)***

[HINT: The cumulative density function of the minimum, $X_{(1)}$, of a random sample of size n is

$$F_{X_{(1)}}(t) = 1 - (1 - F_X(t))^n$$

with $F_X(t) = \mathrm{P}[X \leq t]$

(iv) Prove that the expected value of the MLE is $\frac{n\theta+1}{n+1}$. ***(7 marks)***

(v) Given that variance of the MLE is

$$\mathrm{Var}[\widehat{\theta} \mid \theta] = \frac{(1-\theta)^2 n}{(n+1)^2(n+2)} .$$

Which of $\widehat{\theta}$ or $\tilde{\theta}$ is preferred in MSE? ***(6 marks)***

2 Let $\{X_1, \ldots, X_n\}$ be Gaussian and independent with $\mathrm{E}[X_i] = \mu_i$ and $\mathrm{Var}[X_i] = \frac{a_i}{\lambda}$ and assume that $\{\mu_1, \ldots, \mu_n\}$ and $\{a_1, \ldots, a_n\}$ are known constants.

[Additional information: Let $X \sim \mathrm{N}(x \mid 0, 1)$ then $\mathrm{P}[X \geq 1.282] = 0.1$, $\mathrm{P}[X \geq 1.645] = 0.05$, $\mathrm{P}[X \geq 1.956] = 0.025$].

(a) Write down the distribution of $Y_i = X_i - \mu_i$. ***(2 marks)***

(b) Write down the likelihood of λ, based on $\boldsymbol{Y} = \{Y_1, \ldots, Y_n\}$ and find the MLE. ***(5 marks)***

(c) (i) We wish to test $\mathcal{H}_0 \equiv \{\lambda = \lambda_0\}$ v $\mathcal{H}_1 \equiv \{\lambda > \lambda_0\}$. Prove that the most powerful test of size α rejects $\mathcal{H}_0$ if

$$T(\boldsymbol{Y}) = \sum_{i=1}^{n} \frac{Y_i^2}{a_i}$$

is small enough. ***(5 marks)***

(ii) In a particular sample of size $n = 100$, we obtained $T(\boldsymbol{y}) = 40$. Provide an approximate 90% confidence interval for λ. Let $\lambda_0 = 2$, does the interval provides evidence in favour of $\mathcal{H}_0$? ***(5 marks)***

3 An engineer is in charge of the production line of spare parts for a certain car. To monitor the process he inspects every part produced and stops the process when exactly r defective parts are observed. Then the machine is re-calibrated taking into account the probability of observing a defective part.

The engineer uses the following probabilistic model for reporting to quality control:

$$f(x \mid \theta, r) = \binom{x+r-1}{x} \theta^r (1-\theta)^x \ , x = 0, 1, \ldots \ ; \quad 0 < \theta < 1 \ .$$

where x = the number of non-defective parts in the sample before observing r defective, and $r \geq 1$ is a fixed integer.

[Additional information: Let $X \sim \mathrm{N}(x \mid 0, 1)$ then $\mathrm{P}[X \geq 1.282] = 0.1$, $\mathrm{P}[X \geq 1.645] = 0.05$, $\mathrm{P}[X \geq 1.956] = 0.025$].

(a) Show that $r/(r+x)$ is the MLE of θ. ***(3 marks)***

(b) The department of quality control has decided to set $r = 5$. In a particular day the engineer observes $x = 15$.

(i) Show that the 2-unit likelihood region for θ can be written as

$$R_2 = \{\theta \ : \ a \log \theta + b \log(1-\theta) \leq c \} \ ,$$

and calculate the values for a, b and c. ***(5 marks)***

(ii) An expert's opinion is used to conduct a Bayesian analysis for θ. The expert states that his prior beliefs can be described by a Beta distribution with mean $1/2$ and variance $1/8$. Write down the expert's posterior distribution and calculate its mean. ***(8 marks)***

(c) The machine is re-calibrated if there is enough evidence to suggest that the ratio $\psi = \frac{1}{\theta} < 3$. Use an approximate confidence interval of size 0.9 for ψ to decide whether the machine should be calibrated with the data as in part (b). ***(7 marks)***

4 A chemist is interested in estimating the mean current produced by a chemical reaction. To this end, she conducts two sets of independent measurements, using two different apparatuses. The first device measures the current without bias and has unitary precision, while the second one always yields a doubled measurement but is twice as precise as the first one. After consulting with a statistician, she decides to model the data as follows: Let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ denote the first set of measurements and $\boldsymbol{Y} = \{Y_1, \ldots, Y_k\}$ the second set. Assume,

$$X_i \sim \mathrm{N}(x_i \mid \mu, 1) \qquad \text{and} \qquad Y_j \sim \mathrm{N}\left(y_j \mid 2\mu, \frac{1}{2}\right) ;$$

all independent and with μ the mean current.

[Additional information: Let $X \sim \mathrm{N}(x \mid 0, 1)$ then $\mathrm{P}[X \geq 1.282] = 0.1$, $\mathrm{P}[X \geq 1.645] = 0.05$, $\mathrm{P}[X \geq 1.956] = 0.025$].

(a) (i) Using these assumptions write down the likelihood for μ. *(2 marks)*

(ii) Prove that the MLE is given by

$$\widehat{\mu} = \frac{n\bar{x} + 4k\bar{y}}{n + 8k} .$$

(4 marks)

(b) For the particular experiment conducted, the following statistics were obtained:

$$n = 8 , \quad \bar{x} = \frac{1}{4} , \qquad k = 4 , \quad \bar{y} = \frac{1}{2} .$$

(i) The chemist is interested in whether the mean current is zero or not. Provide a test of size $\alpha = 0.05$ using the data at hand. *(5 marks)*

(ii) The statistician has proposed

$$\tilde{\mu} = \frac{n\bar{x} + k\bar{y}}{n + 2k}$$

as alternative estimator. Which estimator would you recommend the chemist to use and why? *(6 marks)*

 Question 4 continued on next page

4 (continued)

(c) Assuming you can write the likelihood as

$$\mathrm{L}(\mu; \boldsymbol{x}, \boldsymbol{y}) \propto \exp\left[-\frac{p}{2}(\mu - \widehat{\mu})^2\right]$$

with $p = n + 8k$, and that 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. Write down the posterior mean and variance. ***(6 marks)***

(ii) Using her prior knowledge, the chemist sets $m = 2$ and $v = 2$. Provide a posterior probability interval of size 0.95 and comment on whether the data supports the hypothesis that the mean current is zero. ***(5 marks)***

End of Question Paper