The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2013–2014**

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 25, Q2 25, Q3 25, Q4 25] Total marks 100.

1 Let X be a *single observation* from the density

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

The test "Reject $\mathcal{H}_0$ if and only if $X \geq 1/2$" is used to contrast $\mathcal{H}_0 \equiv \{\theta \leq 1\}$ against $\mathcal{H}_1 \equiv \{\theta > 1\}$.

(a) Find the size of the test. ***(3 marks)***

(b) Sketch the power function. ***(3 marks)***

(c) Assume that a random sample $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ is available

(i) Prove that

$$\hat{\theta} = \frac{-n}{\sum_{i=1}^{n} \log X_i}$$

is the MLE. ***(6 marks)***

(ii) Prove that a confidence interval for θ of approximate level γ is given by

$$\left[\frac{\hat{\theta}}{1 + \frac{z_{(1+\gamma)/2}}{\sqrt{n}}}, \frac{\hat{\theta}}{1 - \frac{z_{(1+\gamma)/2}}{\sqrt{n}}} \right]$$

where z_α is the α-th quantile of a standard Normal distribution. ***(8 marks)***

(iii) The random sample

$$\boldsymbol{x} = \left\{ \begin{array}{cccccccc} 0.88 & 0.93 & 0.84 & 0.39 & 0.73 & 0.54 & 0.75 & 0.99 \\ 0.64 & 0.80 & 0.66 & 0.29 & 0.61 & 0.92 & 0.34 & \end{array} \right\}$$

is observed. Using the approximate confidence interval above, is there evidence to support $\mathcal{H}_0 \equiv \{\theta \leq 1\}$?

[*HINT*: $\sum_{i=1}^{15} \log x_i = -6.524$.]

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

(5 marks)

2 As a member of a chess club, I have played many tournaments in the past. In each tournament, I was allowed to progress and play one further match if and only if I had won the previous one. I recorded how many games I won in each tournament as x_i.

Using a random sample from my records, $\boldsymbol{x} = \{x_1, \ldots, x_n\}$, I would like to make inferences about my probability of wining a single game. I think that a Geometric distribution,

$$f(x_i \mid \delta) = \delta\,(1-\delta)^{x_i} \;; \quad x_i = 0, 1, 2, \ldots \;; \quad 0 < \delta < 1 \;,$$

can be used in this situation.

(a) Describe two assumptions implied in using this model. And explain the role of δ. ***(5 marks)***

(b) Show that the k-unit likelihood region for δ can be expressed as

$$R_k = \{\delta \;:\; A \log \delta + B \log(1-\delta) \geq C - k\}$$

and provide the values for A, B and C. ***(6 marks)***

(c) Prior to starting the tournament my beliefs about δ can be described using a Beta distribution with mean $1/2$ and variance $1/12$.

(i) Show that the posterior distribution of δ after n tournaments is $\text{Be}(\delta \mid a, b)$ and give explicit values for a and b. ***(6 marks)***

(ii) I am interested in, $\psi = \frac{1-\delta}{\delta}$, my odds of winning a single game. Calculate the posterior mean of ψ if my sampled results are $\boldsymbol{x} = \{3, 4, 8, 5, 1, 5, 6, 8, 4, 8\}$. ***(8 marks)***

3 A biologist measures the diameter of two different types of cells, X and Y, as follows:

- n of the X type of cell at standard resolution, λ
- m of the Y type of cell at low resolution, $\lambda/4$.

The mean diameter of the X cells is known to be a. Y cells have known mean diameter b.

Assuming that all the measurements are independent and can be modelled as

$$X_i \sim \mathrm{N}\left(x_i \mid a, \frac{1}{\lambda}\right) \quad \text{and} \quad Y_j \sim \mathrm{N}\left(y_j \mid b, \frac{4}{\lambda}\right) .$$

(a) Show that the likelihood can be written as

$$\mathrm{L}(\lambda; \boldsymbol{x}, \boldsymbol{y}) \propto \lambda^{(m+n)/2} \exp\left[-\lambda \frac{S^2}{2}\right]$$

with

$$S^2 = \sum_{i=1}^{n}(x_i - a)^2 + \frac{1}{4}\sum_{j=1}^{m}(y_i - b)^2 .$$

(5 marks)

(b) Show that S^2 is sufficient for λ. *(3 marks)*

(c) Show that the UMP test of $\mathcal{H}_0 \equiv \{\lambda = \lambda_0\}$ versus $\mathcal{H}_1 \equiv \{\lambda = \lambda_1 > \lambda_0\}$ rejects $\mathcal{H}_0$ if S^2 is small enough. *(9 marks)*

(d) After taking $m = 20$ and $n = 60$ measurements, she obtained

$$\sum_{i=1}^{n}(x_i - a)^2 = 35 \quad \text{and} \quad \sum_{j=1}^{m}(y_i - b)^2 = 88 .$$

Provide a confidence interval of approximate level 90% for $\sigma^2 = \frac{1}{\lambda}$.

[*Additional information*: Let $X \sim \mathrm{N}(x \mid 0, 1)$ then $\mathrm{P}[X \geq 1.282] = 0.1$, $\mathrm{P}[X \geq 1.645] = 0.05$, $\mathrm{P}[X \geq 1.956] = 0.025$]. *(8 marks)*

4 Let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ be a random sample from a Gaussian distribution with mean zero and variance $2/\lambda$. The prior is a Gamma distribution,

$$\pi(\lambda) = \frac{b^a}{\Gamma[a]} \lambda^{a-1} e^{-b\lambda} ,$$

where the parameters (a, b) will be set to reflect our prior beliefs.

(a) Write down the likelihood for λ. ***(5 marks)***

(b) Prove that the posterior distribution of λ is Gamma with parameters $(a^\star, b^\star)$ and give explicit values for $(a^\star, b^\star)$. ***(7 marks)***

(c) (i) Suppose that we think that the prior mean is $1/2$ and the prior variance $1/4$. Find the corresponding values for a and b. ***(2 marks)***

(ii) Find the posterior mean and variance after observing the random sample $\boldsymbol{x} = \{0.41, 0.18, 0.58, 0.16, -0.74, 0.07, 0.01\}$. ***(7 marks)***

(iii) Sketch the highest posterior density interval of probability 0.95. ***(4 marks)***

End of Question Paper