The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2015–2016**

MAS472 Computational Inference **2 hours**

Candidates may bring to the examination a calculator that conforms to University regulations.
Marks will be awarded for your best ***three*** *answers. Total marks 60.*

Please leave this exam paper on your desk
Do not remove it from the hall

Registration number from U-Card (9 digits)
to be completed by student

Blank

1 (i) Five observations of the random variable X are recorded:

$$\{13.6, 17.0, 20.2, 16.1, 15.1\}$$

(a) Sketch the empirical cumulative distribution function of X based on the observed sample. ***(3 marks)***

(b) Using the following five random draws from the $U[0,1]$ distribution

$$\{0.02, 0.54, 0.46, 0.77, 0.66\},$$

sample from the empirical cumulative distribution function to produce a single bootstrap value of the sample median of 5 observations of X. ***(4 marks)***

(c) Explain how you would estimate the standard error of the estimate of the median of X. ***(3 marks)***

(ii) The waiting times, denoted $t_1, \ldots, t_n$, between arrivals in a queue are claimed to be independent and to follow an exponential distribution with mean 1. The observed times are noted to be very similar, and it is suspected that the claim of independence may be false.

(a) Assuming that arrival times really do have an exponential distribution, explain how a Monte Carlo test of size 0.05 could be conducted to test the hypothesis of independence, using the sample variance as a test statistic ***(7 marks)***

(b) What is the minimum number of random test statistics required to ensure a size of precisely 0.05? Why would it not be advisable to generate only this minimum number? ***(3 marks)***

2 (i) Let X be a positive random variable with probability density function given by

$$g(x) = \begin{cases} \dfrac{2}{\pi(1+x^2)} & \text{for } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

(a) By using the substitution $x = \tan u$ or otherwise, show that the cumulative distribution function (CDF) of X is

$$G(x) = \begin{cases} \dfrac{2}{\pi}\tan^{-1}(x) & \text{for } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

(5 marks)

(b) Use the inversion method to generate 3 random realisations of X using the uniform random numbers

$$U_1 = 0.2256, U_2 = 0.6342, U_3 = 0.1255.$$

(3 marks)

(c) The 1-2 inverted beta distribution has probability density function

$$f(x) = \begin{cases} \dfrac{2}{(1+x)^3} & \text{for } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Explain how to use rejection sampling to generate random variables from this distribution using a sequence of $U[0,1]$ random variables $U_1, \ldots, U_n$, and using $g(x)$ as the envelope function. ***(8 marks)***

(ii) We are asked to conduct a Bayesian analysis to find

$$Q = \mathbb{E}(h(X)|y).$$

The posterior distribution $\pi(x|y)$ is not known analytically, but we can calculate it up to a constant of proportionality

$$\pi(x|y) \propto \pi(y|x)\pi(x).$$

Explain how to estimate Q using importance sampling, using samples drawn randomly from the prior distribution, i.e., using

$$X_i \sim \pi(x).$$

Give the expression for your importance sampling estimator of Q, being sure to carefully specify the importance weights.

(4 marks)

3 (i) If U is uniformly distributed over $(0, 1)$ then $x = -\ln U$ has the exponential distribution with density $f(x) = \mathrm{e}^{-x}$ on $(0, \infty)$.

(a) If exponentials $X_1, X_2, \ldots, X_{2n}$ are generated as above from independent uniform variables $U_1, U_2, \ldots, U_{2n}$, derive the variance of the sample mean $\bar{X}_{2n} = \frac{1}{2n} \sum_{i=1}^{2n} X_i$. ***(3 marks)***

(b) Letting $Y_i = -\ln(1 - U_i)$ and $\bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$, prove that the variance of $\frac{\bar{X}_n + \bar{Y}_n}{2}$ is $\frac{I}{2n}$, where

$$I = \int_0^1 \ln(x) \ln(1 - x) \, \mathrm{d}x.$$

(6 marks)

(c) Given that $I = 2 - \pi^2/6$, show that $(\bar{X}_n + \bar{Y}_n)/2$ is a better estimator of the mean of the exponential distribution than $\bar{X}_{2n}$, and explain what type of procedure this is an example of.

(3 marks)

(ii) Survival times for 4 mice who took an experimental drug are recorded as $\{6, 6, 10, 20\}$ days. A Weibull distribution with probability density function

$$f_T(t) = \alpha\beta(\beta t)^{\alpha-1} \exp(-(\beta t)^\alpha)$$

is fitted to these data. The maximum likelihood estimators are $\hat{\alpha} = 2.0$ and $\hat{\beta} = 0.08$.

(a) Derive the profile log-likelihood function, $l_p(\alpha)$, for α.

(5 marks)

(b) By considering the profile deviance function, test the null hypothesis that $\alpha = 1$. You may assume that $l_p(\hat{\alpha}) = -12.2$, and that

$$\chi_1^2(0.95) = 3.84, \quad \chi_2^2(0.95) = 5.99, \quad \chi_3^2(0.95) = 7.82.$$

(3 marks)

4 The life times of light bulbs are modelled as exponential random variables with mean $1/\lambda$. Two separate experiments are performed to estimate λ. In the first, n bulbs are tested until they all fail, and the failure time t_i of each bulb $i = 1, 2, \dots, n$ is recorded. In the second experiment, m bulbs are tested for h hours, and r, the number of bulbs failing by the end of the experiment is recorded. The $n + m$ life times may be assumed to be statistically independent.

We will now use the EM algorithm to estimate λ. Let $S_1, \dots, S_m$ be the unrecorded failure times of the bulbs in the second experiment.

(i) Show that the likelihood of the completed data $y = (t_1, \dots, t_n, S_1, \dots, S_m)$ is

$$(m+n)\log\lambda - \lambda\left(\sum_{i=1}^{n} t_i + \sum_{j=1}^{m} S_j\right).$$

(3 marks)

(ii) By considering

$$F(t) = \mathbb{P}(S \leq t | S \leq h)$$

for $t \leq h$, prove that

$$\mathbb{E}(S|S \leq h, \lambda) = \frac{1}{\lambda} - \frac{h\mathrm{e}^{-\lambda h}}{1 - e^{-\lambda h}}.$$

(7 marks)

(iii) Derive an expression for the quantity

$$Q(\lambda|\lambda^{(r)})$$

used in the E-step of the EM algorithm. ***(7 marks)***

(iv) Use the M-step to derive a formula for the next estimate of λ, denoted by $\lambda^{(r+1)}$, in terms of $\lambda^{(r)}$. ***(3 marks)***

End of Question Paper