The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2019–2020**

MAS472 Computational Inference **Approximately 1.5 hours**

This is an open book exam.
Answer ***all*** *questions.*
The submission deadline is 10 am (BST), twenty-four hours after it is released. Late submission will not be considered without extenuating circumstances. It is expected that you will be able to complete this exam in approximately one and a half hour and it is recommended that you submit the work within four hours. You will not be penalised for taking longer, however.
Unless it is explicitly stated otherwise, it is intended that calculations are performed by hand (possibly with the aid of a calculator). To gain full marks, you will need to show your working. You will not get full marks if you simply write down output from a computer package.
By uploading your solutions you declare that your submission consists entirely of your own work, that any use of sources or tools other than material provided for this module is cited and acknowledged and that no unfair means have been used.

1 Let $X \sim Ray(1)$ be a random variable from a Rayleigh distribution with scale parameter equal to 1, which has the probability density function

$$f_X(x) = \begin{cases} x \exp\left(-\frac{x^2}{2}\right) & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Explain how you can generate a single sample of X using a random number $U \sim Unif(0, 1)$. ***(3 marks)***

(ii) Let $Y \sim Exp(1)$ be a random variable from the Exponential distribution with rate parameter equal to 1, which has the probability density function

$$g_Y(x) = \begin{cases} \exp(-y) & \text{for } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Explain how to use rejection sampling to generate a single random sample of X, given samples of $Y \sim Exp(1)$ and $U \sim Unif(0, 1)$.
Hint:

$$c = \sup_x \frac{f_X(x)}{g_Y(x)} \approx 2.20.$$

(3 marks)

(b) Sample a single random sample from the $Ray(1)$ distribution using rejection sampling. Use the following simulated values from the $Exp(1)$ distribution:

$$Y_1 = 0.4, \quad Y_2 = 1.2,$$

and from the $Unif(0, 1)$ distribution:

$$U_1 = 0.6, \quad U_2 = 0.3.$$

(1 mark)

(c) For a general set of two candidate values, Y_1 and Y_2, what is the probability that both would be accepted as random samples from the $Ray(1)$ distribution? ***(1 mark)***

(d) Instead of using a random draw $Y \sim Exp(1)$, would it be beneficial to generate a sample from the $Ray(1)$ distribution using a draw of $Z \sim N(0, 1)$ and a random number $U \sim Unif(0, 1)$? ***(2 marks)***

 Question 1 continued on next page

1(continued)

(iii) Suppose we are interested in the variable

$$W = h(X_1, \ldots, X_{10}) = \max_{i=1,\ldots,10} X_i,$$

where the X_i are independent and identically distributed Rayleigh variables with scale parameter equal to 1.

(a) Carefully explain how to obtain an estimate of $\mathbb{E}(W)$, using only Uniformly distributed random variables. ***(4 marks)***

(b) A random sample, $W_1, \ldots, W_{100}$, is generated of the variable W. The sample mean is calculated to be 2.39, and the sample variance is 0.259. Give an approximate 95% confidence interval for $\mathbb{E}(W)$. ***(3 marks)***

(c) If the true variance of W is given by $\mathbb{V}\text{ar}(W) = 0.242$, how many samples from a Uniform distribution are required for the confidence interval for $\mathbb{E}(W)$ to have width of less the 10^{-2}? ***(3 marks)***

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

$$\{0.02, 1.04, 3.63, 1.04, 1.67\},$$

and we are interested in the mean of X, given by $m(X)$.

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

(b) Produce a single bootstrap sample, m_1^*, of $m(X)$ by sampling from the ECDF of X, using the following five random draws from the $Unif(0,1)$ distribution:

$$\{0.04, 0.52, 0.97, 0.17, 0.63\}.$$

(3 marks)

(c) Explain the approach you would then take to perform a bootstrap estimate of the standard error of $m(X)$. ***(4 marks)***

(ii) Let X be the random variable from a $Beta(2,3)$ distribution, with probability density function

$$f(x) = \frac{x(1-x)^2}{B(2,3)},$$

for $x \in [0,1]$ and where $B(\cdot,\cdot)$ is the Beta function.

(a) Describe how you would use importance sampling to estimate $\mathbb{E}(h(X))$, where $h(x) = \exp(-x)$, using a $Gam(2,1)$ proposal distribution.

Hint: the probability density function of a $Gam(2,1)$ random variable is $g(x) = x\exp(-x)$, for $x > 0$. ***(4 marks)***

(b) This importance sampling approach is implemented to generate 100 samples $X_1, \ldots, X_{100}$. The importance weights, $w(X_i)$, are summarised by the sample variance being equal to 10^4, and

$$\sum_{i=1}^{100} w(X_i)^2 \approx 10^6.$$

Comment on the suitability of the proposal distribution in this sampling approach. ***(1 mark)***

2(continued)

(iii) Given a dataset of 10 observation pairs $(x_1, y_1), \ldots, (x_{10}, y_{10})$, the following analysis was carried out in R.

```
> head(obs)
      x     y
1  1.00 17.04
2  6.44 26.96
3 11.89 28.10
4 17.33 53.02
5 22.78 77.92
6 28.22 79.35
>
> error.lm1 <- rep(NA, 10)
> error.lm2 <- rep(NA, 10)
>
> for(i in 1:10){
+   obs.reduced <- obs[-i,]
+
+   lm1 <- lm(y ~ x + I(log(x)), data = obs.reduced)
+   lm2 <- lm(y ~ x + I(x^2), data = obs.reduced)
+
+   lm1.pred <- predict(lm1, newdata = obs[i, ])
+   lm2.pred <- predict(lm2, newdata = obs[i, ])
+
+   error.lm1[i] <- (lm1.pred - obs[i, ]$y)^2
+   error.lm2[i] <- (lm2.pred - obs[i, ]$y)^2
+ }
>
> mean(error.lm1)
[1] 148.4239
> mean(error.lm2)
[1] 179.8139
```

(a) Identify the name of the above approach and describe the general method implemented. ***(2 marks)***

(b) What is the purpose of this method and the motivation for using it in this scenario? ***(2 marks)***

(c) Given only the information in this analysis, write the mathematical model that should be preferred for this dataset. ***(1 mark)***

End of Question Paper