The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2018–2019**

MAS472 Computational Inference **2 hours**

Candidates may bring to the examination a calculator that conforms to University regulations.
Answer all questions. 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 Consider the probability density function of an exponential distributed random variable with mean $\frac{1}{\lambda}$

$$f(x) = \lambda e^{-\lambda x}.$$

(i) The inversion method can be used to convert a sequence $U_1, U_2, \ldots$ of $U[0,1]$ random variables to a sequence of exponential random variables by setting $X_i = g(U_i)$ for some choice of g.

Derive the function g.

Hint: The CDF of an $\text{Exp}(\lambda)$ random variable is $F(x) = 1 - e^{-\lambda x}$.

(3 marks)

(ii) Give an unbiased Monte Carlo estimator for the expected value of some function $h(X)$ when $X \sim Exp(\lambda)$ in terms of the uniform random variables $U_1, \ldots, U_n$, i.e., an estimator for $\mathbb{E}h(X)$.

(2 marks)

(iii) If $\mathbb{V}\text{ar}[h(X)] = 1$, how many samples (i.e., what value of n) would you need in order to compute a 95% confidence interval for $\mathbb{E}h(X)$ that has width less than 10^{-2}?

(4 marks)

(iv) Briefly discuss the advantages and disadvantages of using Monte Carlo integration compared with using a deterministic numerical integration scheme such as the mid-ordinate rule for estimating $\mathbb{E}h(X)$. How does your answer depend upon the dimension of X?

(3 marks)

(v) Suppose you are now told

$$h(x) = e^{-x^2}.$$

Describe how you could use importance sampling to estimate

$$\mathbb{E}h(X) = \int h(x) f(x) \mathrm{d}x,$$

using a gamma distribution as the importance/proposal distribution. Be sure to give the expression for the importance weights and the importance sampling estimator of $\mathbb{E}h(X)$.

Hint: The probability density function of a $\Gamma(\alpha, \beta)$ random variable is

$$g(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}.$$

(4 marks)

(vi) What would the optimal proposal distribution be for computing $\mathbb{E}h(X)$ when $h(x) = e^{-x^2}$?

(4 marks)

2 (i) Survival times for 5 rats who took an experimental drug are recorded as $\{4, 7, 12, 20, 38\}$ 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 $\hat{\alpha} = 1.4$ and $\hat{\beta} = 0.05$.

(a) Show that the profile log-likelihood function for α is

$$l_p(\alpha) = 5\log\alpha + 5\log\left(\frac{5}{\sum t_i^\alpha}\right) + (\alpha - 1)\sum\log t_i - 5.$$

(5 marks)

(b) By considering the profile deviance function, test the null hypothesis that $\alpha = 1$. You may assume that $l_p(\hat{\alpha}) = -18.5$, 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)

 Question 2 continued on next page

2 (continued)

(ii) We are given a data set of n observation pairs $(x_1, y_1), \ldots, (x_n, y_n)$. Our aim is to build a model to predict new values of y from values of x. We have two candidate models, $f_\psi(x)$ and $g_\theta(x)$, both of which are parametric models depending upon a parameter ψ and θ respectively. The sum of squared errors for each model is defined to be

$$S_f(\psi) = \sum_{i=1}^{n}(f_\psi(x_i) - y_i)^2, \qquad S_g(\theta) = \sum_{i=1}^{n}(g_\theta(x_i) - y_i)^2,$$

and the models are trained by choosing the parameter value which minimises the sum of squared errors, i.e.,

$$\hat{\psi} = \arg\min_{\psi} S_f(\psi), \qquad \hat{\theta} = \arg\min_{\psi} S_g(\theta).$$

(a) Explain why choosing between the models solely on the basis of the residual sum of squares, $S_f(\hat{\psi})$ and $S_g(\hat{\theta})$, may be a bad idea if interest lies in predicting y for new values of x.

(2 marks)

(b) Describe algorithmically, i.e., by writing out an algorithm, how you would use K-fold cross-validation to choose between the two models.

(4 marks)

(c) Discuss how the choice of K may affect your algorithm, and suggest what value you would use and why.

(2 marks)

(d) If we assume

$$y_i = f_\psi(x_i) + \epsilon_i$$

where $\epsilon_i \sim t_2$, i.e., has a t-distribution with two degrees of freedom, explain how you could use a Monte Carlo test to test the null hypothesis $H_0 : \psi = 0$ vs the alternative $H_1 : \psi \neq 0$.

Hint: You can use the sum of square errors as a test statistic:

$$T_{obs} = \sum_{i=1}^{n}(y_i - f_0(x_i))^2.$$

(4 marks)

3 (i) Suppose you are given as data a set of independent identically distributed samples from $F(\cdot)$, i.e., you are given data $\{X_1, \ldots, X_n\}$ where each X_i has distribution F.

The empirical cumulative distribution function (ECDF), based on the sample data $\{X_1, \ldots, X_n\}$, is

$$\hat{F}_X(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{X_i \leq x}.$$

(a) Describe a property of $\hat{F}_n(x)$ that makes it a good estimator of $F(x)$.

(2 marks)

(b) Using the plug-in principle, find an estimator, $\hat{\theta}$, of

$$\theta = \mathbb{E}_F(X^2).$$

(3 marks)

(c) Describe a bootstrap procedure to estimate the standard error of your estimator $\hat{\theta}$.

(4 marks)

(ii) Suppose $f(x)$ and $g(x)$ are probability density functions defined on $\mathbb{R}$. Let

$$M = \sup_{x \in \mathbb{R}} \left(\frac{f(x)}{g(x)} \right).$$

(a) Describe how rejection sampling can be used to generate observation from f using a set of random draws from g.

(2 marks)

(b) Suppose f is the half-Normal density given by

$$f(x) = \sqrt{\frac{2}{\pi}} \mathrm{e}^{-\frac{1}{2}x^2}, \quad x \geq 0.$$

If g is the exponential density

$$g(x) = \lambda \mathrm{e}^{-\lambda x}, \lambda > 0, x \geq 0,$$

show that

$$M = \sqrt{\frac{2\mathrm{e}^{\lambda^2}}{\pi \lambda^2}}.$$

Hint: $\frac{1}{2}x^2 - \lambda x = \frac{1}{2}(x-\lambda)^2 - \frac{1}{2}\lambda^2.$

(5 marks)

(c) What value of λ would optimize the acceptance rate?

(4 marks)

End of Question Paper