Imperial College
London

M4/5S9

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)

May-June 2017

This paper is also taken for the relevant examination for the Associateship of the Royal College of Science

Date: Monday 08 May 2017

Time: 10:00 - 12:00

Time Allowed: 2 Hours

This paper has 4 Questions.

Candidates should use ONE main answer book.

Supplementary books may only be used after the relevant main book(s) are full.

All required additional material will be provided.

Stochastic Simulation

- DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO.

- Affix one of the labels provided to each answer book that you use, but DO NOT USE THE LABEL WITH YOUR NAME ON IT.

- Credit will be given for all questions attempted, but extra credit will be given for complete or nearly complete answers to each question as per the table below.

Raw Mark	Up to 12	13	14	15	16	17	18	19	20
Extra Credit	0	½	1	1 ½	2	2 ½	3	3 ½	4

- Each question carries equal weight.

- Calculators may not be used.

1. Consider a region in $\mathbb{R}^2$ defined by:

$$C_h = \left\{ (u,v) \in \mathbb{R}^2 \ : \ 0 \leq u \leq \sqrt{h\left(\frac{v}{u}\right)} \right\}$$

where $h(x) \geq 0 \ \forall x \in \mathbb{R}$ and $\displaystyle\int_{\mathbb{R}} h(x)dx < \infty$.

(a) (i) Show that C_h has finite area.

(ii) Show that, if the bivariate vector (U,V) is distributed uniformly on C_h, then the random variable $X = V/U$ has probability density function

$$f_X(x) = \frac{h(x)}{\displaystyle\int_{\mathbb{R}} h(y)dy}.$$

Now define

$$h(x) = \begin{cases} \exp\{-\lambda x^2\} & x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

(b) Calculate values for a, b and c, such that

$$R_h = \left\{ (u,v) \in \mathbb{R}^2 \ : \ 0 \leq u \leq a, \ b \leq v \leq c \right\}$$

is the smallest rectangle such that $C_h \subseteq R_h$.

The following pseudo-code outlines a Ratio-of-Uniforms procedure for sampling from $f_X(x)$, which uses squeezing in Steps 3 and 4:

1. Choose values for the tuning parameters α and β.
2. Generate U_1, $U_2 \sim \mathcal{U}(0,1)$.
3. Set $U = aU_1$, $V = b + (c-b)U_2$, where a, b and c are given in part (b).
4. If $V \leq T_1(U;\alpha)$, set $X = V/U$.
5. If $V \geq T_2(U;\beta)$, return to Step 2.
6. If $V \leq T_3(U)$, set $X = V/U$, otherwise return to Step 2.

(c) (i) Derive the expression $T_3(U)$, which appears in Step 6.

(ii) By using a Taylor series expansion for $\exp(y)$, derive expressions for the squeezing functions $T_1(U;\alpha)$ and $T_2(U;\beta)$.

(d) Explain how the above Ratio-of-Uniforms procedure could be used to generate a homogeneous Poisson process of rate λ.

2. Consider the problem of estimating the quantity $\eta = \mathbb{E}_{f_X}[X]$, where X is a random variable with some known probability density function $f_X(\cdot)$.

Suppose that $\hat{\eta}_1$ and $\hat{\eta}_2$ are two distinct, unbiased and negatively correlated estimators for η, both of which have variance σ^2.

(a) Show that the arithmetic mean of $\hat{\eta}_1$ and $\hat{\eta}_2$ is an unbiased estimator with variance less than $\sigma^2/2$.

(b) Prove that if $U \sim \mathcal{U}(0,1)$, and $\phi : [0,1] \to \mathbb{R}$ is a monotonic increasing function, then $\phi(U)$ and $\phi(1-U)$ will be negatively correlated.

Hint: you may wish to consider defining a point $t \in [0,1]$ such that

$$1 - t = \inf\Big\{u : \phi(u) > \mathbb{E}[\phi(U)]\Big\}.$$

Suppose now that X has a truncated Cauchy distribution, i.e.

$$f_X(x) = \begin{cases} \dfrac{k}{\pi(1+x^2)}, & 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

(c) Calculate the constant of proportionality, k.

(d) (i) State an expression for an unbiased estimator for η, in terms of an i.i.d. sample $X_1, \ldots, X_n \sim f_X$; denote this estimator $\hat{\eta}_3$.

(ii) Now express your estimator $\hat{\eta}_3$ in terms of an i.i.d. sample $U_1, \ldots, U_n \sim \mathcal{U}(0,1)$.

(iii) Derive expressions for two further, different unbiased estimators for η. These estimators should be written as a function of an i.i.d. sample $U_1, \ldots, U_n \sim \mathcal{U}(0,1)$.

(iv) State how the variance of each of your estimators in (d)(iii) compares with the variance of $\hat{\eta}_3$.

Suppose we use Importance Sampling to estimate η, choosing an auxiliary density $g(x)$.

(e) Derive the Importance Sampling estimator $\hat{\eta}_{IS}$ when the auxiliary density corresponds to a $\mathcal{U}(0,1)$ distribution.

3. The central Laplace distribution has probability density function

$$f_X(x) = \frac{1}{2\sigma} \exp\left\{-\frac{|x|}{\sigma}\right\}, \qquad x \in \mathbb{R},$$

and it is known that this is also the distribution of the random variable

$$X = \begin{cases} -\sigma \log(1 - 2|W|) & W \geq 0 \\ \sigma \log(1 - 2|W|) & W < 0 \end{cases}$$

where $W \sim \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right)$.

(a) Suppose we have access to a stream of $n+1$ i.i.d. Exponentially-distributed random variables

$$Y_1, \ldots, Y_{n+1} \sim \mathsf{Exp}(1).$$

(i) By defining

$$S_k = \sum_{i=1}^{k} Y_k, \qquad k = 1, \ldots, n+1,$$

show that the joint density of $S = (S_1, \ldots, S_{n+1})$ can be written

$$f_S(s_1, \ldots, s_{n+1}) = \exp\left\{-s_{n+1}\right\}.$$

(ii) Using part (i), and defining

$$V_k = S_k / S_{n+1}, \qquad k = 1, \ldots, n, \qquad\qquad \text{and} \qquad V_{n+1} = S_{n+1},$$

show that the joint density of $V = (V_1, \ldots, V_{n+1})$ can be written

$$f_V(v_1, \ldots, v_{n+1}) = \exp\left\{-v_{n+1}\right\} v_{n+1}^n.$$

(iii) Using part (ii), show that $V_1, \ldots, V_n$ can be treated as a set of n ordered $\mathcal{U}(0,1)$ random variates.

(iv) Hence, briefly explain how we can use $Y_1, \ldots, Y_{n+1}$ to obtain an ordered sample

$$X_{(1)}, \ldots, X_{(n)} \sim f_X.$$

[Continued on next page.]

The President of the United States of America tweets the following:

> *"I've just observed an ordered data set $x_{(1)}, \ldots, x_{(n)}$ from the central Laplace distribution. The ratio $x_{(n)}/x_{(1)}$ is -0.5. So central! Amazing!"*

In order to fact-check this tweet, you decide to carry out a two-tailed Monte Carlo test with Type I error α, using the observed test statistic $t = x_{(n)}/x_{(1)}$. Your null hypothesis, H_0, is that the observed data are distributed according to $f_X(x)$. The distribution of $T = X_{(n)}/X_{(1)}$ under H_0 is unknown but continuous.

You generate m distinct ordered datasets from f_X, and use these to obtain simulated values $t_1^*, \ldots, t_m^*$ of the test statistic. These simulated values are then sorted, to provide an ordered set of simulated test statistics, which we denote $t_{(1)}^*, \ldots, t_{(m)}^*$.

(b) Give expressions for k_1 and k_2 such that we would reject H_0 if either

$$t < t_{(k_1)}^* \qquad \text{or} \qquad t > t_{(k_2)}^*.$$

(c) Show that the Monte Carlo test rejects H_0 with probability

$$p = \sum_{r=0}^{k_1-1} \binom{m}{r} \left\{ u^r (1-u)^{m-r} + u^{m-r} (1-u)^r \right\},$$

where $u = \mathbb{P}(T \leq t)$.

Having implemented the MC test, suppose we do <u>not</u> reject H_0. Define

$$T_i = t_i^*, \qquad i = 1, \ldots, m, \qquad \text{and} \qquad T_{m+1} = t,$$

such that $\underline{T} = (T_1, \ldots, T_{m+1})$ is a vector of $m+1$ i.i.d., unordered replicates of the test statistic T and denote the corresponding sample mean by $\bar{T}$.

(d) Explain how Jackknife resampling can be used to estimate the variance of $\bar{T}$. Provide an expression for the Jackknife estimator for the variance of $\bar{T}$.

4. Suppose we wish to sample from a distribution with pdf $\pi(x)$, defined on a continuous univariate state space $\mathcal{X} \subseteq \mathbb{R}$.

(a) We can sample from $\pi(x)$ using the following static (i.e. non-adaptive) Metropolis-Hastings procedure to construct a Markov chain $\{X_n\}_{n\geq 0}$ with $\pi(x)$ as its stationary distribution:

Step 1: Initialise the chain by sampling from some initial distribution: $X_0 \sim \pi^{(0)}$. Set $n = 1$.

Step 2: Given $X_{n-1} = x$, generate a candidate value $Y = y$ from the proposal density

$$q_h(y|x) = \frac{1}{h} k\left(\frac{y-x}{h}\right), \qquad x, y \in \mathcal{X}, h > 0,$$

Step 3: Set $X_n = y$ with probability $\alpha(x, y)$, otherwise set $X_n = x$.

Step 4: Replace n by $n + 1$ and return to Step 2.

(i) Give an expression for $\alpha(x, y)$ and state the property of the resulting Markov chain that guarantees the existence of a stationary distribution.

(ii) What condition must $k(\cdot)$ satisfy in order for the above Metropolis-Hastings procedure to reduce to a Metropolis procedure.

(iii) The given proposal density, q_h, allows the user to control the performance of the Metropolis procedure through the fixed parameter h. Assuming the condition in part (ii) holds, describe and explain the effect on the acceptance probability $\alpha(x, y)$ that comes from decreasing h.

Suppose we are now interested in using adaptive MCMC methods to sample from the multivariate distribution with pdf $\pi_d(x)$, which is defined on the continuous state space $\mathcal{X}^d \subseteq \mathbb{R}^d$, $d > 1$.

(b) The Adaptive Metropolis (AM) algorithm is a version of the Metropolis procedure that uses the following proposal density at iteration n:

$$q_n(y|x) = \begin{cases} f_{\mathcal{N}}(x, (0.1)^2 I_d/d) & n \leq 2d \\ (1-\beta) f_{\mathcal{N}}(y|x, 2.38^2 S_n/d) + \beta f_{\mathcal{N}}(x, 0.1^2 I_d/d) & n > 2d, \end{cases}$$

where $f_{\mathcal{N}}(\cdot|\mu, \Sigma)$ is a multivariate Gaussian density with mean μ and covariance matrix Σ, and where I_d is the d-dimensional identity matrix and $\beta \in (0, 1)$ is a user-specified constant.

(i) Describe the quantity denoted by S_n, and briefly explain why the value of $2.38^2/d$ is chosen to scale S_n in the above density.

(ii) Briefly explain why the non-adaptive density $f_{\mathcal{N}}(x, 0.1^2 I_d/d)$ is included in the proposal mechanism.

[Continued on next page.]

(c) As an alternative to the AM procedure in part (b), one can employ component-wise adaptive scaling methods, such as the Single Component Adaptive Metropolis (SCAM) algorithm and the Adaptive Metropolis-within-Gibbs (AMWG) algorithm.

Briefly explain the primary difference between the SCAM and AMWG procedures.

(d) Suppose that, at the start of iteration n, the existing output of an adaptive Metropolis procedure is denoted $\underline{x}_n = (x'_0, \ldots, x'_{n-1}) \in \mathcal{X}^{d \times n}$, where $x_j = (x_j^1, \ldots, x_j^d) \in \mathcal{X}^d$ for $j = 0, \ldots, n-1$, and where x'_j denotes the transpose of x_j. Denote the row vector containing only component i of this output by $\underline{x}_n^i = (x_0^i, \ldots, x_{n-1}^i) \in \mathcal{X}^n$, and define the sample mean $\bar{x}_{n-1}^i$ and sample variance g_n^i for this univariate component of the output in the usual way:

$$\bar{x}_{n-1}^i = \frac{1}{n} \sum_{j=0}^{n-1} x_j^i, \qquad g_n^i = \frac{1}{n-1} \sum_{j=0}^{n-1} \left(x_j^i - \bar{x}_n^i \right)^2.$$

Show that g_n^i satisfies the following relationship:

$$g_n^i = \frac{n-2}{n-1} g_{n-1}^i + \left(\bar{x}_{n-1}^i\right)^2 + \frac{1}{n-1}\left(x_{n-1}^i\right)^2 - \frac{n-1}{n}\left(\bar{x}_{n-1}^i + \frac{x_{n-1}^i}{n-1}\right)^2.$$

(e) Roberts & Rosenthal ("Examples of Adaptive MCMC", 2009) use integrated autocorrelation time and average square jump distance as indicators of performance for adaptive MCMC procedures.

(i) Which aspect of performance is being measured by these quantities?

(ii) For each of these indicators, state whether good performance of the procedure is indicated by high or low values?

	EXAMINATION SOLUTIONS 2016-17	Course M3/4S9
Question 1		Marks & seen/unseen

Part (a) (i) — seen

$$\text{Area}(C_h) = \int\int_{C_h} du dv$$

Use a change of variables: $(U, V) \to \left(U, X = \frac{V}{U}\right)$ — 1

$$V = XU \quad \Longrightarrow \quad \frac{dv}{dx} = u \ , \ \ dv = u\,dx,$$

So, $\text{Area}(C_h) = \int_{\mathbb{R}} \left[\int_0^{\sqrt{h(x)}} u du \right] dx = \frac{1}{2} \int_{\mathbb{R}} h(x) dx < \infty$ — 1

Part (a) (ii) — seen

The Jacobian for the above change of variables is

$$\left| \frac{\partial(u, x)}{\partial(u, v)} \right| = \frac{1}{u}$$

$$\Longrightarrow \ f_{U,X}(u, x) = u f_{U,V}(u, v)$$

$$= \frac{u}{\text{Area}(C_h)} \text{ if } (U, V) \text{ is uniform on } C_h$$ — 1

Now, we marginalise to obtain the density for X: — 1

$$f_X(x) = \int_0^{\sqrt{h(x)}} f_{U,X}(u, x)\, du$$

$$= \frac{1}{\text{Area}(C_h)} \int_0^{\sqrt{h(x)}} u\, du = \frac{\frac{1}{2} h(x)}{\frac{1}{2} \int_{\mathbb{R}} h(x)}$$

Part (b) — seen

Find the bounding rectangle $R_h = \{(u, v) \in \mathbb{R}^2 : 0 \leq u \leq a, b \leq v \leq c\}$:

$$a = \sup_{x \geq 0} \sqrt{h(x)} = 1$$ — 2

$$b = \inf_{x \leq 0} x \sqrt{h(x)} = 0$$ — 2

$$c = \sup_{x \geq 0} x \sqrt{h(x)} = \sup_{x \geq 0} \left\{ x e^{-\lambda x^2/2} \right\}$$

To find c, we must find max of $g(x) = x e^{-\lambda x^2/2}$:

$$g'(x) = e^{-\lambda x^2/2} \left(1 - \lambda x^2\right)$$
$$= 0 \text{ when } x = \pm \lambda^{-1/2}$$

We can only take $x = \lambda^{-1/2}$, as we are looking for $x \geq 0$

We verify that $x = \lambda^{-1/2}$ is a maximum through checking that $g''(x) < 0$ when $x = \lambda^{-1/2}$ — 1

So $\quad c = \sup_{x \geq 0} g(x) = g\left(\lambda^{-1/2}\right) = (\lambda e)^{-1/2}$ — 1

	Setter's initials	Checker's initials	Page number

	EXAMINATION SOLUTIONS 2016-17	Course M3/4S9
Question 1		Marks & seen/unseen

Part (c) (i) — seen sim

Step 5 of the given procedure is the post-squeezing membership test for the Ratio of Uniforms procedure.
In the RoU procedure, we set $X = V/U$ if $(U,V) \in C_h$, i.e. if

$$U \le \sqrt{h(V/U)} = \exp\left\{-\frac{\lambda}{2}\frac{V^2}{U^2}\right\}$$ 1

$$\implies \log U \le -\frac{\lambda}{2}\frac{V^2}{U^2}$$

$$\implies -\frac{2U^2}{\lambda}\log U \ge V^2$$

$$\implies V \le \left[-\frac{2U^2}{\lambda}\log U\right]^{1/2} \implies T_3(U) = \left[-\frac{2U^2}{\lambda}\log U\right]^{1/2}$$ 2

Part (c) (ii) — seen sim

From the Taylor series expansion for $\exp(y)$, we know that

$$e^y \ge 1 + y \qquad \forall y \in \mathbb{R}$$ 1

$$\implies y \ge \log\{1+y\} \qquad \forall y > -1 \qquad (\star)$$

Writing $y = \alpha U - 1$ in $(\star)$, we have

$$\alpha U - 1 \ge \log\{\alpha U\} \qquad \forall U > 0$$

$$\implies \alpha U - 1 - \log\alpha \ge \log U \qquad \forall U > 0$$

$$\implies \log\alpha - \alpha U + 1 \le -\log U \qquad \forall U > 0$$

$$\implies \left[\frac{2U^2}{\lambda}(1+\log\alpha) - \frac{2\alpha U^3}{\lambda}\right]^{1/2} \le T_3(U) \qquad \forall U > 0$$

So we can take $\quad T_1(U;\alpha) = \left[\frac{2U^2}{\lambda}(1+\log\alpha) - \frac{2\alpha U^3}{\lambda}\right]^{1/2}$

[1 mark for evidence of understanding that we need

$$T_1(U;\alpha) \le T_3(U) \qquad \forall U > 0$$

1 mark for a correct derivation of a sensible $T_1(U;\alpha)$.] 2

If we instead write $y = \beta/U - 1$ in $(\star)$, we have

$$\beta/U - 1 \ge \log\{\beta/U\} \qquad \forall U > 0$$

$$\implies \beta/U - (1+\log\beta) \ge -\log U \qquad \forall U > 0$$

$$\implies \left[\frac{2\beta U}{\lambda} - \frac{2U^2}{\lambda}(1+\log\beta)\right]^{1/2} \ge T_3(U) \quad \forall U > 0$$

So we can take $\quad T_2(U;\beta) = \left[\frac{2\beta U}{\lambda} - \frac{2U^2}{\lambda}(1+\log\beta)\right]^{1/2}$

[1 mark for evidence of understanding that we need

$$T_2(U;\beta) \ge T_3(U) \qquad \forall U > 0$$

1 mark for a correct derivation of a sensible $T_2(U;\beta)$.] 2

	Setter's initials	Checker's initials	Page number

	EXAMINATION SOLUTIONS 2016-17	Course M3/4S9
Question 1		Marks & seen/unseen
Part (d)	To generate a Poisson process, we could simulate exponential waiting times $$T_i \sim f_T \propto \exp\{-\lambda t\},$$ which could be achieved using RoU as above, with $h(x) = \exp\{-\lambda x\}$ instead of $h(x) = \exp\{-\lambda x^2\}$	unseen 2
	Setter's initials Checker's initials	Page number

	EXAMINATION SOLUTIONS 2016-17	Course M3/4S9
Question 2		Marks & seen/unseen

Part (a) — seen

Unbiasedness:

$$\mathbb{E}\Big[\frac{1}{2}(\hat\eta_1+\hat\eta_2)\Big] = \frac{1}{2}\Big(\mathbb{E}[\hat\eta_1]+\mathbb{E}[\hat\eta_2]\Big) = \frac{1}{2}(\eta+\eta) = \eta \qquad 1$$

Lower variance:

$$\mathsf{Var}\Big[\frac{1}{2}(\hat\eta_1+\hat\eta_2)\Big] = \frac{1}{4}\mathsf{Var}[\hat\eta_1] + \frac{1}{4}\mathsf{Var}[\hat\eta_2] + \frac{1}{2}\mathsf{Cov}[\hat\eta_1,\hat\eta_2]$$

$$= \frac{1}{2}\Big(\sigma^2 + \mathsf{Cov}[\hat\eta_1,\hat\eta_2]\Big) = \frac{\sigma^2}{2}\Big(1+\mathsf{Corr}[\hat\eta_1,\hat\eta_2]\Big)$$

So, $\quad \mathsf{Corr}[\hat\eta_1,\hat\eta_2] < 0 \implies \mathsf{Var}\Big[\frac{1}{2}(\hat\eta_1+\hat\eta_2)\Big] < \sigma^2/2$ 1

Part (b) — seen

Let $\quad \theta = \mathbb{E}[\phi(U)] = \mathbb{E}[\phi(1-U)],$ as $U \sim \mathcal{U}(0,1)$

Then

$$\begin{aligned}\mathsf{Cov}\Big[\phi(U),\phi(1-U)\Big] &= \mathbb{E}\Big[\Big(\phi(U)-\theta\Big)\Big(\phi(1-U)-\theta\Big)\Big]\\ &= \mathbb{E}\Big[\phi(U)\Big(\phi(1-U)-\theta\Big)\Big] - \theta\mathbb{E}\Big[\phi(1-U)-\theta\Big]\\ &= \mathbb{E}\Big[\phi(U)\Big(\phi(1-U)-\theta\Big)\Big]\\ &= \int_0^1 \phi(u)\Big(\phi(1-u)-\theta\Big)\,du\end{aligned} \qquad 1$$

Now, $\quad 1-t = \inf\Big\{u : \phi(u) > \theta\Big\}$, so

$$u\in[0,t] \implies 1-u\in[1-t,1] \implies \phi(1-u) > \theta, \qquad (\star)$$

and

$$u\in(t,1] \implies 1-u\in[0,1-t) \implies \phi(1-u) \le \theta, \qquad (\star\star)$$

so we can split our integral at t: 1

$$\mathsf{Cov}\Big[\phi(U),\phi(1-U)\Big] = \int_0^t \phi(u)\Big(\phi(1-u)-\theta\Big)\,du + \int_t^1 \phi(u)\Big(\phi(1-u)-\theta\Big)\,du$$

and use $(\star)$ and $(\star\star)$ to provide upper bounds for each component.

$$\begin{aligned}(\star) &\implies \phi(1-u)-\theta > 0\\ &\implies \int_0^t \phi(u)\Big(\phi(1-u)-\theta\Big)\,du < \phi(t)\int_0^t \Big(\phi(1-u)-\theta\Big)\,du\end{aligned} \qquad 1$$

$$\begin{aligned}(\star\star) &\implies \theta-\phi(1-u) \ge 0\\ &\implies \int_t^1 \phi(u)\Big(\theta-\phi(1-u)\Big)\,du \ge \phi(t)\int_t^1 \Big(\theta-\phi(1-u)\Big)\,du\\ &\implies \int_t^1 \phi(u)\Big(\phi(1-u)-\theta\Big)\,du \le \phi(t)\int_t^1 \Big(\phi(1-u)-\theta\Big)\,du\end{aligned} \qquad 1$$

(cont. on next page)

Combining these results, we get

$$\text{Cov}\Big[\phi(U), \phi(1-U)\Big] < \phi(t) \underbrace{\int_0^1 \Big(\phi(1-u) - \theta\Big)\, du}_{=0}$$

So $\quad \text{Cov}\Big[\phi(U), \phi(1-U)\Big] < 0$

$$\implies \text{Corr}\Big[\phi(U), \phi(1-U)\Big] < 0$$ 1

Part (c) — unseen

$$f_X(x) = \frac{k}{\pi}\left(1+x^2\right)^{-1}, \qquad 0 \le x \le 1$$

$$\int_0^1 f_X(x)\, dx = 1 \implies \frac{1}{k} = \frac{1}{\pi}\int_0^1 \left(1+x^2\right)^{-1}\, dx$$ 1

$$= \frac{1}{\pi}\left[\tan^{-1} x\right]_0^1 = \frac{1}{\pi}\frac{\pi}{4} \implies k = 4$$ 1

Part (d) (i) — seen

A simple unbiased estimator for η is the sample mean:

$$\hat{\eta}_3 = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad X_i \sim f_X, i = 1, \dots, n$$

[A mark should be given for any estimator in terms of $X_1, \dots, X_n$ that is unbiased for η] 1

Part (d) (ii) — unseen

From part (c), we know the pdf

$$f_X(x) = \frac{4}{\pi}\frac{1}{1+x^2} \qquad 0 \le x \le 1$$

and straightforward integration yields the corresponding cdf:

$$F_X(x) = \frac{4}{\pi}\tan^{-1} x \qquad 0 \le x \le 1.$$

From the probability integral transform, we know that, for continuous X, $F_X^{-1}(X) \sim \mathcal{U}(0,1)$, so if $U \sim \mathcal{U}(0,1)$, then $X = \tan\left(\frac{\pi U}{4}\right)$ will have the required distribution, and we can write

$$\hat{\eta}_3 = \frac{1}{n}\sum_{i=1}^{n} \tan\left(\frac{\pi U_i}{4}\right)$$

[2 marks should be given for correct calculation of the inverse CDF, and the correct subsequent application to the estimator given in (d)(i)] 2

[Inversion sampling has been seen; its use in developing a second unbiased estimator is unseen.]

(cont. on next page)

	EXAMINATION SOLUTIONS 2016-17	Course M3/4S9
Question 2		Marks & seen/unseen

Part (d) (iii) — seen

In (d)(ii), we use the inverse cdf to write $X = F^{-1}(U)$.
We know that, for $U \sim \mathcal{U}(0,1)$, and for any (integrable) function $\phi : [0,1] \to \mathbb{R}$,

$$\mathbb{E}\big[\phi(U)\big] = \mathbb{E}\big[\phi(1-U)\big].$$

$$\implies \mathbb{E}\big[F^{-1}(U)\big] = \mathbb{E}\big[F^{-1}(1-U)\big] = \eta.$$

So,

$$\hat{\eta}_4 = \frac{1}{n}\sum_{i=1}^{n} F^{-1}(1-U_i) = \frac{1}{n}\sum_{i=1}^{n} \tan\left(\frac{\pi}{4} - \frac{\pi U_i}{4}\right)$$

is also an unbiased estimator for η.
From part (a), we know that, since $\hat{\eta}_3$ and $\hat{\eta}_4$ are both unbiased for η, then

$$\hat{\eta}_5 = \frac{1}{2}\Big(\hat{\eta}_3 + \hat{\eta}_4\Big)$$ is also unbiased for η.

[For each estimator:
1 mark for stating an unbiased estimator distinct from $\hat{\eta}_3$ — 2
1 mark for a correct justification or derivation.] — 2

Part (d) (iv)

For the estimators given above,

- $\text{Var}\Big[\hat{\eta}_4\Big] = \text{Var}\Big[\hat{\eta}_3\Big]$
- $\text{Var}\Big[\hat{\eta}_5\Big] < \text{Var}\Big[\hat{\eta}_3\Big]\Big/2.$

Note that this follows from parts (a) and (b) as $F^{-1}(u)$ is a monotonic increasing function of u.

[For each estimator:
1 mark for stating correctly how the variance of the estimators they give in part (d)(iii) compare with that of $\hat{\eta}_3$.] — 2

NB: Whilst the tools for answering (d) are 'seen', a complete answer requires a deeper understanding of the material than would be tested by simple regurgitation of 'seen' content.

Part (e) — seen

The Importance Sampling estimator $\hat{\eta}_{IS}$ is given by

$$\hat{\eta}_{IS} = \frac{1}{n}\sum_{i=1}^{n} \frac{f_X(U_i)}{g(U_i)} U_i \qquad U_i \sim \mathcal{U}(0,1)$$

with $g(x) = 1 \; \forall x \in [0,1]$ and $f_X(x)$ as calculated in part (c).
Thus,

$$\hat{\eta}_{IS} = \frac{1}{n}\sum_{i=1}^{n} \frac{4U_i}{\pi(1+U_i^2)} \qquad U_i \sim \mathcal{U}(0,1)$$ — 2

	Setter's initials	Checker's initials	Page number

	EXAMINATION SOLUTIONS 2016-17	Course M3/4S9
Question 3		Marks & seen/unseen

Part (a)(i) — seen

First, note that $Y_1, \ldots, Y_{n+1} \overset{iid}{\sim} \text{Exp}(1)$

$$\implies f_Y(y_1, \ldots, y_{n+1}) = \exp\left\{-\sum_{i=1}^{n+1} y_i\right\}$$ 1

Now, using

$$S_k = \sum_{i=1}^{k} Y_k, \quad k = 1, \ldots, n+1$$

$$\implies \begin{cases} Y_1 = S_1 \\ Y_k = S_k - S_{k-1}, \quad k = 2, \ldots, n+1, \end{cases}$$

we can find the joint density:

$$f_S(s_1, s_2, \ldots, s_{n+1}) = f_Y(s_1, s_2 - s_1, \ldots, s_{n+1} - s_n)|J|$$ 1

where

$$|J| = \begin{vmatrix} \frac{\partial y_1}{\partial s_1} & \frac{\partial y_1}{\partial s_2} & \cdots & \frac{\partial y_1}{\partial s_{n+1}} \\ \frac{\partial y_2}{\partial s_1} & \frac{\partial y_2}{\partial s_2} & \cdots & \frac{\partial y_2}{\partial s_{n+1}} \\ \vdots & & & \vdots \\ \frac{\partial y_{n+1}}{\partial s_1} & \frac{\partial y_{n+1}}{\partial s_2} & \cdots & \frac{\partial y_{n+1}}{\partial s_{n+1}} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{vmatrix} = 1$$ 1

So,

$$f_S(s_1, s_2, \ldots, s_{n+1}) = \exp\left\{-s_1 - \sum_{k=2}^{n+1}(s_k - s_{k-1})\right\}$$
$$= \exp\{-s_{n+1}\}$$

Part (a)(ii) — seen

Now for a second transformation of variables: using

$$\left.\begin{array}{l} V_k = S_k/S_{n+1} \quad k = 1, \ldots, n, \\ V_{n+1} = S_{n+1}, \end{array}\right\} \implies \begin{cases} S_k = V_{n+1}V_k \quad k = 1, \ldots, n, \\ S_{n+1} = V_{n+1}, \end{cases}$$

we obtain the joint density

$$f_V(v_1, v_2, \ldots, v_{n+1}) = f_S(v_1 v_{n+1}, v_2 v_{n+1}, \ldots, v_{n+1})|J|$$ 1

where

$$|J| = \begin{vmatrix} \frac{\partial s_1}{\partial v_1} & \frac{\partial s_1}{\partial v_2} & \cdots & \frac{\partial s_1}{\partial v_{n+1}} \\ \frac{\partial s_2}{\partial v_1} & \frac{\partial s_2}{\partial v_2} & \cdots & \frac{\partial s_2}{\partial v_{n+1}} \\ \vdots & & & \vdots \\ \frac{\partial s_{n+1}}{\partial v_1} & \frac{\partial s_{n+1}}{\partial v_2} & \cdots & \frac{\partial s_{n+1}}{\partial v_{n+1}} \end{vmatrix} = \begin{vmatrix} v_{n+1} & 0 & \cdots & v_1 \\ 0 & v_{n+1} & \cdots & v_2 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{vmatrix} = v_{n+1}^n$$ 1

So, using part (i),

$$f_V(v_1, \ldots, v_n, v_{n+1}) = \exp\{-v_{n+1}\}v_{n+1}^n$$

<table>
<tr><td></td><td>EXAMINATION SOLUTIONS 2016-17</td><td>Course
M3/4S9</td></tr>
<tr><td>Question 3</td><td></td><td>Marks &
seen/unseen</td></tr>
</table>

Part (a)(iii) — seen

We find the joint density for $(V_1, \ldots, V_n)$ and show that it is equal to the density for n ordered standard uniforms.

The joint pdf of n ordered uniforms is $f_{U_{(1)},\ldots,U_{(n)}}(u_{(1)}, \ldots, u_{(n)}) = n!$ — 1

The joint density for $(V_1, \ldots, V_n)$ can be found through marginalising:

$$f_V(v_1, \ldots, v_n) = \int_0^\infty f_V(v_1, \ldots, v_n, v_{n+1})\, dv_{n+1}$$ — 1

$$= \int_0^\infty \exp\{-v_{n+1}\} v_{n+1}^n \, dv_{n+1}$$

Integration by parts gives — 1

$$= \underbrace{\left[-\exp\{-v_{n+1}\} v_{n+1}^n\right]_0^\infty}_{=0} + n \int_0^\infty \exp\{-v_{n+1}\} v_{n+1}^{n-1} \, dv_{n+1}$$

and repeating another $n-1$ times yields

$$f_V(v_1, \ldots, v_n) = n! \int_0^\infty \exp\{-v_{n+1}\} \, dv_{n+1} = n!$$ — 1

as required.

Part (a)(iv) — unseen

Given $Y_1, \ldots, Y_{n+1} \sim \text{Exp}(1)$, we can use the transformations in parts (i)-(iii) to generate n ordered random variables

$$U_{(1)}, \ldots, U_{(n)} \sim \mathcal{U}(0, 1)$$

By defining $W_{(i)} = U_{(i)} - 1/2$, $i = 1, \ldots, n$, we can then obtain $X_{(1)}, \ldots, X_{(n)} \sim f_X$ by using the transformation given in the question. — 1

[Alternatively, assuming the cdf for the given distribution to be available, we could use the ordered uniforms in an inversion sampling procedure]

Part (b) — seen sim

For a two-tailed Monte Carlo test with Type I error α, H_0 is rejected if either

$$t < t^*_{(k_1)} \qquad \text{or} \qquad t > t^*_{(k_2)},$$

where

$$\frac{k_1}{m+1} = 1 - \frac{k_2}{m+1} = \frac{\alpha}{2}$$

$$\implies k_1 = (m+1)\frac{\alpha}{2}, \qquad k_2 = (m+1)\left(1 - \frac{\alpha}{2}\right)$$ — 2

<table>
<tr><td></td><td>Setter's initials</td><td>Checker's initials</td><td>Page number</td></tr>
</table>

Part (c) — seen sim

For our two-tailed Monte Carlo test, the probability of rejecting H_0 is

$$p = \mathbb{P}\Big(r \text{ simulated values of } T \le t,\ 0 \le r \le k_1 - 1\Big) + \mathbb{P}\Big(r \text{ simulated values of } T \ge t,\quad 0 \le r \le m - k_2\Big)\ (\star)$$

or, equivalently,

$$p = \mathbb{P}\Big(r \text{ simulated values of } T \le t,\ 0 \le r \le k_1 - 1\Big) + \mathbb{P}\Big(r \text{ simulated values of } T \le t,\quad k_2 \le r \le m\Big) \qquad (\star\star)$$

2

We deal with $(\star)$ and $(\star\star)$ separately; both are valid approaches. — unseen

$(\star)$:

$$p = \sum_{r=0}^{k_1-1} \mathbb{P}\Big(r \text{ sim. values of } T \le t\Big) + \sum_{r=0}^{m-k_2} \mathbb{P}\Big(r \text{ sim. values of } T \ge t\Big)$$

$$= \sum_{r=0}^{k_1-1} \Big\{\mathbb{P}\Big(r \text{ sim. values of } T \le t\Big) + \mathbb{P}\Big(r \text{ sim. values of } T \ge t\Big)\Big\}$$

as $m - k_2 = k_1 - 1$.

The two probabilities inside the braces are binomial probabilities, which can be rewritten:

$$p = \sum_{r=0}^{k_1-1} \left\{ \binom{m}{r} u^r (1-u)^{m-r} + \binom{m}{r} (1-u)^r u^{m-r} \right\}$$

and this can be rearranged to give the required expression.

$(\star\star)$:

$$p = \sum_{r=0}^{k_1-1} \mathbb{P}\Big(r \text{ sim. values of } T \le t\Big) + \sum_{r=k_2}^{m} \mathbb{P}\Big(r \text{ sim. values of } T \le t\Big)$$

$$= \sum_{r=0}^{k_1-1} \binom{m}{r} u^r (1-u)^{m-r} + \sum_{r=k_2}^{m} \binom{m}{r} u^r (1-u)^{m-r}$$

$$= \sum_{r=0}^{k_1-1} \binom{m}{r} u^r (1-u)^{m-r} + \sum_{r=m+1-k_1}^{m} \binom{m}{r} u^r (1-u)^{m-r}$$

Now, substituting $s = m - r$ in the second summation (and switching the limits in the notation):

$$= \sum_{r=0}^{k_1-1} \binom{m}{r} u^r (1-u)^{m-r} + \sum_{s=0}^{k_1-1} \binom{m}{m-s} u^{m-s} (1-u)^s$$

and since $\binom{m}{r} = \binom{m}{m-r}$, this can be rewritten

$$p = \sum_{r=0}^{k_1-1} \left\{ \binom{m}{r} u^r (1-u)^{m-r} + \binom{m}{r} (1-u)^r u^{m-r} \right\}$$

which can be rearranged to give the required expression.

2

	EXAMINATION SOLUTIONS 2016-17	Course M3/4S9
Question 3		Marks & seen/unseen
Part (d)	To perform Jackknife resampling, repeat the following for $i = 1, \ldots, m+1$: • construct an m-length vector $\underline{T}_{-i}$ by dropping the i^{th} element of $\underline{T}$ • use $\underline{T}_{-i}$ to estimate the sample mean: $$\bar{T}_{-i} = \frac{1}{m} \sum_{\substack{j=1 \\ j \neq i}}^{m+1} T_j$$ This gives $m+1$ estimates of the sample mean, $\bar{T}_{-i}$, $i = 1, \ldots, m+1$. The variance of $\bar{T}$ can then be estimated using a corrected sample variance of the $m+1$ sample means: $$\widehat{\text{Var}}[\bar{T}] = \frac{m}{m+1} \sum_{i=1}^{m+1} \left(\bar{T}_{-i} - \bar{T}_{(\cdot)} \right)^2.$$ where $$\bar{T}_{(\cdot)} = \frac{1}{m+1} \sum_{i=1}^{m} \bar{T}_{-i},$$ *[Can alternatively just use $\bar{T}$ instead of $\bar{T}_{(\cdot)}$ in expression for $\widehat{Var}[\bar{T}]$].*	seen 1 1 2
	Setter's initials Checker's initials	Page number

	EXAMINATION SOLUTIONS 2016-17	Course M3/4S9
Question 4		Marks & seen/unseen
Part (a)(i)	$$\alpha(x,y) = \min\left\{\frac{\pi(y)\ q_h(x \mid y)}{\pi(x)\ q_h(y \mid x)}\ ,\ 1\right\}$$	seen 2
	The resulting Markov chain will be reversible, guaranteeing the existence of its stationary distribution.	1
	Reversibility in the MC is achieved by design, as the form of $\alpha(x,y)$ ensures that the transition densities satisfy the detailed balance equations. *[Must mention either reversibility or detailed balance for the full mark]*	
Part (a)(ii)		seen
	The function $k(\cdot)$ must be symmetric (in x and y) in order for the given MH procedure to reduce to a Metropolis procedure.	1
Part (a)(iii)		seen
	Since the proposal q_h is assumed to be symmetric, the acceptance probability is simply the ratio of target densities $\pi(y)/\pi(x)$.	
	If h is small, then q_h will often propose candidate values in regions where the ratio of target densities $\pi(y)/\pi(x)$ is high; this is because the proposal mechanism will be more likely to propose candidate values close to the current value.	1
	Thus, decreasing h will increase the acceptance probability.	1
Part (b)(i)		seen/unseen
	S_n is the empirical estimate, at iteration n, of the covariance matrix of the target distribution, based on the simulated chain so far.	1 (seen)
	The covariance matrix in the multivariate Gaussian density performs the same role as h in part (a) - it scales the proposal density.	
	The multivariate Gaussian proposal $$f_{\mathcal{N}}(y \mid x, h\Sigma_\pi)$$ specifies a $d-$dimensional random-walk Metropolis procedure.	*Max 2 from 3:*
	As the dimension $d \to \infty$, the Markov chain produced by this procedure converges to a diffusion process with step size proportional to $1/d$.	1 (unseen)
	The diffusion speed of this process is maximised by $$h = 2.38^2/d,$$ which corresponds to a Metropolis acceptance rate of ≈ 0.234.	1 (unseen)
	The use of $2.38/d^2$ to scale the proposal distribution is an effort to approximate this optimality for finite d.	1 (seen)
	Setter's initials Checker's initials	Page number

Part (b)(ii) — seen

The mixture with a non-adaptive proposal density is chosen to ensure that the procedure does not get stuck at problematic values of S_n. 1
An example of a problematic S_n is one that is singular. 1

Part (c) — seen

The SCAM procedure adaptively scales its univariate proposal for component $i \in \{1, \ldots, d\}$ based on the empirical variance of the corresponding component of the simulated chain so far. 1
In contrast, the AMWG procedure adaptively scales its univariate proposal for component i based on the empirical acceptance rate so far, for component i. 1

Part (d) — unseen

g_n^i is the sample variance for $(x_0^i, \ldots x_{n-1}^i)$:

$$\begin{aligned} g_n^i &= \frac{1}{n-1} \sum_{j=0}^{n-1} \left(x_j^i - \bar{x}_n^i\right)^2 \\ &= \frac{1}{n-1} \sum_{j=0}^{n-1} \left(x_j^i\right)^2 - \frac{2}{n-1} \bar{x}_n^i \sum_{j=0}^{n-1} x_j^i + \frac{n}{n-1} \left(\bar{x}_n^i\right)^2 \\ &= \frac{1}{n-1} \sum_{j=0}^{n-1} \left(x_j^i\right)^2 - \frac{n}{n-1} \left(\bar{x}_n^i\right)^2 \qquad (\star) \end{aligned}$$ 1

Similarly,

$$g_{n-1}^i = \frac{1}{n-2} \sum_{j=0}^{n-2} \left(x_j^i\right)^2 - \frac{n-1}{n-2} \left(\bar{x}_{n-1}^i\right)^2$$

$$\implies \frac{n-2}{n-1} g_{n-1}^i + \frac{1}{n-1} \left(x_{n-1}^i\right)^2 = \frac{1}{n-1} \sum_{j=0}^{n-1} \left(x_j^i\right)^2 - \left(\bar{x}_{n-1}^i\right)^2$$

$$\implies \frac{1}{n-1} \sum_{j=0}^{n-1} \left(x_j^i\right)^2 = \frac{n-2}{n-1} g_{n-1}^i + \frac{1}{n-1} \left(x_{n-1}^i\right)^2 + \left(\bar{x}_{n-1}^i\right)^2$$ 1

Substituting this into $(\star)$,

$$g_n^i = \frac{n-2}{n-1} g_{n-1}^i + \frac{1}{n-1} \left(x_{n-1}^i\right)^2 + \left(\bar{x}_{n-1}^i\right)^2 - \frac{n}{n-1} \left(\bar{x}_n^i\right)^2$$ 1

Also,

$$\begin{aligned} \bar{x}_n^i \quad &= \quad \frac{1}{n} \sum_{j=0}^{n-1} x_j^i \quad = \quad \frac{n-1}{n} \left(\frac{1}{n-1} \sum_{j=0}^{n-2} x_j^i \right) + \frac{x_{n-1}^i}{n} \\ &= \quad \frac{n-1}{n} \bar{x}_{n-1}^i + \frac{x_{n-1}^i}{n} \end{aligned}$$ 1

and substituting this into the above gives the required expression

	EXAMINATION SOLUTIONS 2016-17	Course M3/4S9
Question 4		Marks & seen/unseen
Part (e)(i)		seen
	Both the integrated autocorrelation time and average square jumping distance measure provide a measure of Markov chain mixing. They measure how well the chain explores the state space, and provide an indication of the level of dependence present in the simulated chain. The IACT can also be used as a measure of convergence speed for the algorithm.	1
Part (e)(ii)		seen
	For the integrated autocorrelation times, low values indicate good performance.	1
	For the average square jump distance, high values indicate good performance.	1
	Setter's initials Checker's initials	Page number

Imperial College London
Department of Mathematics

Examiner's Comments

Exam: M3/4S9, Stochastic Simulation **Session: 2016-2107**

Question 1

Please use the space below to comment on the candidates' overall performance in the exam. A brief paragraph highlighting common mistakes and parts of questions done badly (or well) is sufficient. Do not refer to individual candidates. The purpose of this exercise is to provide guidance to the external examiners, and to the candidates themselves, on how you feel the cohort faired. Your comments will be available to students online.

Overall, the cohort's performance on Question 1 was reasonable, with the majority of candidates obtaining over half the marks available. Generally speaking, most marks were picked up in the first half of the question [in parts (a)-(c)(i)]. In part (c)(ii), the majority were able to establish the Taylor series result required. For several candidates, there was a lack of clarity as to what quantity required bounding, with some choosing to bound the usual membership criterion for the Ratio of Uniforms, instead of the criterion for V, established in part (c)(i). The unseen material in part (d) also challenged the candidates, with fewer than expected spotting a possible link between the density in the question and an Exponential density.

In retrospect, this question was perhaps a little long, with the working required for part (c)(ii) in particular being too cumbersome.

Marker: J S Martin

Signature: ______ **Date:** 15/05/17

Please return with exam marks (one report per marker)

Imperial College London
Department of Mathematics

Examiner's Comments

Exam: M3/4S9, Stochastic Simulation **Session: 2016-2107**

Question 2

Please use the space below to comment on the candidates' overall performance in the exam. A brief paragraph highlighting common mistakes and parts of questions done badly (or well) is sufficient. Do not refer to individual candidates. The purpose of this exercise is to provide guidance to the external examiners, and to the candidates themselves, on how you feel the cohort faired. Your comments will be available to students online.

There was a large spread of performances in the second question. Part (b) proved a bit of a sticking point for some of the candidates, who left it out entirely; this was unexpected as it was all seen material. Generally speaking, those that attempted part (b) did reasonably well in that part. On the whole, the remainder of the question was well-attempted by most. There was some confusion in part (d)(ii) ('express the estimator [from part (d)(i)] in terms of an iid [Uniform] sample' – several candidates chose to derive a second estimator in terms of a Uniform sample, instead of re-expressing their existing estimator. A surprising error that cropped up in a number of scripts was the expression of the expectation as the integral of the pdf.

Marker: J S Martin

Signature: ______________ **Date:** 15/05/17

Please return with exam marks (one report per marker)

Imperial College London
Department of Mathematics

Examiner's Comments

Exam: M3/4S9, Stochastic Simulation **Session: 2016-2107**

Question 3

Please use the space below to comment on the candidates' overall performance in the exam. A brief paragraph highlighting common mistakes and parts of questions done badly (or well) is sufficient. Do not refer to individual candidates. The purpose of this exercise is to provide guidance to the external examiners, and to the candidates themselves, on how you feel the cohort faired. Your comments will be available to students online.

Again, there was a large spread of performances in this question. A number of errors were picked up in part (a), where candidates appear to have been attempting to replicate working from the notes from memory, without the requisite understanding. In general, those that attempted parts (b) and (c), though there was some confusion about the limits of summation required to calculate the probability of the test statistic appearing in the upper critical region.

Marker: J S Martin

Signature: ____________ **Date:** 15/05/17

Please return with exam marks (one report per marker)

Imperial College London
Department of Mathematics

Examiner's Comments

Exam: M3/4S9, Stochastic Simulation **Session: 2016-2107**

Question 4

Please use the space below to comment on the candidates' overall performance in the exam. A brief paragraph highlighting common mistakes and parts of questions done badly (or well) is sufficient. Do not refer to individual candidates. The purpose of this exercise is to provide guidance to the external examiners, and to the candidates themselves, on how you feel the cohort faired. Your comments will be available to students online.

Question 4 was reasonably well answered – the cohort (of 6) seem to have a solid grasp of both the theory and implementation of both standard MCMC and adaptive MCMC procedures. Part (d) involved some tricky manipulation, but was generally well-handled.

Marker: J S Martin

Signature: ______________ **Date:** 15/05/17

Please return with exam marks (one report per marker)