Imperial College London

Department of Mathematics

M3S9 Stochastic Simulation

Question	Examiner's Comments
Q 1	The question contained material that most students were familiar with, thus showing good performance. For parts (a), (b), (c) and (d)(i) variation in the marks was mainly due to deriving a rigourous mathematical formulation and explaining the definitions carefully. Part (d)(ii) required applying a theorem from the lecture notes to a novel problem and most of the students performed very well in uncovering this link and providing a full proof as required.
Q 2	This question turned out to be difficult and challenged the students. For part (a) (i) and (ii), most of the students had problems in deriving the complete proofs as required. However, performance for part (iii) was strong, showing that students were well familiar with the rejection sampling algorithm. Altogether, lower performance for (i) and (ii) was surprising as the material was classified as seen. Part (b) was well performed and required mainly familiarity with the lecture notes. Part (c) contained unseen material and this was reflected in the performance accordingly. Only few students performed well, applying the course content to a new problem, but most students were unable to derive meaningful solutions or leaving blank answers altogether.
Q 3	Similarly to question 2, this question challenged the students and resulted in a large variation in the marks. Part (a) was overall well performed. Part (a) (iii) contained unseen material and it was good to see that most of the students correctly applied the pre-testing principle for the algorithm. Part (b) showed larger variation, especially for part (i). A large proportion of students were unable to correctly formulate the complete density for the proposal distribution. Further, for (ii) and (iii) variation in performance was mainly due to completeness and rigour.

Imperial College London

Department of Mathematics

M4S9 Stochastic Simulation

Question	Examiner's Comments
Q 4	For this mastery question, part (a) challenged the students, whereas parts (b) and (c) showed smaller variation in the marks. Part (c) shows that although students were well familiar with the mastery material they faced more difficulties for parts (a) and (b) due to incomplete mathematical rigour.

Imperial College
London

M4/5S9

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)

May-June 2018

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

Stochastic Simulation

Date: Friday, 18 May 2018

Time: 10:00 AM - 12:00 PM

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.

- 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.
- Each question carries equal weight.
- Calculators may not be used.

1. (a) What is a sequence of pseudo-random numbers and how does it differ from a sequence of true random numbers?

   (b) Define a congruential pseudo-random number generator. Describe the properties of the generator and show how it can be used to generate uniform random variates.

   (c) Let F_X denote a distribution function of a continuous random variable $X \in \mathbb{R}$ and let F_X^{-1} denote the inverse of F_X. Show that if U is a uniform random variable on $(0,1)$, then $F_X^{-1}(U)$ follows the distribution function F_X. Also, show that $F_X(X)$ is uniformly distributed on $(0,1)$.

   (d) Applying the result of Question 1.(c),

      (i) show how uniform random variables can be transformed to Cauchy random variables and vice versa. The distribution function for the Cauchy distribution is given by $F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan(x)$, for $x \in \mathbb{R}$.

      (ii) describe a method to generate pairs of random variables with distribution function F_X that are negatively correlated. Show that the generated variables are negatively correlated.

2. (a) Consider two probability density functions $f(x)$ and $g(x)$, for $x \in \mathbb{R}$. Let M denote a constant such that $f(x) \leq Mg(x) < \infty$, for $x \in \mathbb{R}$.

   Let $(X_1, U_1 M g(X_1)), (X_2, U_2 M g(X_2)), \ldots$ denote a sequence of random vectors, where $U_i \sim U(0,1)$ and $X_i \sim g$, for $i \geq 1$. Let $j \geq 1$ denote the smallest value of the index i for which $U_i M g(X_i) \leq f(X_i)$.

      (i) What is the distribution of the X_j? Show the derivation of the result.

      (ii) What is the distribution of the j? Show the derivation of the result.

      (iii) Outline an algorithm that halts at iteration j and outputs X_j. What is this algorithm called?

   (b) Outline two different methods for generating $n > 1$ ordered uniform random variates on $(0,1)$ without resorting to generic sorting routines.

   (c) Using uniform order statistics, outline an algorithm to generate variables S_i, for $i = 1, \ldots, n$, that are uniformly distributed over the simplex

$$A_n = \{(x_1, \ldots, x_n) : x_i \geq 0, \sum_{i=1}^{n} x_i \leq 1\},$$

   for $n > 1$. Show that the algorithm is correct.

3. (a) Consider the following algorithm.

(1) Generate $X, Y \sim U(-1, 1)$.

(2) If $X^2 + Y^2 \leq 1$, return (X, Y). Else go to step (1).

(i) What is the probability that a generated pair (X, Y) is returned by the algorithm?

(ii) What is the distribution of a returned pair (X, Y) of the algorithm?

(iii) Testing the condition $|X| + |Y| \leq 1$, outline an extension of the algorithm that reduces the number of squaring operations in step (2) by 50%. Show why the reduction results.

(b) Consider a Metropolis-Hastings algorithm for generating a Markov chain $X_0, X_1, X_2 \ldots$, initialised at $X_0 \geq a$, that converges to a continuous-valued density that is proportional to $f(X)$, for $X \geq a$, and zero elsewhere. To generate a proposal Y at iteration n, assume a proposal distribution given by

$$q(Y|X_{n-1}) \propto \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y - X_{n-1})^2}{2\sigma^2}\right),$$

that is truncated to the interval $Y \geq a$.

The algorithm outline is as follows.

(1) Initialise the chain, $X_0 \geq a$, and set $n = 1$.

(2) Given X_{n-1} generate a proposal Y from the proposal density $q(Y|X_{n-1})$.

(3) Set $X_n = Y$ with probability α. Otherwise, set $X_n = X_{n-1}$.

(4) Replace n by $n + 1$ and go to step (2).

(i) Show that the proposal acceptance probability α is given by

$$\alpha = \min\left(\frac{f(Y)\left(1 - \Psi\left(\frac{a - X_{n-1}}{\sigma}\right)\right)}{f(X_{n-1})\left(1 - \Psi\left(\frac{a-Y}{\sigma}\right)\right)}, 1\right),$$

where Ψ denotes the cumulative distribution function for the normal distribution with zero mean and unit variance.

(ii) Comment how different values for σ^2 affect convergence of the algorithm.

(iii) Describe how the generated X_n, for $n = 1, \ldots, N$, may be used to estimate the mean of the distribution f? Focus your discussion on convergence and dependence of the chain.

4. (a) (i) Let f denote a probability density function on $\mathbb{R}$. Show that if a random vector $(X, U) \in \mathbb{R}^2$ is uniformly distributed on

$$A = \{(x, u) : x \in \mathbb{R}, 0 \leq u \leq f(x)\},$$

then X has density f on $\mathbb{R}$.

(ii) Based on this result explain why slice sampling works correctly.

(b) Consider the following procedure for slice sampling from a distribution that has the density function $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$, for $x \in \mathbb{R}$. Fill in the missing parts of the following algorithm and outline the complete algorithm.

(1) Initialise $x_0 \in \mathbb{R}$. Set $n = 1$.

(2) Generate $u \sim$

(3) Generate $x_n \sim U(a, b)$, where

$$a = \qquad \text{and } b = \qquad .$$

(4) Replace n by $n + 1$ and go to step (2).

(c) Compare slice sampling to Gibbs and Metropolis-Hastings algorithms.

BSc, MSc and MSci EXAMINATIONS (MATHEMATICS)

May – June 2018

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

Stochastic Simulation [SOLUTIONS]

Date: Friday, 18th May 2018

Time: 10:00–11:30 (M3 version); 10:00–12:00 (M4 version)

Time Allowed: 1.5 Hours for M3 paper; 2 Hours for M4 paper

This paper has *3 Questions (M3 version); 4 Questions (M4version).*

Candidates should start their solutions to each question in a new main answer book.

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

Statistical tables will not be provided.

- 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.
- Each question carries equal weight.
- Calculators may not be used.

1. (a) A sequence of pseudo-random numbers is a deterministic sequence of numbers, that approximates the same relevant statistical properties of a sequence of random numbers. The nth random number depends on the previous $n-1$ numbers and an initial nonrandom seed. seen ⇓ 1

(b) seen ⇓

$$X_{n+1} = (aX_n + b) \mod (m)$$

* X_0 is a non-random seed
* m modulus
* b shift
* a multiplier

Properties:

* X_n depends on X_{n-1} for $n \geq 0$
* $\exists i, k \in \{0, 1, \ldots, m\}$ s.t. $X_i = X_{i+k}$
* $X_i, X_{i+1}, \ldots, X_{i+k-1}$ repeats (period of length k)

$$U_n = \frac{X_n}{m} \in [0,1) \forall n$$

5

(c) For all $x \in \mathbb{R}$, seen ⇓

$$P(F_X^{-1}(U) \leq x) = P(U \leq F_X(x)) = F_U(F_X(x)) = F_X(x).$$

For all $0 \leq u \leq 1$,

$$P(F_X(X) \leq u) = P(X \leq F_X^{-1}(u)) = F_X(F_X^{-1}(u)) = u.$$

4

(d) (i) seen ⇓

$$F_X^{-1}(U) = \tan\left(\pi\left(U - \frac{1}{2}\right)\right)$$

Generate (i) $U \sim U(0,1)$ and (ii) set $X = \tan\left(\pi\left(U - \frac{1}{2}\right)\right)$.
Generate (i) $X \sim$ Cauchy and (ii) set $U = \frac{1}{2} + \frac{1}{\pi}\arctan(X)$. 3

(ii) Given $U \sim U(0,1)$, define $V = 1 - U$. Then generate pairs $F_X^{-1}(U)$ and $F_X^{-1}(V)$; they are negatively correlated. sim. seen
Note that $g(U) = F_X^{-1}(U)$ is a monotonically increasing function on $(0,1)$. Then $\mathrm{cor}(g(U), g(1-U)) < 0$.
Denote

$$E[g(U)] = E[g(1-U)] = \theta = \int_0^1 g(x)dx$$

and

$$\inf\{u : g(u) > \theta\} = 1 - t.$$

$$
\begin{aligned}
\mathrm{cov}(g(U), g(1-U)) &= E[(g(U)-\theta)\,(g(1-U)-\theta)] \\
&= E[g(U)\,(g(1-U)-\theta)] - \theta E[g(1-U)-\theta] \\
&= \int_0^1 g(u)\,(g(1-u)-\theta)\,du \\
&= \int_0^t g(u)\,(g(1-u)-\theta)\,du + \int_t^1 g(u)\,(g(1-u)-\theta)\,du \\
&< g(t)\int_0^1 (g(1-u)-\theta)\,du = 0.
\end{aligned}
$$

Thus $\mathrm{cor}(g(U), g(1-U)) < 0$. $\boxed{7}$

2. (a) (i) Let Y and U denote the random variates generated from g and uniform distribution, respectively. Let us define $h(Y) = \frac{f(Y)}{Mg(Y)}$ and express the condition equivalently as $U \le h(Y)$. Further, let G denote the distribution function of g. $\boxed{\text{seen} \Downarrow}$

The distribution function for the first value in the sequence that satisfies the condition $U \le h(Y)$ is given by $P(Y \le y | U \le h(Y))$. We know that

$$P(Y \le y | U \le h(Y)) = \frac{P(U \le h(Y) | Y \le y) P(Y \le y)}{P(U \le h(Y))}.$$

Here, we show that

$$\begin{aligned} P(U \le h(Y) | Y \le y) &= \frac{P(U \le h(Y), Y \le y)}{G(y)} \\ &= \frac{1}{G(y)} \int_{-\infty}^{y} P(U \le h(Y) | Y = w \le y) g(w) dw \\ &= \frac{1}{G(y)} \int_{-\infty}^{y} \frac{f(w)}{Mg(w)} g(w) dw \\ &= \frac{F(y)}{MG(y)}, \end{aligned}$$

where $F(y)$ denotes the probability distribution function of $f(y)$, and

$$P(U \le h(Y)) = \int_{-\infty}^{\infty} \frac{f(y)}{Mg(y)} g(y) dy = \frac{1}{M}.$$

Thus, we get

$$P(Y \le y | U \le h(Y)) = \frac{F(y)}{MG(y)} \frac{G(y)}{1/M} = F(y),$$

meaning that the distribution function of the X_j is given by F.

(ii) Let p denote the probability that the condition is accepted. The distribution for j is $P(j = k) = p(1-p)^{k-1}$, where $k = 1, 2, \ldots$ and $p = P(U \le h(Y)) = \int_{-\infty}^{\infty} \frac{f(y)}{Mg(y)} g(y) dy = \frac{1}{M}$.

(iii) (1) Generate $Y = y \sim g$

(2) Generate $U = u \sim U(0, 1)$

(3) If $u \le \frac{f(y)}{Mg(y)}$, set $X = y$. Otherwise go to step (1).

This algorithm is referred to as rejection sampling algorithm. $\boxed{8}$

(b) 1. Set $U_{(n+1)} = 1$ and generate $U_i \sim U(0, 1)$, for $i = 1, \ldots, n$. Recursively set $U_{(k)} = U_{(k+1)} U_k^{\frac{1}{k}}$, for $k = n, n-1, \ldots, 1$. $\boxed{\text{seen} \Downarrow}$

2. Generate $X_i \sim \text{Exp}(1)$, for $i = 1, \ldots, n+1$, and set $S_k = \sum_{i=1}^{k} X_i$. Define $U_{(k)} = \frac{S_k}{S_{n+1}}$, for $k = 1, \ldots, n$. $\boxed{6}$

(c) Generate $U_{(1)} \le \cdots \le U_{(n)}$ using either of the two methods described in 2.(b). $\boxed{\text{unseen} \Downarrow}$

Define $U_{(0)} = 0$ and set $S_k = U_{(k)} - U_{(k-1)}$, for $k = 1, \ldots, n$.

We know that $(U_1, \ldots, U_n)$ is distributed uniformly in the unit cube $(0, 1)^n$. The joint density of $(U_{(1)}, \ldots, U_{(n)})$ is uniformly distributed in the simplex

$$B_n = \{(x_1, \ldots, x_n) : 0 < x_1 < \ldots < x_n < 1\}.$$

The transformation $s_k = u_{(k)} - u_{(k-1)}$ has an inverse $u_{(k)} = \sum_{i=1}^{k} s_i$. The Jacobian of the transformation corresponds to one. Thus, $(S_1, \ldots, S_n)$ is uniformly distributed on the set A_n. $\boxed{6}$

3. (a) (i) The area of the unit-circle is given by π and the area of the rectangle is $2^2 = 4$. Thus the probability is given by $\frac{\pi}{4}$. seen $\Downarrow$

(ii) The algorithm outputs random points that are uniformly distributed inside the unit-circle.

2

(iii) (1) Generate $X, Y \sim U(-1, 1)$. unseen $\Downarrow$

(2) Compute $Z = |X| + |Y|$. If $Z \leq 1$, return (X, Y). Else, go to step (3).

(3) If $X^2 + Y^2 \leq 1$, return (X, Y). Else go to step (1).

Note that proportionally $\frac{\sqrt{2}^2}{2^2} = \frac{1}{2}$ of the candidates are accepted directly in step (2). Thus the number of squaring operations is reduced by 50%.

3

(b) (i) The density function for the proposals is given by sim. seen

$$q(Y|X_{n-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{\exp\left(-\frac{(Y-X_{n-1})^2}{2\sigma^2}\right)}{1 - \Psi\left(\frac{a-X_{n-1}}{\sigma}\right)}.$$

Hence the proposal acceptance probability is given by

$$\begin{aligned}\alpha &= \min\left(\frac{f(Y)q(X_{n-1}|Y)}{f(X_{n-1})q(Y|X_{n-1})}, 1\right) \\ &= \min\left(\frac{f(Y)\left(1 - \Psi\left(\frac{a-X_{n-1}}{\sigma}\right)\right)}{f(X_{n-1})\left(1 - \Psi\left(\frac{a-Y}{\sigma}\right)\right)}, 1\right).\end{aligned}$$

(ii) If σ^2 is chosen too large then the proposal mechanism will explore the state space well but will often propose proposals in regions where the the target density is low such that the probability of rejection is high.
If σ^2 is chosen too small, on the other hand, most of the proposals will be accepted but this results in poor exploration of the state space and significant dependence in the resulting Markov chain.

(iii) Before the algorithm converges the samples cannot be used to represent the distribution $f(X)$. Assume after a specified number of iterations $B < N$ convergence has occurred. Monte Carlo estimate for the mean is given by $\frac{1}{N-B}\sum_{m=1}^{N-B} X_{B+m}$. In the limit $N \to \infty$ it is known that the estimate approaches the true expectation with probability one, provided that $\int |x| f(x)dx < \infty$. For moderate values of N a common step is to take every kth iteration and discard the rest to reduce dependence of the Markov chain. The value for k may be chosen by inspecting the autocorrelation sequence.

15

4. (a) (i) The joint distribution is given by $p(U,X) = \frac{1}{\int\int_0^{f(x)} du dx} = 1$, for $(U,X) \in A$. Marginalisation of U of the joint density gives $p(X) = \int_0^{f(X)} p(U,X) dU = f(X)$. unseen ⇓ 3

(ii) The idea in slice sampling sampling is to generate a Markov chain (X_i, U_i) that converges as $i \to \infty$ to a uniform distribution under the plot of the density function f. Then by the result of Question 4.(a)(i) the variables X_n follow the distribution f and the corresponding values for U_n may be discarded. unseen ⇓ 3

(b) Let $f(x)$ denote $N(x|0,1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$. unseen ⇓

(1) Initialise $x_0 \in \mathbb{R}$. Set $n = 1$.

(2) Generate $u \sim U(0, f(x_{n-1}))$.

(3) Generate $x_n \sim U(a, b)$, where

$$a = -\sqrt{-2\log(\sqrt{2\pi}u)} \text{ and } b = \sqrt{-2\log(\sqrt{2\pi}u)}.$$

(4) Set $n \leftarrow n + 1$ and go to step (2). 6

(c) Gibbs sampling is not feasible when the conditional distributions of interest may not be computed in closed analytical form. For slice sampling the conditional distributions are not necessary to derive. unseen ⇓

Metropolis-Hastings requires choosing and tuning a proposal distribution that may not be trivial in most cases. This is not necessary for slice sampling method.

Compared to both approaches slice sampling is able to reduce dependence of the chain by making large jumps in the state space. Thus the method may lead to faster convergence with smaller dependence of the Markov chain. 8