Imperial College London

MATH97085

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May-June 2020

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

Stochastic Simulation

Date: 26th May 2020

Time: 09.00am - 11.00am (BST)

Time Allowed: 2 Hours

Upload Time Allowed: 30 Minutes

This paper has 4 Questions.

Candidates should start their solutions to each question on a new sheet of paper.

Each sheet of paper should have your CID, Question Number and Page Number on the top.

Only use 1 side of the paper.

Allow margins for marking.

Any required additional material(s) will be provided.

Credit will be given for all questions attempted.

Each question carries equal weight.

SUBMIT YOUR ANSWERS AS <u>ONE PDF</u> TO THE RELEVANT DROPBOX ON BLACKBOARD INCLUDING A COMPLETED COVERSHEET WITH YOUR CID NUMBER, QUESTION NUMBERS ANSWERED AND PAGE NUMBERS PER QUESTION.

.

1. (a) The congruential generator to generate pseudo-standard uniform variates, U_i, defines

$$U_i = \frac{X_i}{m},$$

where,

$$X_i = (aX_{i-1} + b) \ \text{mod}\ m, \quad i = 1, 2, 3, \ldots$$

with integer m and $X_0, a, b \in \{0, 1, \ldots, m-1\}$.

(i) What is the value of b for a multiplicative generator? (1 mark)

(ii) Define the period, k, of such a generator. What is the largest value of k when $b \neq 0$? (3 marks)

(iii) Consider the following R code:

```
cg = function(n, m, a, b, x0)
{
x = c(x0, rep(0,(n-1)))
for(i in 2:n)
      x[i] = (a * x[i - 1]+b) %% m
data.frame(x = x/m)
}
x1 = cg(500, 2^(40), 5^3, 0, 1)
x2 = cg(500, 2^(45), 11, 0, 1)
x3 = cg(500, 2^(12), 2^6, 1, 2^2)
```

Find the period for each of the generators which produce x1, x2, and x3. (4 marks)
The figures below show lag-one scatter plots of the data produced by the R code above.

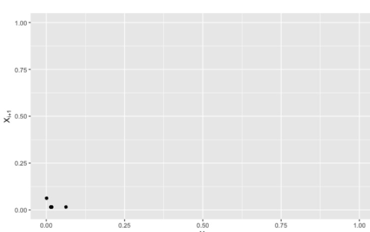
Figure A

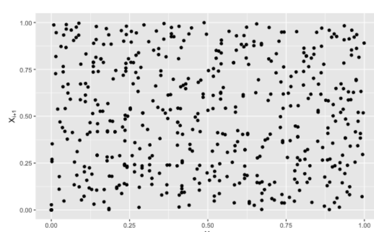
Figure B

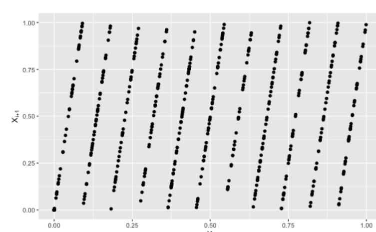
Figure C

To which sequence (out of x1, x2, and x3) do each of the Figures (out of A, B and C) correspond? (2 marks)

(b) Given a sequence of n variates which are claimed to be from a $U(0, 1)$ distribution, describe how you would perform a chi-squared goodness-of-fit test on the numbers (between 0 and 99) formed by taking the first two decimal places of the generated sequence. (5 marks)

(c) Describe two methods to investigate the dependence structure of a sequence of generated $U(0, 1)$ variates. (5 marks)

(Total: 20 marks)

2. (a) Describe in detail the rejection method for generating samples from a probability density function (pdf) $f_X(\cdot)$ using a rejection envelope $g_Y(\cdot)$. (3 marks)

(b) Describe in detail the Metropolis-Hastings (M-H) scheme to generate samples from a pdf $f_X(\cdot)$ using a transition kernel $q(\cdot \mid \cdot)$. (4 marks)

(c) Describe in detail the inversion algorithm to generate samples from a pdf $f_X(\cdot)$. (2 marks)

(d) Consider schemes to simulate samples from a pdf $f_X(x) \propto f_X^*(x)$, where,

$$f_X^*(x) = \begin{cases} \exp(-|x|), & -1 < x < 1; \\ 0, & \text{otherwise.} \end{cases}$$

(i) Determine the acceptance probability of a rejection scheme to generate samples from $f_X(x)$ using a uniform envelope, i.e.

$$g_Y(y) = \begin{cases} 0.5, & -1 < y < 1; \\ 0, & \text{otherwise.} \end{cases}$$

(2 marks)

(ii) Consider the following R code to implement a M-H scheme to simulate from $f_X(\cdot)$,

```
f = function(x){
  c=2*(1-exp(-1))
  ifelse(x>-1 & x<1, exp(-abs(x))/c, 0)
}
qpdf = function(y, sigma){
  rnorm(1, mean=y, sd=sigma)
}
MH = function(n){
  x=vector("double",n); x[1]=0; u=runif(n); sigma=0.1
  for(i in 2:n){
    y=qpdf(x[i-1],sigma)
    x[i] = ifelse(u[i]<min(****,1), y, x[i-1])
  }
  return(x)
}
```

(I) What transition kernel is being used in this implementation? (1 mark)

(II) What R code should replace **** for this code to work? (1 mark)

(III) Describe the effect of changing the variable sigma. (2 marks)

(e) Why is it beneficial for a simulation scheme to generate independent samples? (2 marks)

(f) With justification, which approach, out of rejection, inversion and M-H would you choose to apply in order to generate samples from the pdf outlined in part (d). (3 marks)

(Total: 20 marks)

3. Consider the following integral,

$$\theta = \int h(x)\,\mathrm{d}x = \int \phi(x) f_X(x)\,\mathrm{d}x.$$

(a) Given $X_1, \dots, X_n \sim f_X(\cdot)$.

(i) Write down the Monte-Carlo estimator $\hat{\theta}$ for θ in terms of the $X_i, i = 1, \dots, n$. (1 mark)

(ii) Prove that $\hat{\theta}$ is unbiased for θ. (2 marks)

(iii) Find an expression for the variance of $\hat{\theta}$ in terms of $\phi(\cdot), f_X(\cdot)$ and θ. (2 marks)

(b) Consider estimating the integral,

$$\theta = \int_0^1 \exp(-x^3)\,\mathrm{d}x.$$

(i) Describe the hit-or-miss method for estimating θ.

Consider the following R code:

```
a=0;b=1;c=1;n=10000
u=runif(n, min=a,max=b)
v=runif(n, min=0,max=c)
thetahat = ****
```

What R code should replace **** in the code above in order for thetahat to be the hit-or-miss Monte-Carlo estimate of θ? (4 marks)

(ii) Describe how you would construct the crude Monte-Carlo estimator for θ given a sequence of independent variates $U_1, \dots, U_n \sim U(0, 1)$. (2 marks)

(iii) Using inversion, one can generate from the following pdf,

$$f_X(x) = \begin{cases} \dfrac{k\cos(kx)}{\sqrt{1-\mathrm{e}^{-2}}}, & 0 < x < 1; \\ 0, & \text{otherwise,} \end{cases}$$

where $k = \arccos(\mathrm{e}^{-1})$.

Consider the following R implementation of an inversion scheme to generate from $f_X(\cdot)$,

```
n=10000;k=acos(exp(-1));k1=sqrt(1-exp(-2))
u=runif(n)
x= ****
```

What R code should replace **** in the section of code above so that x contains variates with pdf $f_X(\cdot)$? (4 marks)

(iv) Determine an estimator of θ given a sequence of independent variates $X_1, \dots, X_n \sim f_X(\cdot)$, described in part (iii). (3 marks)

(v) Without performing any calculations, but giving your reasoning, which estimator out of those calculated in parts (i), (ii) and (iv), would you expect to have the lowest variance? (2 marks)

(Total: 20 marks)

4. (a) Given random variables X and Y with joint probability density (pdf) function $f_{X,Y}(x, y)$, and conditional probability density functions $f_{Y|X}(y|x)$ and $f_{X|Y}(x|y)$, outline the two-stage Gibbs sampler to generate $(X_t, Y_t), t = 1, 2, \ldots$ from the joint pdf. (3 marks)

(b) It is supposed that X and Y have a joint distribution for which the conditional pdfs are $X|Y = y \sim \text{Exp}(y)$ and $Y|X = x \sim \text{Exp}(x)$.

(i) Show that these conditional distributions would be consistent with a joint distribution whose pdf is $f_{X,Y}(x, y) \propto \exp(-xy),\ x, y > 0$.
Explain why this function $f_{X,Y}(x, y)$ is not actually a valid pdf. (3 marks)

(ii) What issue with Gibbs sampling does the lack of validity of $f_{X,Y}(x, y)$ highlight? (1 mark)

(iii) Now suppose that the X and Y have a joint distribution with pdf $f_{X,Y}(x, y) \propto \exp(-xy)$ for $x \in (0, 1), y \in (0, 1)$ and $f_{X,Y}(x, y) = 0$ otherwise.
In this case, outline the Gibbs sampler to generate from $f_{X,Y}(x, y)$. (4 marks)

(c) Let the joint pdf of X and U be given by

$$f_{X,U}(x, u) = \begin{cases} 1, & 0 < u < \frac{8}{\pi}\sqrt{x(1-x)},\ 0 < x < 1; \\ 0, & \text{otherwise.} \end{cases}$$

(i) Verify that

$$f_X(x) = \begin{cases} \frac{8}{\pi}\sqrt{x(1-x)}, & 0 < x < 1; \\ 0, & \text{otherwise.} \end{cases}$$

(1 mark)

(ii) Prove that $U|X = x \sim U(0, a)$ and $X|U = u \sim U(b, c)$ where you should specify a in terms of x, and b and c in terms of u. (5 marks)

(iii) An implementation of the Gibbs sampler to simulate from the joint pdf is given below.

```
N=5000
X=c(runif(1));U=c(runif(1))
for(t in 1:N){
  a=**1**
  U[t+1]=runif(1,0,a)
  b=**2**
  c=**3**
  X[t+1]=runif(1,b,c)
}
```

What `R` code should replace `**1**`, `**2**` and `**3**` in order for `X` to contain variates with pdf $f_X(\cdot)$? (3 marks)

(Total: 20 marks)

Module: MATH96054/MATH97085
Setter: McCoy
Checker: Hallsworth
Editor: Hallsworth
External: Jennison
Date: May 18, 2020
Version: Draft version for checking

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)

May – June 2020

MATH96054/MATH97085 Stochastic Simulation (Solutions)

The following information must be completed:

Is the paper suitable for resitting students from previous years: Yes

Category A marks: available for basic, routine material (excluding any mastery question) (40 percent = 32/80 for 4 questions):
1(a) 3 marks; 1(c) 5 marks; 2(a)(b) 7 marks; 2(c) 2 marks; 2(d)(ii) 2 marks; 3(a) 5 marks. Total 24 marks.

Category B marks: Further 25 percent of marks (20/ 80 for 4 questions) for demonstration of a sound knowledge of a good part of the material and the solution of straightforward problems and examples with reasonable accuracy (excluding mastery question):
2(d)(i) 2 marks; 2(iii) 2 marks; 2(e) 2 marks; 3(b)(ii) 2 marks; 3(b)(iv) 3 marks; 3(b)(v) 2 marks. Total 13 marks.

Category C marks: the next 15 percent of the marks (= 12/80 for 4 questions) for parts of questions at the high 2:1 or 1st class level (excluding mastery question):
1(b) 5 marks; 2(f) 5 marks; 3(b)(i) 4 marks. Total 12 marks.

Category D marks: Most challenging 20 percent (16/80 marks for 4 questions) of the paper (excluding mastery question):
1(a)(iii) 7 marks; 3(b)(iii) 4 marks. Total 11 marks.

Signatures are required for the final version:

Setter's signature	Checker's signature	Editor's signature
....	....	....

—— Temporary front page ——

BSc, MSc and MSci EXAMINATIONS (MATHEMATICS)

May – June 2020

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

Stochastic Simulation (Solutions)

Date: Tuesday, 26th May 2020

Time: AM

Time Allowed: 1.5 Hours for MATH96 paper; 2 Hours for MATH97 papers

This paper has *3 Questions (MATH96 version); 4 Questions (MATH97 versions)*.

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.

seen ⇓

1. (a) (i) $b = 0$ for a multiplicative generator.

1 A

(ii) The period k is the smallest positive integer such that

$$X_{i+k} = X_i, \ \forall i$$

2 A

The sequence $\{X_i, \ldots, X_{i+k-1}\}$ repeats.
If $b \neq 0$ the largest value of k is m (as $i \mod m \in \{0, 1, \ldots, m-1\}, i \in \mathbb{N}$).

1 A

(iii) A multiplicative generator with $m = 2^\beta$ has period $m/4$ if $a \mod 8 \equiv 3$ or 5 and X_0 is odd.

meth seen ⇓

x1: $m = 2^{40}, a = 5^3, b = 0, x_0 = 1$, so this is multiplicative and $a \mod 8 = 5$ so period $= 2^{38}$.
x2: $m = 2^{45}, a = 11, b = 0, x_0 = 1$, so this is multiplicative and $a \mod 8 = 3$ so period $= 2^{43}$.
x3: $m = 2^{12}, a = 2^6, b = 1, x_0 = 2^2$, so,

$$\begin{aligned} x_1 &= (2^6 2^2 + 1) \mod 2^{12} = 2^8 + 1, \\ x_2 &= (2^6(2^8 + 1) + 1) \mod 2^{12} = 2^6 + 1, \\ x_3 &= (2^6(2^6 + 1) + 1) \mod 2^{12} = 2^6 + 1, \end{aligned}$$

4 D

so $x_{i+1} = x_i, \ i > 1$ and hence the period =1.

unseen ⇓

Figure A is the plot for sequence x3 as it eventually gets stuck at the value $(2^6+1)/2^{12}$.
Sequence x2 has $a = 11$ and hence is associated with Figure C (all points on lines with gradient 11)
Finally, x1 is associated with Figure B.

2 D

sim. seen ⇓

(b) Let $v_i, i = 1, \ldots, n$ be the number represented by the first two decimal places of u_i, i.e. $v_i = \lfloor 100u_i \rfloor$, alternatively, let d_{1i} and d_{2i} represent the first two decimal places of the observed $u_i, i = 1, \ldots, n$ respectively and let $v_i = 10d_{1i} + d_{2i}$. Let,

$$O_j = \sum_{i=1}^{n} I_{[v_i=j]}, \quad j = 0, \ldots, 99; \text{ where } I_{[v_i=j]} = \begin{cases} 1 & \text{if } v_i = j; \\ 0 & \text{otherwise.} \end{cases}$$

If the sequence follows a standard uniform distribution then we would expect each of the numbers $0, \ldots, 99$ to occur with equal probability so $E_i = n/100$.
Form the test statistics

$$\chi^2 = \sum_{i=0}^{99} \frac{(O_i - E_i)^2}{E_i}.$$

Under the null hypothesis that the sequence is uniformly distributed the test statistic will approximately follow a Chi-Squared distribution with $100-1 = 99$ degrees of freedom. The value of the test statistic is then compared to the percentage points of the χ^2_{99} distribution to see if there is evidence to reject the null hypothesis. 5 C

seen ⇓

(c) The dependence structure of a sequence, $U_1, \ldots, U_n$ of generated $U(0,1)$ variates can be tested by the following:

* Correlogram: Plot of the sample autocorrelation sequence

$$\hat{\rho_k} = \frac{\frac{1}{n}\sum_{i=1}^{n-k}(U_i - \bar{U})(U_{i+k} - \bar{U})}{\frac{1}{n}\sum_{i=1}^{n}(U_i - \bar{U})^2} :$$

against k. An approximate 95% confidence interval (CI) for the estimated autocorrelation sequence is

$$-\frac{1}{n} \pm \frac{1.96}{\sqrt{n}};$$

Values of $\hat{\rho_k}$ that lie outside the CI give evidence to reject the hypothesis that the sequence is independent.

* Serial test for digits: Let n_{jk} be the number of times the digit j is followed by the digit k. In this case we have

$$O_0 = n_{00}, O_1 = n_{01}, \ldots, O_9 = n_{09}, O_{10} = n_{10}, \ldots, O_{99} = n_{99}$$

$$E_0 = \tfrac{n}{100}, E_1 = \tfrac{n}{100}, \ldots, E_{99} = \tfrac{n}{100}.$$

This can be tested with a Chi-Squared test.

* Gap Test: Choose a digit, say 3, then record the length of the subsequence lying between occurrences of the digit 3.

$$X_1, X_2, X_3, 3, \underbrace{X_5, X_6, \ldots, X_{k+4}}_{\text{gap length}=\,k}, 3, \ldots$$

If the sequence is uniform and independent, the distribution of gap lengths, K, should be Geometric($\frac{1}{10}$), which can be tested with a Chi-Squared test. 5 A seen $\Downarrow$

2. (a) Generating X with a known pdf $f_X(\cdot)$ using rejection employs the use of a rejection envelope $g_Y(\cdot)$ (which we can generate from). The pdf $g_Y(\cdot)$ must have the following properties:

   * the support of g_Y encompasses that of f_X, i.e. $f_X(x) > 0 \Rightarrow g_Y(x) > 0$,
   * there exists $M > 0$ such that $\forall x$ s.t. $f_X(x) > 0$,

$$\frac{f_X(x)}{g_Y(x)} \leq M < \infty.$$

Rejection Sampling Algorithm

1. Generate $Y = y \sim g_Y(.)$.
2. Generate $U = u \sim U(0,1)$.
3. If $u \leq \frac{f_X(y)}{M g_Y(y)}$ set $X = y$.
4. Otherwise GOTO 1.

3 A

(b) Generating X from known pdf $f_X(\cdot)$ using Metropolis-Hastings (M-H), employs a transition kernel $q(y|x) \geq 0$, which satisfies

$$\int q(y|x)dy = 1.$$

$q(y \mid x)$ and should be chosen to ensure irreducibility (the conditional pdf should be non-zero over the range of X).

Metropolis-Hastings Procedure (continuous state space):

1. Initialise the chain: start from an arbitrary X_0, possibly sampled from an initial pdf/prior. Set $n = 1$.
2. Given $X_{n-1} = x$, generate a candidate value $Y = y$ from the proposal density $q(y|x)$.
3. Set $X_n = y$ with probability $\alpha(x, y)$, where

$$\alpha(x,y) = \min\left\{\frac{f_X(y)q(x|y)}{f_X(x)q(y|x)}, 1\right\},$$

   otherwise, set $X_n = x$.
4. Replace n by $n+1$ and return to Step 2.

Note that as X_0 is not generated directly from f_X the algorithm needs a burn-in period. Also, MH does not produce independent samples. 4 A

meth seen $\Downarrow$

(c) Generating X from known pdf $f_X(\cdot)$ using inversion requires calculating the inverse cdf $F_X^{-1}(x)$.

Inversion Algorithm:

1. Generate $U = u \sim U(0,1)$
2. Set $X = x = F_X^{-1}(u)$

2 A

(d) (i)

$$M = \sup_x \frac{f_X^*(x)}{g_Y(x)} = \sup_{-1<x<1} 2\mathrm{e}^{-|x|} = 2.$$

Acceptance probability, θ, is given by

$$\theta = \frac{\int_{-1}^{1} f_X^*(x)\,\mathrm{d}x}{M} = \frac{2\int_0^2 \mathrm{e}^{-x}\,\mathrm{d}x}{2}$$
$$= \left[-\mathrm{e}^{-x}\right]_0^1 = 1 - \mathrm{e}^{-1}.$$

$\boxed{2\ \mathrm{B}}$

(ii) (I) The transition kernel is a Normal density centred at the current value with variance 0.01,

$$q(y|x) = \frac{10}{\sqrt{2\pi}}\mathrm{e}^{-50(y-x)^2},\ y \in \mathbb{R}.$$

$\boxed{1\ \mathrm{B}}$

(II) `****` should be replaced with `f(y)/f(x[i-1])`. Note that $q(y|x)$ is symmetric, so does not need to be included as the terms cancel. $\boxed{1\ \mathrm{B}}$

(III) `sigma` controls how close the distribution of the proposal around the current value. Reducing `sigma` means that the proposed values are accepted with higher probability, will be more correlated and slowly mixing with less opportunity to explore the space. If `sigma` is set too high, then the acceptance probability may be too low and be inefficient. $\boxed{2\ \mathrm{A}}$

(e) Many statistical procedures rely on the independence assumption either directly or through the Central Limit Theorem. $\boxed{2\ \mathrm{B}}$

(f) Answer should justify with a discussion of computational complexity and timing considerations. Accept any well reasoned answer based on e.g. avoiding complex calculations, computational time or a discussion of excluding M-H because of lack of independence. $\boxed{3\ \mathrm{C}}$

seen ⇓

3. (a) (i)

$$\theta = \mathsf{E}\left(\phi(x)\right), \quad \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\phi(X_i).$$

1 A

(ii)

$$\mathsf{E}(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}(\phi(X_i)) = \frac{1}{n}\sum_{i=1}^{n}\theta = \theta.$$

2 A

So $\hat{\theta}$ is unbiased for θ.

(iii)

$$\begin{aligned}\mathsf{var}(\hat{\theta}) &= \frac{1}{n^2}\sum_{i=1}^{n}\mathsf{var}(\phi(X_i)) = \frac{1}{n}\left(E(\phi^2(X)) - \theta^2\right) \\ &= \frac{1}{n}\left[\int \phi^2(x) f_X(x)\,\mathrm{d}x - \theta^2\right]\end{aligned}$$

2 A

part seen ⇓

(b) (i) Given $\theta = \int_a^b h(x)\,\mathrm{d}x$ with $0 \le h \le c$. We generate points uniformly in the rectangle $(a, b) \times (0, c)$ and estimate the area from the proportion that lie under $h(x)$.

Hit-or-Miss Algorithm:

1. Sample $\quad U = u_i \sim U(a, b), \qquad V = v_i \sim U(0, c), \qquad i = 1, \ldots, n$
2. Define

$$\hat{\theta} = \underbrace{c(b-a)}_{\text{area of rectangle}} \underbrace{\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(v_i \le h(u_i))}_{\text{frequency of ``dart'' under } h}$$

We have $h(x) = \exp(-x^3), x \in (0, 1)$.

So $a = 0, b = 1$ and $c = \sup_{x\in(0,1)} \exp(-x^3) = 1$ and `****` should be replaced with the `R` code

```
c*(b-a)*mean(v <= exp(-u^3)).
```

Alternatively, `c*(b-a)*length(v[v<=exp(-u^3)])/n`,

or `c*(b-a)*sum(v <= exp(-u^3))/n`.

4 C

meth seen ⇓

(ii) We have

$$\theta = \int_0^1 \exp(-x^3)\,\mathrm{d}x = \int_a^b \phi(x) f_X(x)\,\mathrm{d}x,$$

where $f_X(x) = 1,\ x \in (0, 1)$ and $\phi(x) = \exp(-x^3),\ x \in (0, 1)$. The crude MC estimator is given by,

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\exp(-U_i^3).$$

2 B

(iii) We have, for $x \in (0,1)$,

$$F_X(x) = \int_0^x \frac{k\cos(ky)}{\sqrt{1-\mathrm{e}^{-2}}}\,\mathrm{d}y = \frac{1}{\sqrt{1-\mathrm{e}^{-2}}}\left[\sin(ky)\right]_0^x,$$
$$= \frac{\sin(kx)}{\sqrt{1-\mathrm{e}^{-2}}}.$$

To find $F_X^{-1}(\cdot), x \in (0,1)$,

$$U = \frac{\sin(kx)}{\sqrt{1-\mathrm{e}^{-2}}} \Rightarrow X = F_X^{-1}(U) = \frac{1}{k}\arcsin(U\sqrt{1-\mathrm{e}^{-2}}).$$

To generate from $f_X(\cdot)$ using inversion:

1. Generate $U = u \sim U(0,1)$.
2. Set $X = x = \arcsin(u\sqrt{1-\mathrm{e}^{-2}})/k$.

Hence ***** should be replaced with the commands `x=asin(u*k1)/k`. $\boxed{4\ \mathrm{D}}$

(iv) We have

$$\theta = \int_0^1 \exp(-x^3)\,\mathrm{d}x = \int_a^b \phi(x) f_X(x)\,\mathrm{d}x,$$

where

$$f_X(x) = \frac{k\cos(ky)}{\sqrt{1-\mathrm{e}^{-2}}},\ x \in (0,1),\ \text{ and hence }\ \phi(x) = \frac{\exp(-x^3)\sqrt{1-\mathrm{e}^{-2}}}{k\cos(kx)},\ x \in (0,1).$$

The MC estimator is given by

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \frac{\exp(-X_i^3)\sqrt{1-\mathrm{e}^{-2}}}{k\cos(kX_i)}.$$

$\boxed{3\ \mathrm{B}}$

(v) We would expect the estimator outlined in (iv) to have the lowest variance. We know that hit-or-miss always has higher variance than crude MC. The density outlined in (iv) mimics the shape of $\exp(-x^3)$ better than the uniform (cf Importance sampling). $\boxed{2\ \mathrm{B}}$

4. (a) Two-stage Gibbs sampler given the conditional pdfs for $f_{X,Y}(x|y)$ and $f_{Y|X}(y|x)$.
Set $X_0 = x_0$. Set $t = 1$.

1. Generate $Y_t = y_t \sim f_{Y|X}(\cdot|x_{t-1})$;
2. Generate $X_t = x_t \sim f_{X|Y}(\cdot|y_t)$;
3. Set $t = t + 1$ and goto 1. $\boxed{3}$

(b) We have $f_{X|Y}(x|y) = y\exp(-xy), x > 0$ and $f_{Y|X}(y|x) = x\exp(-xy), y > 0$.

(i)

$$f_{X,Y}(x,y) \propto \exp(-xy) \Rightarrow f_X(x) \propto \int_0^\infty \exp(-xy)\,\mathrm{d}y = \left[\frac{\exp(-xy)}{-x}\right]_0^\infty = \frac{1}{x},\ x > 0,$$

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \propto x\exp(-xy),\ x > 0.$$

So $Y|X = x \sim \mathrm{Exp}(x)$. Similarly $X|Y = y \sim \mathrm{Exp}(y)$ as required for consistency with joint distribution. $\boxed{2}$

This is not a valid pdf as $\int_0^\infty 1/y\,\mathrm{d}y$ is not finite. $\boxed{1}$

(ii) This can be a danger as you can implement the Gibbs sampler with conditional distributions without having to verify that that they correspond to a valid joint distribution. $\boxed{1}$

(iii) If the conditional pdfs are truncated to $(0,1)$, then we have

$$f_{Y|X}(y|x) = \frac{x\exp(-xy)}{\int_0^1 x\exp(-xy)\,\mathrm{d}y} = \frac{x\exp(-xy)}{1-\mathrm{e}^{-x}}$$

Similarly for $f_{X|Y}$. To generate from $f_{Y|X}$ (and $f_{X|Y}$) we can use inversion. For $0 < y < 1$, we have

$$F_{Y|X}(y|x) = \int_0^y \frac{x\exp(-xt)}{1-\mathrm{e}^{-x}}\,\mathrm{d}t = \frac{1-\mathrm{e}^{-xy}}{1-\mathrm{e}^{-x}}.$$

Setting

$$U = \frac{1-\mathrm{e}^{-xY}}{1-\mathrm{e}^{-x}} \Rightarrow U(1-\mathrm{e}^{-x}) = 1-\mathrm{e}^{-xY} \Rightarrow -xY = \log(1-U(1-\mathrm{e}^{-x}))$$
$$\Rightarrow Y = F_{Y|X}^{-1}(U|x) = -\frac{1}{x}\log(1-U(1-\mathrm{e}^{-x}))$$

Gibbs sampler:

1. Set $t = 1$, $X_0 = x_0$.
2. Generate $U = u \sim U(0,1)$, set $Y_t = y_t = -\frac{1}{x_{t-1}}\log(1-u(1-\mathrm{e}^{-x_{t-1}}))$, then $Y_t \sim \mathrm{Exp}_{(0,1)}(x_{t-1})$, where $\mathrm{Exp}_{(0,1)}(\cdot)$ is an exponential distribution truncated to $(0,1)$.
3. Generate $U = u \sim U(0,1)$, set $X_t = x_t = -\frac{1}{y_t}\log(1-u(1-\mathrm{e}^{-y_t}))$, then $X_t \sim \mathrm{Exp}_{(0,1)}(y_t)$.
4. Set $t = t + 1$ and goto 1. $\boxed{4}$

(c) (i)

$$f_X(x) = \int f_{X,U}(x,u)\,\mathrm{d}u = \int_0^{8\sqrt{x(1-x)}/\pi} 1\,\mathrm{d}u = \frac{8}{\pi}\sqrt{x(1-x)}, \quad 0 < x < 1,$$

as required. $\boxed{1}$

(ii)

$$f_{U|X}(u|x) = \frac{f_{U,X}(u,x)}{f_X(x)} = \frac{\pi}{8\sqrt{x(1-x)}}, \quad 0 < u < \frac{8}{\pi}\sqrt{x(1-x)}.$$

i.e $U|X = x \sim U(0,a)$ where $a = \frac{8}{\pi}\sqrt{x(1-x)}$. Now,

$$f_U(u) = \int_{0<u<\frac{8}{\pi}\sqrt{x(1-x)}} 1\,\mathrm{d}x$$

$$\begin{aligned} u < \frac{8}{\pi}\sqrt{x(1-x)} &\Rightarrow \frac{\pi u}{8} < \sqrt{x(1-x)} \Rightarrow \frac{\pi^2u^2}{64} < x(1-x) \Rightarrow x^2 - x + \frac{\pi^2u^2}{64} < 0 \\ &\Rightarrow \left(x - \frac{1-\sqrt{1-\frac{4\pi^2u^2}{64}}}{2}\right)\left(x - \frac{1+\sqrt{1-\frac{4\pi^2u^2}{64}}}{2}\right) < 0 \\ &\Rightarrow b = \frac{1}{2} - \frac{1}{8}\sqrt{16-\pi^2u^2} < x < \frac{1}{2} + \frac{1}{8}\sqrt{16-\pi^2u^2} = c. \end{aligned}$$

So

$$f_U(u) = \int_b^c 1\,\mathsf{d}x = \frac{1}{4}\sqrt{16-\pi^2u^2}, \quad 0 < u < \frac{4}{\pi},$$

and

$$f_{X|U}(u|x) = \frac{4}{\sqrt{16-\pi^2u^2}}, \quad \frac{1}{2} - \frac{1}{8}\sqrt{16-\pi^2u^2} < x < \frac{1}{2} + \frac{1}{8}\sqrt{16-\pi^2u^2}$$

i.e. $X|U = u \sim U(b,c)$ where $b = \frac{1}{2} - \frac{1}{8}\sqrt{16-\pi^2u^2}$ and $c = \frac{1}{2} + \frac{1}{8}\sqrt{16-\pi^2u^2}$. $\boxed{5}$

(iii) From part (ii):

```
**1** = 8*sqrt(X[t]*(1-X[t]))/pi
**2**  =0.5-(1/8)*sqrt(16-U[t+1]^2*pi^2)
**3** = 0.5+(1/8)*sqrt(16-U[t+1]^2*pi^2)
```

$\boxed{3}$

Commentary

Q1 Mainly a straightforward question, though students may find a(iii) challenging, for (b) they have seen the test for a single digit.

Q2 The first two parts are bookwork, though we concentrated more on the discrete case, so the MH scheme may be a little more challenging.

Q3 Mainly application of standard results, but the inversion is a little fiddly and the `R` code sections may be harder for them as this has not appeared in previous exams.

Q4 Based on Chapter 7 of "Introducing Monte Carlo Methods with R, Use R" by Robert and Casella. They have only seen the basic Gibbs sampler in lectures.

If your module is taught across multiple year levels, you might have received this form for each level of the module. You are only required to fill this out once for each question.

Please record below, some brief but non-trivial comments for students about how well (or otherwise) the questions were answered. For example, you may wish to comment on common errors and misconceptions, or areas where students have done well. These comments should note any errors in and corrections to the paper. These comments will be made available to students via the MathsCentral Blackboard site and should not contain any information which identifies individual candidates. Any comments which should be kept confidential should be included as confidential comments for the Exam Board and Externals. If you would like to add formulas, please include a sperate pdf file with your email.

ExamModuleCode	Question	Comments for Students	
MATH97085	1	This question was generally well done. The only significant comment is that many candidates failed to distinguish between methods designed to test for uniformity of output and those designed to test for serial dependence.	
MATH97085	2	This question was found straightforward. Rather few answers provided a careful argument about why we would want independent samples, and too many candidates asserted (incorrectly) that inversion was not an option with the density in (d).	
MATH97085	3	Question was generally done well. There were some mistakes in the inversion technique for the trigonometric pdf and the calculation of the full details of the hit-or-miss algorithm was not given in all cases	
MATH97085	4	This questions proved challenging for many students, particularly part (c) where many students were unable to derive the values for b and c.	