ExamModuleCode	Question Number	Comments for Students
M3S9	1	This question was well done and found straightforward. It was, however, quite long, and few candidates had the stamina to verify the result in the final part by transformation of variables and marginalization.
M3S9	2	Quite an intimidating, though fair, question. I had the impression on marking that many candidates panicked a bit at the level of detail required in part (b), and spent a long time calculating the values of M in part (a).
M3S9	3	On the whole, this was quite well done, but care was needed to verify that the algorithm works (if we choose a suitable Q) in the last part.
M45S9	4	This was found a very intimidating question, and few candidates were able to make much progress.

Imperial College
London

M3S9

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)

May-June 2019

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

Stochastic Simulation

Date: Tuesday 28 May 2019

Time: 10.00 - 11.30

Time Allowed: 1 Hour 30 Minutes

This paper has 3 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.
- Calculators may not be used.

Imperial College
London

M4/5S9

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)

May-June 2019

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

Stochastic Simulation

Date: Tuesday 28 May 2019
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.

- 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.
- Calculators may not be used.

1. (a) Consider a probability density function $f(x)$, for $x \in \mathbb{R}$, and a region

$$C = \left((u,v)|0 \le u \le \sqrt{f\left(\frac{v}{u}\right)}\right).$$

(i) Show that if a point (U,V) is uniformly distributed over C, then $\frac{V}{U}$ has density $f(x)$.

(ii) Using the result of Q1(a)(i), outline a ratio-of-uniforms algorithm to generate samples with density $f(x)$ stating any conditions on $f(x)$ needed for the algorithm to work correctly.

(iii) Use the algorithm of Q1(a)(ii) to generate random variates from the Cauchy distribution, for which $f(x) = \frac{1}{\pi(1+x^2)}$.

(iv) Based on the algorithm of Q1(a)(iii), show that if (U,V) is uniformly distributed in the unit circle, then $\frac{V}{U}$ follows the Cauchy distribution.

(v) Is the algorithm of Q1(a)(ii) valid for $f(x) \propto \frac{1}{(1+2x^2)^{3/4}}$? Justify why.

(b) Let $X_1, X_2, ..., X_n$ denote a sample of n iid. random variates. Let $f(x)$, for $x \in \mathbb{R}$, denote a probability density function of a random variable with mean μ and variance σ^2.

(i) Using large sample techniques, present a statistical test to check whether the sample has come from a distribution with first moment μ.

(ii) Present a Chi-squared goodness of fit test to assess if the sample follows the density $f(x)$.

(c) Assume $X \sim$ Normal$(0,1)$ and $Y \sim$ Gamma(α, β), where $\alpha = a/2$ and $\beta = \frac{1}{2}$. Further assume X and Y are independent and $a \in \{1, 2, 3, \dots\}$.

(i) Show that $Z = X/\sqrt{\frac{Y}{a}}$ has the density $c\frac{1}{\left(1+\frac{Z^2}{a}\right)^{\frac{a+1}{2}}}$, where $c = \frac{\Gamma(\frac{a+1}{2})}{\Gamma(\frac{a}{2})\sqrt{\pi a}}$.

(ii) Based on the result of Q1(c)(i) and a set of $n = a+1$ iid. Normal$(0,1)$-distributed random variates, $X_1, X_2, \dots, X_n$, outline an algorithm to generate Z. Show that the algorithm is correct.

Recall that the probability density functions for the normal and Gamma distributions are $Normal(x|0,1) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ *and* $Gamma(y|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}y^{\alpha-1}e^{-\beta y}$, *respectively.*

2. (a) Consider the following rejection sampling algorithm to generate a random variate that has density $f(x)$, for $x \in \mathbb{R}$. Here, $g(x)$ denotes the proposal density.

   1. Compute $M = \sup_x \frac{f(x)}{g(x)}$.
   2. Generate $y \sim g(y)$ and $u \sim U(0,1)$.
   3. If $uMg(y) \leq f(y)$, return y. Else go to step (2).

   (i) Consider $f(x) = \frac{2\Gamma(2\alpha)}{\Gamma(\alpha)^2} x^{\alpha-1}(1-x)^{\alpha-1}$, for $x \in (1/2, 1)$, and zero otherwise, assuming $0 < \alpha < 1$. Consider two distributions $g_1(x) \propto x^{\alpha-1}$ and $g_2(x) \propto (1-x)^{\alpha-1}$ as the proposal distributions for the rejection algorithm. Which distribution of the two would you choose to maximise the acceptance probability of the method? Justify your reasoning, showing the derivations.

   (ii) Consider $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ and $g(x) = \frac{\theta}{\pi(\theta^2 + x^2)}$, for $\theta \in \mathbb{R}$. What is the value for θ that maximises the acceptance probability of the method? Show the derivation of the result.

   (iii) Assume two sequences of functions $l_n(x)$ and $u_n(x)$ such that $\lim_{n\to\infty} l_n(x) = f(x)$ and $\lim_{n\to\infty} u_n(x) = f(x)$ and $l_n(x) \leq f(x) \leq u_n(x)$. Modify the rejection algorithm using l_n and u_n such that f is never computed in step (3). Outline the algorithm for the modified method.

   (iv) Using the series expansion $f(x) = f(0) + \sum_{i=1}^{\infty} \frac{x^i}{i!} f^{(i)}(0)$, where $f^{(i)}(0)$ denotes the ith derivative of the function $f(x)$ evaluated at $x = 0$, outline an extension of the algorithm of Q2(a)(iii) to generate random variates from the density $f(x) = \frac{e^{-x}}{1-e^{-\mu}}$, for $0 \leq x \leq \mu$, and zero elsewhere, for $0 < \mu \leq 1$. Assume $g(x)$ is $U(0,\mu)$.

(b) Let $E_{(1)} \leq E_{(2)} \leq \cdots \leq E_{(n)}$ denote the order statistics of a sequence of iid. standard exponential random variables, $E_1, E_2, \ldots, E_n$. Define $E_{(0)} = 0$.

   (i) Show that $(n - i + 1)\left(E_{(i)} - E_{(i-1)}\right)$, for $i = 1, \ldots, n$, are iid. standard exponential random variables.

   (ii) Show that $\sum_{j=1}^{i} \frac{E_j}{n-j+1}$ are distributed as $E_{(i)}$, for $i = 1, \ldots, n$.

   (iii) Show that e^{-E_i} and $U_i \sim U(0,1)$ are identically distributed.

   (iv) Using the results of Q2(b)(i)-(iii), show that $\prod_{j=i}^{n} U_{n-j+1}^{1/(n-j+1)}$, for $i = 1, \ldots, n$, are distributed as order statistics of n iid $U(0,1)$ variates, $0 \leq U_{(1)} \leq U_{(2)} \leq \cdots \leq U_{(n)}$. Define $U_{(n+1)} = 1$.

   (v) Based on the result of Q2(b)(iv), outline an algorithm to generate an ordered sample of n uniform variates.

   Note that the probability density function of a standard exponential random variable is $f(x) = e^{-x}$, for $x \geq 0$.

3. Consider a discrete-time homogeneous Markov chain with a finite state space, $\{X_0, X_1, X_2, \dots\}$, where X_i, for $i \in \mathbb{N}$, can take $1 < K < \infty$ different values.

(a) Define the properties of the chain that characterise the distribution of the X_i.

(b) Outline an algorithm to simulate $\{X_0, X_1, X_2, \dots\}$, given $X_0 = i_0$, using

(i) the inversion sampling method and

(ii) the rejection sampling method.

(c) Prove that the method Q3(b)(i) works correctly.

(d) At $1 < n < \infty$, what is the distribution of X_n given $X_0 = j$ and the probabilities $P(X_0 = j)$ for $j = 1, \dots, K$? Show the derivation of the result.

(e) What is the definition of

(i) a stationary distribution of $\{X_0, X_1, X_2, \dots\}$?

(ii) an irreducible $\{X_0, X_1, X_2, \dots\}$?

(iii) an aperiodic $\{X_0, X_1, X_2, \dots\}$?

(iv) a reversible $\{X_0, X_1, X_2, \dots\}$?

(f) Outline a convergent Metropolis algorithm to generate random variates from a discrete probability distribution defined over K states. Using the definitions of Q3(e), show that the algorithm works correctly.

4. Let $\mathbf{p} \in \Delta^I$ denote a vector for which $\sum_{i=1}^I p_i = 1$ and $p_i \geq 0$, for $i = 1, \ldots, I$. Consider $\boldsymbol{\psi}_k \in \Delta^V$, for $k = 1, \ldots, K$, and $\boldsymbol{\theta}_m \in \Delta^K$, for $m = 1, \ldots, M$. Let $z_d^{(m)}$ and $w_d^{(m)}$, for $m = 1, \ldots, M$ and $d = 1, \ldots, D_m$, denote random variables generated from categorical distributions; $z_d^{(m)} \sim$ Categorical$(\boldsymbol{\theta}_m)$ and $w_d^{(m)} \sim$ Categorical$(\boldsymbol{\psi}_{z_d^{(m)}})$. Further, assume $\boldsymbol{\psi}_k \sim$ Dirichlet$(\gamma \mathbf{1}_V)$ and $\boldsymbol{\theta}_m \sim$ Dirichlet$(\alpha \mathbf{1}_K)$, for $k = 1, \ldots, K$ and $m = 1, \ldots, M$.

Denote $\Theta = \{\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_K, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M, z_1^{(1)}, \ldots, z_{D_M}^{(M)}\}$, $\mathcal{D} = \{w_1^{(1)}, \ldots, w_{D_M}^{(M)}\}$ and $\Psi = \{\alpha, \gamma\}$.

(a) Showing the derivations, what is the distribution of

(i) $\boldsymbol{\theta}_m$, conditioning on $\{\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_K, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{m-1}, \boldsymbol{\theta}_{m+1}, \ldots, \boldsymbol{\theta}_M, z_1^{(1)}, \ldots, z_{D_M}^{(M)}\}$, $\mathcal{D}$ and Ψ?

(ii) $\boldsymbol{\psi}_k$, conditioning on $\{\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{k-1}, \boldsymbol{\psi}_{k+1}, \ldots, \boldsymbol{\psi}_K, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M, z_1^{(1)}, \ldots, z_{D_M}^{(M)}\}$, $\mathcal{D}$ and Ψ?

(iii) $z_d^{(m)}$, conditioning on $\{\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_K, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M, z_1^{(1)}, \ldots, z_{d-1}^{(m)}, z_{d+1}^{(m)}, \ldots, z_{D_M}^{(M)}\}$, $\mathcal{D}$ and Ψ?

(b) Let Y_i, for $i = 1, \ldots, I$, denote iid. Gamma variables with density Gamma$(\beta_i, 1)$, respectively. See Q1(c) for the expression of the density. Show that if $Y = \sum_{i=1}^I Y_i$ and $X_i = \frac{Y_i}{Y}$ then $(X_1, \ldots, X_I) \sim$ Dirichlet$(\boldsymbol{\beta})$.

(c) Using the inversion sampling technique, outline an algorithm to simulate from the distribution of Q4(a)(iii).

(d) Based on the results of Q4(a)-(c), outline a Gibbs sampling algorithm to generate random variates from $p(\Theta|\mathcal{D}, \Psi)$. You may assume access to a random variate generator for the Gamma density.

(e) Denoting $\mathbf{z} = \{z_1^{(1)}, \ldots, z_{D_M}^{(M)}\}$, what is the distribution $p(\mathcal{D}, \mathbf{z}|\Psi)$? Show the derivation of the result.

(f) Denote $\mathbf{z}_{-md}$ such that the $z_d^{(m)}$ is excluded from the $\mathbf{z}$. Using the result of Q4.(e), what is the conditional distribution $p(z_d^{(m)}|\mathbf{z}_{-md}, \mathcal{D}, \Psi)$? Show the derivation of the result.

(g) Using inversion sampling technique and the result of Q4.(f), outline a collapsed Gibbs sampling algorithm to generate random variates from $p(\mathbf{z}|\mathcal{D}, \Psi)$.

Recall: The probability mass function for $x \sim$ Categorical$(\mathbf{p})$ is $f(x = i) = p_i$, for $i = 1, \ldots, I$. Also, the probability density function for $\mathbf{p} \sim$ Dirichlet$(\boldsymbol{\beta})$ is $\frac{1}{Z}\prod_{i=1}^I p_i^{\beta_i - 1}$, for $\beta_i > 0$, for $i = 1, \ldots, I$, and $Z = \frac{\prod_{i=1}^I \Gamma(\beta_i)}{\Gamma(\sum_{i=1}^I \beta_i)}$, where $\Gamma(\cdot)$ denotes the Gamma function and $\mathbf{1}_I$ denotes I-dimensional vector of ones.

1. (a) (i) **(A bookwork 3 marks)** Define $\left(u, x = \frac{v}{u}\right)$. The Jacobian of the inverse transformation is u and the density of (u, v) is uniform over the region C and zero otherwise. The uniform density is 1/Area(C). Here, Area(C) $= \int\int_C dudv = \int\int_0^{\sqrt{f(x)}} ududx = \int 1/2f(x)dx$, using $dv = udx$ and $0 \leq u \leq \sqrt{f(x)}$. The density of (u, x) is u times the density of (u, v). Finally, the density of x is the marginal density $\int_0^{\sqrt{f(x)}} u/\text{Area}(C)du = f(x)$.

(ii) **(A bookwork 2 marks)** Enclose the area C by a rectangle B. Generate samples over the rectangle and apply the rejection principle. The boundary of B is $[0, \sup\sqrt{f(x)}]$ and $[\inf x\sqrt{f(x)}, \sup x\sqrt{f(x)}]$.

1. Compute the boundary of B.
2. Generate $U \sim U(0, \sup\sqrt{f(x)})$ and $V \sim U(\inf x\sqrt{f(x)}, \sup x\sqrt{f(x)})$ and compute $X = \frac{V}{U}$.
3. If $U^2 \leq f(X)$, return X. Else go to (2).

The algorithm is valid if $f(x)$ and $x^2 f(x)$ are bounded in the domain of x.

(iii) **(B sim seen 2 marks)**

1. Generate $U \sim U(0, 1/\sqrt{\pi})$, $V \sim U(-1/\sqrt{\pi}, 1/\sqrt{\pi})$, $X = \frac{V}{U}$.
2. If $U^2 \leq \frac{1}{\pi(1+X^2)}$, return X. Else go to (1).

(iv) **(D unseen 2 marks)** Rewrite the acceptance step (2) of the algorithm, $U^2 + V^2 \leq 1/\pi$ and modify the ranges to generate the U and V from $U(-1, 1)$.

(v) **(D unseen 3 marks)** No, because the boundaries do not exist for B. We can see this either by searching for the boundaries or noting that $x^2 f(x)$ is unbounded. Thus we are not able to find an enclosing rectangle for C.

(b) **(A bookwork 5 marks)**

(i) Using Central Limit Theorem, $\hat{X} = \frac{1}{n}\sum_{i=1}^n X_i$ has an asymptotic Normal$\left(\mu, \frac{\sigma^2}{n}\right)$ distribution. Design the null hypothesis H_0 with corresponding μ and σ^2 and choose significance level α.

(ii) Define n categories that group the sample variates. Let $O_1, \ldots, O_n$ denote the observed frequencies noting that $\sum_i O_i = n$. Also define the expected frequencies under some H_0 for the distribution of variates by E_i, respectively.

To measure departure of the observed from the expected, use the following statistic $S = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$. If n is large then S follows chi-squared distribution with $n-1$ degrees of freedom. Thus we compare the observed statistic to the chi-squared distribution as a method of testing the hypothesis.

(c) **(B sim seen 3 marks)**

(i) Define a vector $\left(Z = X/\sqrt{\frac{Y}{a}}, W = Y\right)$ whose inverse transformation is $\left(X = Z\sqrt{\frac{W}{a}}, Y = W\right)$. The Jacobian of the transformation is $\sqrt{\frac{W}{a}}$. The density of (Z, W) is $c_0 e^{-\frac{wz^2}{2a}} w^{\frac{a}{2}-1} e^{-\frac{w}{2}} \sqrt{\frac{w}{a}}$, where $c_0 = 1/\left(\Gamma(\frac{a}{2}) 2^{\frac{a}{2}} \sqrt{2\pi}\right)$, for $w > 0, z \in \mathbb{R}$. Marginalise w from the joint density to obtain the result; for fixed z, w has density Gamma $\left(\frac{a+1}{2}, \frac{1+z^2/a}{2}\right)$.

(ii) The algorithm output is $Z = X_n/\sqrt{\frac{\sum_{j=1}^{a} X_j^2}{a}}$. To verify correctness of the algorithm, we need to show that $\sum_{j=1}^{a} X_j^2$ follows Gamma$(a/2, 1/2)$ distribution. We first show that X_i^2 has Gamma$(1/2, 1/2)$ distribution and then using moment generating function we show that the sum of a iid squared $N(0,1)$ variates follows Gamma$(a/2, 1/2)$.

2. (a) (i) **(A bookwork 2 marks)** The acceptance probability of the method depends on the value for M; the smaller the value, the larger the probability of acceptance. For g_2 the value for M leads to division by zero at $x = 1$ and thus for an infinite value for M. Thus g_1 should be used.
For g_1, we have $M = \sup_x \frac{2\Gamma(2\alpha)}{\Gamma(\alpha)^2}(1-x)^{\alpha-1}$, for $x \in (1/2, 1)$. The maximum is obtained for $x = 1/2$.

(ii) **(C sim seen 2 marks)** We minimise the value of $\log M$ wrt. θ. Setting the derivative to zero yields $-x + \frac{2x}{\theta^2 + x^2} = 0$. This gives the values $x = 0$ and $x = \pm\sqrt{2-\theta^2}$. The latter case can only occur for $\theta^2 \leq 2$. At $x = 0$, $M = \theta\sqrt{\frac{\pi}{2}}$. At $x = \pm\sqrt{2-\theta^2}$, $M = \frac{\sqrt{2\pi}}{e\theta} e^{\theta^2/2}$. For $\theta < \sqrt{2}$, the maximum of M is obtained at $x = \pm\sqrt{2-\theta^2}$ and the minimum at $x = 0$. For $\theta \geq \sqrt{2}$, the maximum is obtained at $x = 0$. Thus, the M has only one minimum at $\theta = 1$ and for this choice $M = \sqrt{\frac{2\pi}{e}}$.

(iii) **(D unseen 3 marks)** We modify step (3) of the method.

1. Compute $M = \sup_x \frac{f(x)}{g(x)}$.
2. Generate $y \sim g(y)$ and $u \sim U(0,1)$.
3. Set $n = 0$
4. Increase n by one, while $uMg(y) > u_n(y)$, and if $uMg(y) \leq l_n(y)$, return y. Else go to step (2).

(iv) **(B bookwork 2 marks)** For odd k and $x > 0$, we use $\sum_{i=0}^{k}(-1)^i \frac{x^i}{i!} \leq e^{-x} \leq \sum_{i=0}^{k-1}(-1)^i \frac{x^i}{i!}$. We have $l_n(x) = \frac{1}{1-e^{-\mu}} \sum_{i=0}^{n}(-1)^i \frac{x^i}{i!}$ and $u_n(x) = \frac{1}{1-e^{-\mu}} \sum_{i=0}^{n-1}(-1)^i \frac{x^i}{i!}$, for odd n.

(b) (i) **(D unseen 3 marks)** The joint density for $E_{(1)}, \ldots, E_{(n)}$ is $n! e^{-\sum_{i=1}^n x_i}$, for $0 \le x_1 \le x_2 \le \cdots \le x_n$. We note that $n! e^{-\sum_{i=1}^n x_i} = n! e^{-\sum_{i=1}^n (n-i+1)(x_i - x_{i-1})}$. Define $Y_i = (n-i+1)(E_{(i)} - E_{(i-1)})$ and $y_i = (n-i+1)(x_i - x_{i-1})$. We have $x_1 = \frac{y_1}{n}$, $x_2 = \frac{y_1}{n} + \frac{y_2}{n-1}, \ldots, x_n = \frac{y_1}{n} + \cdots + \frac{y_n}{1}$. The Jacobian of the transformation is $\frac{1}{n!}$. Thus $Y_1, \ldots, Y_n$ has density $e^{-\sum_{i=1}^n y_n}$, for $y_i \ge 0$.

(ii) **(C unseen 2 marks)** Define $E_{(1)} = \frac{E_1}{n}$, $E_{(2)} = E_{(1)} + \frac{E_2}{n-1}, \ldots, E_{(n)} = E_{(n-1)} + \frac{E_n}{1}$. The result follows from the result of Q(1)(b)(i).

(iii) **(A bookwork 2 marks)** The distribution function for $x \sim \text{Exp}(1)$ is $F(x) = 1 - e^{-x}$. Thus $P(e^{-X} \le x) = P(X \ge -\log(x)) = 1 - F(-\log(x)) = x$. Thus e^{-E_i} has uniform distribution.

(iv) **(B sim seen 2 marks)** In the Q2.(b)(i)-(ii), replace U_i by e^{-E_i} and $U_{(i)}$ by $e^{-E_{(i)}}$ obtaining the result.

(v) **(A 2 marks)** Generate $U_1, \ldots, U_n$ independent $U(0,1)$'s. Then iteratively define $U_{(n)} = U_n^{\frac{1}{n}}$ and $U_{(k)} = U_{(k+1)} U_k^{\frac{1}{k}}$, for $k = n-1, \ldots, 1$. Then $U_{(1)}, \ldots, U_{(n)}$ are ordered uniform variates.

3. (a) **(A bookwork 2 marks)** Together with the Markov property, given by, $p(X_n = j | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) = P(X_n = j | X_{n-1} = i_{n-i})$ and time homogeneity, such that, $P(X_{n+1} = j | X_n = i)$ equals for all n, we need to specify distribution for the initial state X_0 and the transition probabilities between states.

(b) (i) **(A bookwork 2 marks)** Define $p_{ij} = P(X_{n+1} = j | X_n = i)$ and $P_{ik} = \sum_{j=1}^k p_{ij}$.

1. Set $n = 1$.
2. Set $i = X_{n-1}$, $k = 1$, and generate $U = u \sim U(0,1)$.
3. If $u \le P_{ik}$, set $X_n = k$.
4. Otherwise $k \leftarrow k+1$, go to step (3).
5. Set $n \leftarrow n+1$ and go to step (2).

(ii) **(B sim seen 2 marks)** We need proposal probabilities q_{ij} such that $p_{ij} \le M_i q_{ij}$, where $M_i \ge 1$ is the rejection constant for state i.

1. Set $n = 1$.
2. Set $i = X_{n-1}$.
3. Generate $U = u \sim U(0,1)$ and a proposal k with distribution determined by q_{ij}, for $j = 1, \ldots, K$ (using for example inversion).
4. If $u M_i q_{ik} \le p_{ik}$, set $X_n = k$. Otherwise, go to step (3).
5. Set $n \leftarrow n+1$ and go to step (2).

(c) **(B sim seen 2 marks)** Define the generalised inverse, $F^{-1}(u) = \min\{x, F(x) \geq u\}$. For X with distribution function F, we have $F_U(u) = P(U \leq u) = P(F(X) \leq u) = P(X \leq F^{-1}(u)) = u$. Also $F(x) = P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$. Thus $F^{-1}(U)$ has distribution function F. Here, given i, define F by the probabilities P_{ij}.

(d) **(A bookwork 2 marks)** Define n-step transition probabilities $p_{ij}^{(n)} = P(X_n = j | X_0 = i)$ and let P denote a matrix with elements p_{ij}, similarly, $P^{(n)} = \left(p_{ij}^{(n)}\right)$.

$$
\begin{aligned}
p_{ij}^{(n)} &= P(X_n = j \,|\, X_0 = i) \\
&= \sum_k P(X_n = j, X_1 = k \,|\, X_0 = 1) \\
&= \sum_k P(X_n = j \,|\, X_1 = k) P(X_1 = k \,|\, X_0 = i) \\
&= \sum_k p_{ik} p_{kj}^{(n-1)} \\
\Rightarrow P^{(n)} &= P^n \quad (n\text{th power of P})
\end{aligned}
$$

Let $\pi_i^{(0)} = P(X_0 = i)$ and use $\underline{\pi}^{(0)}$ to denote the row vector, i.e. the initial distribution over the discrete state space.

$$
\begin{aligned}
\pi_i^{(n)} = P(X_n = i) &= \sum_k P(X_n = i, X_{n-1} = k) \\
&= \sum_k P(X_n = i \,|\, X_{n-1} = k) P(X_{n-1} = k) \\
&= \sum_k \pi_k^{(n-1)} P_{ki} \\
\Rightarrow \underline{\pi}^{(n)} &= \underline{\pi}^{(n-1)} P \\
\Rightarrow \underline{\pi}^{(n)} &= \underline{\pi}^{(0)} P^n
\end{aligned}
$$

(e) **(B bookwork 5 marks)**

(i) π is a stationary distribution iff (1) $\pi_i \geq 0, \forall i$, (2) $\sum_i \pi_i = 1$ and (3) $\pi_j = \sum_i \pi_i p_{ij}, \forall j$ $(\pi = \pi P)$

(ii) If it is possible to traverse from any state i to any state j and traverse back from j to i, the chain is irreducible.

(iii) Define a period of a state as the greatest common divisor $\{n : p_{ii}^{(n)} > 0\}$. A state is aperiodic if its period is 1 and the Markov chain is aperiodic if all its states in the corresponding state space are aperiodic.

(iv) The detailed balance condition is $\pi_j p_{ji} = \pi_i p_{ij}$. A chain that satisfies this property is reversible.

(f) **(C sim seen 5 marks)** We choose any symmetric transition matrix Q, with elements q_{ij}. Then, we define our transition mechanism as follows:

Suppose the chain is in state i...

* we select state j as a candidate for the next state of the chain, with probability q_{ij},
* and we then move to state j with probability

$$\min\left\{1, \frac{\pi_j}{\pi_i}\right\},$$

otherwise, we stay at state i. Here π denotes the probability mass function of the discrete probability distribution.

This defines P in the following way:

$$\begin{aligned}
p_{ij} &= \min\left\{1, \frac{\pi_j}{\pi_i}\right\} q_{ij}, \qquad i \neq j \\
p_{ii} &= \underbrace{q_{ii}}_{\text{prob. choose state } i} + \sum_{j \neq i} \underbrace{\max\left\{0, 1 - \frac{\pi_j}{\pi_i}\right\} q_{ij}}_{\text{prob. not moving to state } j}
\end{aligned}$$

To guarantee convergence, we need to show that P is irreducible, aperiodic and that $\underline{\pi} = \underline{\pi} P$. For the P designed above, we have

$$\begin{aligned}
\pi_i p_{ij} &= \pi_i \min\left\{1, \frac{\pi_j}{\pi_i}\right\} q_{ij} \\
&= \min\{\pi_i, \pi_j\} q_{ij} \\
&= \min\{\pi_i, \pi_j\} q_{ji} \quad \text{since } Q \text{ symmetric} \\
&= \pi_j \min\left\{1, \frac{\pi_i}{\pi_j}\right\} q_{ji} \\
&= \pi_j p_{ji}
\end{aligned}$$

In addition we choose a Q that is irreducible and aperiodic.

4. **(Mastery Question 20 marks)**

(a) (i) Collecting terms that depend on $\boldsymbol{\theta}_m$, we identify $\boldsymbol{\theta}_m|\text{rest} \sim \text{Dirichlet}(n_{mk} + \alpha)$, where n_{mk} is the number of times $z_d^{(m)}$ takes value k for some m for $d = 1, \ldots, D_m$.

(ii) Collecting terms that depend on ψ_k, we identify $\psi_k|\text{rest} \sim \text{Dirichlet}(g_{kv} + \gamma)$, where g_{kv} is the number of times $z_d^{(m)}$ takes value k for $w_d^{(m)} = v$ over $m = 1, \ldots, M$ and $d = 1, \ldots, D_m$.

(iii) This is a discrete distribution with proportional probabilities $p(z_d^{(m)} = k|w_d^{(m)} = v, \text{rest}) \propto \theta_{mk}\psi_{kv}$.

(b) The joint density of Y_i's is $c_0 \prod_i y_i^{\beta_i - 1} e^{-\sum_i y_i}$, where c_0 is a normalisation constant. Consider $y = \sum_i y_i$ and $x_i = y_i/y$, with reverse transformation $y_i = yx_i$. The Jacobian of the transformation is y^I and the density of x_i's is Dirichlet(β), marginalising over y.

(c)
1. Compute proportional probabilities $p(z_d^{(m)} = k)$ for $k = 1, \ldots, K$ and their sum $\sum_i p(z_d^{(m)} = i)$. Define $p_k = p(z_d^{(m)} = k)/\sum_i p(z_d^{(m)} = i)$.
2. Generate $u \sim U(0, 1)$ and set $j = 1$.
3. If $u \leq \sum_{j'=1}^{j} p_{j'}$, return j.
4. Otherwise $j \leftarrow j + 1$ and go to step (3).

(d)
1. sample $\boldsymbol{\theta}_m$, for $m = 1, \ldots, M$, from the distribution given in Q4(a)(i) using the result of Q4(a)(b) by generating independent Gamma variates and normalising them.
2. sample ψ_k, for $k = 1, \ldots, K$, using similar technique as in step (1).
3. sample $z_d^{(m)}$, for $m = 1, \ldots, M$ $d = 1, \ldots, D_m$, using the method given in Q4(c).
4. Go to step (1).

After initialising the values for the Θ, the method converges to the posterior distribution and generates random variates from the $p(\Theta|\mathcal{D}, \Psi)$.

(e) We marginalise $\boldsymbol{\theta}_m$ and ψ_k, for $m = 1, \ldots, M$ and $k = 1, \ldots, K$, from the joint density $p(\Theta, \mathcal{D}|\Psi)$. Marginalisation of $\boldsymbol{\theta}_m$ results in $\frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} \frac{\Gamma(n_{mk}+\alpha)}{\Gamma(\sum_{k'}(n_{mk'}+\alpha))}$, using the formula for computing the normalisation constant of the Dirichlet distribution. We use a similar method for marginalising ψ_k. We obtain the $p(\mathcal{D}, \mathbf{z}|\Psi)$ by combining the results using conditional independence.

(f) We know the form of the distribution is a categorical distribution. The probabilities are proportional to $p(z_d^{(m)} = k|\mathbf{z}_{-md}, \mathcal{D}, \Psi) \propto p(z_d^{(m)} = k, \mathbf{z}_{-md}, \mathcal{D}, \Psi)$. We may further simplify the expression using the properties of the Gamma function.

(g) Using the result of Q4(f) we have $p(z_d^{(m)} = k|w_d^{(m)} = v) \propto (n_{mk}^{-w_d^{(m)}} + \alpha) \frac{g_{kv}^{-w_d^{(m)}} + \gamma}{\sum_{v'} g_{kv'}^{-w_d^{(m)}} + \gamma}$, where the upper index denotes omitting the contribution of $w_d^{(m)}$. The sampler proceeds by sampling the $z_d^{(m)}$ using the marginal conditional probability.