Imperial College London

MATH70047

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May 2023

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

Stochastic Simulation

Date: 11 May 2023

Time: 10:00 – 12:00 (BST)

Time Allowed: 2 hrs

This paper has 4 Questions.

Please Answer Each Question in a Separate Answer Booklets

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

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

Any required additional material(s) will be provided.

Allow margins for marking.

Credit will be given for all questions attempted.

Each question carries equal weight.

DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO

1. (a) Consider the Rayleigh probability density function (PDF)

$$p(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right), \qquad x \in [0, \infty).$$

with the parameter $\sigma > 0$.

(i) Show that (using integration by parts or other techniques) the cumulative distribution function (CDF) of $p(x)$, denoted $F_X(x)$ is given by

$$F_X(x) = 1 - \exp\left(-\frac{x^2}{2\sigma^2}\right) \qquad \text{for } x \in [0, \infty).$$

Hint: Completing the square will **not** work here, as the support is positive. (3 marks)

(ii) Given $F_X(x)$ above, show that

$$F_X^{-1}(u) = \sqrt{-2\sigma^2 \log(1-u)}.$$

(2 marks)

(iii) Based on part (ii), write the pseudo-code of the inversion method that samples from the Rayleigh distribution $p(x)$ using a uniform sample $U \sim \text{Unif}(0, 1)$. (2 marks)

(iv) Justify the following sampler for the Rayleigh variable

$$U \sim \text{Unif}(0, 1) \qquad\qquad X = \sqrt{-2\sigma^2 \log(U)}.$$

Explain why this also gives $X \sim p(x)$ as above. (1 mark)

(b) Now, suppose we would like to implement a rejection sampling (RS) scheme with the Rayleigh PDF $p(x)$ as the target density and using the proposal density

$$q_\mu(x) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad \text{for } x \in (-\infty, \infty).$$

Note that the Rayleigh density parameter and the standard deviation of $q_\mu(x)$ are fixed and **equal** (σ).

(i) Show that the value $x^\star$ that maximises the ratio $p(x)/q_\mu(x)$ is given by $x^\star = \sigma^2/\mu$. Provide any necessary constraints on μ. (4 marks)

(ii) Based on (i), show that

$$M_\mu := \sup_{x \in [0,\infty)} \frac{p(x)}{q_\mu(x)} = \frac{\sigma\sqrt{2\pi}}{\mu} \exp\left(\frac{\mu^2}{2\sigma^2} - 1\right).$$

(2 marks)

(iii) Provide the expression of the **acceptance rate** of the RS and show that the optimal μ that maximises the acceptance rate is $\mu_\star = \sigma$ and that $M_{\mu_\star} = \sqrt{2\pi/e}$. (3 marks)

(iv) Provide the pseudo-code of the RS method based on the optimal proposal $q_{\mu_\star}$ where the acceptance ratio is provided explicitly. (3 marks)

(Total: 20 marks)

2. Consider a generic Bayesian probability model with a prior distribution $p(x)$ and a likelihood $p(y|x)$. The marginal likelihood of this model is given as

$$p(y) = \int p(y|x)p(x)\mathrm{d}x.$$

Assume we fix an observation $y = y_0$ and define $p(y_0) = \int p(y_0|x)p(x)\mathrm{d}x$.

(a) (i) Describe the standard Monte Carlo (MC) method for estimating $p(y_0)$. Describe the test function $\varphi(x)$ clearly. (2 marks)

(ii) Prove that the MC estimator in (i) is unbiased. (2 marks)

(iii) Write down an expression for the variance of the standard MC estimator in terms of φ. Simplify this expression as much as possible. (2 marks)

(iv) Describe a scenario where this MC estimator could fail to be accurate. (1 mark)

(b) In this part, we are interested in importance sampling (IS) for this problem.

(i) Describe the IS estimator for estimating $p(y_0)$ using a proposal $q(x)$. (2 marks)

(ii) Prove that the IS estimator is unbiased. (2 marks)

(iii) Derive an expression for the variance of the IS estimator. Simplify this expression as much as possible. (3 marks)

(iv) State a condition, in terms of the functions $p(x)$, $p(y_0|x)$, and $q(x)$, that must be satisfied, in order for this variance to be finite.

(1 mark)

(c) Consider the problem of estimating $p(y_0)$ given

$$\begin{aligned} p(x) &= \mathcal{N}(x; 0, 1), \\ p(y|x) &= \mathcal{N}(y; x, 1). \end{aligned}$$

Consider the following proposal for IS

$$q_\mu(x) = \mathcal{N}\left(x; \mu, \frac{1}{2}\right).$$

Prove that the optimal value $\mu^\star$ that minimises the variance of the IS estimator is given by

$$\mu^\star = \frac{y_0}{2}.$$

Discuss how this optimal importance sampler can be useful if y_0 is unlikely under $p(y)$.

(5 marks)

(Total: 20 marks)

3. This question is about Markov chain Monte Carlo (MCMC) on a continuous state-space.

(a) (i) Given a Markov kernel with corresponding density $K(x'|x)$ and a target density $p(x)$, state the detailed-balance condition. (2 marks)

(ii) Given a three-dimensional target $p(x, y, z)$, describe one step of the Gibbs sampler for this target given the state of the chain $(x_{n-1}, y_{n-1}, z_{n-1})$ by writing the pseudo-code. (2 marks)

(iii) Suppose that the full conditionals for this target density are not available in closed form, but unnormalised full conditionals are available. Suggest a suitable modification that can be made to the Gibbs sampler, and provide the pseudo-code for one full update of your modified sampler. You do not need to specify the form of any required proposal distributions or acceptance probabilities. (2 marks)

(iv) In order to measure the quality of the MCMC algorithm, we can evaluate the autocorrelation of the resulting chain. State one way in which we can reduce the autocorrelation in the final MCMC sample. Give one other statistic that is commonly used to measure the quality of the MCMC sample. (2 marks)

(b) Consider the Metropolis-Hastings (MH) sampler for a generic target density $p(x)$.

(i) Provide the pseudocode of the MH sampler for target density $p(x)$ given an independent proposal density $q(x)$. (2 marks)

(ii) Provide the pseudocode of the MH sampler for target density $p(x)$ given a symmetric proposal density $q(x|x')$. (2 marks)

(iii) Consider the following target density

$$p(x) = \mathcal{N}(x; \mu, \sigma^2).$$

Derive the proposal density $q(x'|x)$ of the Metropolis Adjusted Langevin Algorithm (MALA) for this target density and write the pseudo-code. (2 marks)

(iv) If we skip the Metropolis acceptance step in MALA and just use the proposal as a sampler itself, what would happen to the target distribution of the chain? Would it be still $p(x)$? Answer in words, without equations. (1 mark)

(c) Consider the following Gaussian distribution $p(x|\sigma^2)$ with an inverse Gamma (IG) prior $p(\sigma^2)$ on its variance:

$$p(\sigma^2) = \mathcal{IG}(\sigma^2; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left(-\frac{\beta}{\sigma^2}\right),$$

$$p(x|\sigma^2) = \mathcal{N}(x; 0, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

For this model,

(i) State full conditionals by deriving their (closed-form) distributions. (3 marks)

(ii) Describe the pseudo-code of the Gibbs sampler for sampling from the joint distribution $p(x, \sigma^2)$ assuming that you can sample from the IG distribution. (2 marks)

(Total: 20 marks)

4. This question is based on the mastery material, J. Dahlin and T. B. Schön, *Getting started with particle Metropolis-Hastings for inference in nonlinear dynamical models*. Journal of Statistical Software, 2019.

Consider the following generic state-space model structure

$$\begin{aligned} x_0 &\sim \mu(x_0), \\ x_t|x_{t-1} &\sim f(x_t|x_{t-1}), \\ y_t|x_t &\sim g(y_t|x_t). \end{aligned}$$

(a) Based on the state-space model above:

(i) Describe the filtering problem in a state-space model and the quantity of interest in this setting. (2 marks)

(ii) Write the expression of the full-joint distribution $p(y_{1:n}, x_{0:n})$ in terms of μ, f, g and variables $(x_{0:n}, y_{1:n})$. (2 marks)

(iii) Write the *filtering recursions*, that is, state the prediction and update steps in terms of f and g given the filtering distribution at time $n-1$ to obtain the filtering distribution at time n. (3 marks)

(b) (i) Based on the model above, describe the pseudo-code of the sequential importance sampling-resampling (SISR) method with a Markov proposal $q(x_t|x_{t-1})$. Provide a clear expression for the importance weights at each step. (3 marks)

(ii) Consider the case $q(x_t|x_{t-1}) = f(x_t|x_{t-1})$. How does this change the algorithm you provided in (b)-(i)? What is the algorithm called in this case? (2 marks)

(iii) Suggest a method for calculating the normalised weights at each step, that avoids problems of numerical instability. (1 mark)

(c) Consider now the model with parameter θ given below

$$\begin{aligned} \theta &\sim p(\theta), \\ x_0 &\sim \mu_\theta(x_0), \\ x_t|x_{t-1} &\sim f_\theta(x_t|x_{t-1}), \\ y_t|x_t &\sim g_\theta(y_t|x_t), \end{aligned}$$

and the problem of sampling from $p(\theta|y_{1:T})$ given an observation sequence $y_{1:T}$.

(i) Write down the estimator of the marginal likelihood $p_\theta(y_{1:T}) := p(y_{1:T}|\theta)$ given the SISR algorithm with proposal $f(x_t|x_{t-1})$ as described in Part(b)-(ii). (3 marks)

(ii) Write the pseudo-code of the particle Metropolis-Hastings method in order to sample from the posterior $p(\theta|y_{1:T})$. Describe each term of the acceptance ratio clearly. What is the key property of the marginal likelihood estimator given in Part(c)-(i) above so that this MH algorithm is valid? (4 marks)

(Total: 20 marks)

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2023

This paper is also taken for the relevant examination for the Associateship.

MATH60047/70047

Stochastic Simulation (Solutions)

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

1. (a) (i) We write $\boxed{\text{unseen} \Downarrow}$

$$F_X(x) = \int_0^x p(y)\mathrm{d}y = \int_0^x \frac{y}{\sigma^2} \exp\left(-\frac{y^2}{2\sigma^2}\right) \mathrm{d}y.$$

Choose $u = y^2$, hence $\mathrm{d}u = 2y\mathrm{d}y$. Note that this also changes the limits $y = 0 \implies u = 0$ and $y = x \implies u = x^2$. We can then write the integral as

$$F_X(x) = \int_0^{x^2} \frac{1}{2\sigma^2} \exp\left(-\frac{u}{2\sigma^2}\right) \mathrm{d}u = \left[-\exp\left(-\frac{u}{2\sigma^2}\right)\right]_0^{x^2} = 1 - \exp\left(-\frac{x^2}{2\sigma^2}\right),$$

as wanted. $\boxed{3, \text{B}}$

(ii) Write $F_X(x) = u$ and leave u alone as follows $\boxed{\text{meth seen} \Downarrow}$

$$1 - e^{-\frac{x^2}{2\sigma^2}} = u \implies 1 - u = \exp\left(-\frac{x^2}{2\sigma^2}\right) \implies \log(1-u) = -\frac{x^2}{2\sigma^2},$$
$$\implies -2\sigma^2 \log(1-u) = x^2 \implies \sqrt{-2\sigma^2 \log(1-u)} = x$$

We deduce from above that

$$F^{-1}(u) = \sqrt{-2\sigma^2 \log(1-u)}.$$

$\boxed{2, \text{B}}$

(iii) Inversion is a method to sample a uniform and then pass this through F_X^{-1}. An example pseudo code is given as

1. $U \sim \text{Unif}(0,1)$
2. $X = F_X^{-1}(U) = \sqrt{-2\sigma^2 \log(1-U)}$

It is OK to refer to part (ii) for the explicit expression without rewriting it. $\boxed{2, \text{A}}$

(iv) This sampler works because U and $1-U$ have the same distribution. Therefore, one can replace $1-U$ with simply U and the sampler would do the same thing.

$\boxed{1, \text{B}}$

(b) For this question we have the target $\boxed{\text{meth seen} \Downarrow}$

$$p(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \qquad x \in [0, \infty)$$

and the proposal

$$q_\mu(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

(i) We write down the ratio

$$R(x) = \frac{p(x)}{q_\mu(x)} = \frac{\frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right)}{\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)} = \frac{x\sqrt{2\pi}\cancel{\exp\left(-\frac{x^2}{2\sigma^2}\right)}}{\sigma\cancel{\exp\left(-\frac{x^2}{2\sigma^2}\right)}\exp\left(\frac{x\mu}{\sigma^2}\right)\exp\left(-\frac{\mu^2}{2\sigma^2}\right)}$$
$$= \frac{\sqrt{2\pi}}{\sigma} x \exp\left(-\frac{x\mu}{\sigma^2}\right) \exp\left(\frac{\mu^2}{2\sigma^2}\right).$$

In order to maximise this, we take the log and its derivative

$$\frac{\mathrm{d}\log R(x)}{\mathrm{d}x} = \frac{1}{x} - \frac{\mu}{\sigma^2},$$

and setting this to zero, we obtain

$$x^\star = \frac{\sigma^2}{\mu}.$$

We know that $x \geq 0$ for the support of $p(x)$, hence we impose $x^\star \geq 0$. This implies that we should impose $\mu > 0$ when choosing the proposal. $\boxed{4, \mathrm{D}}$

(ii) For this, we simply compute

$$R(x^\star) = \sup_{x\in[0,\infty)} \frac{p(x)}{q_\mu(x)} = \frac{\sqrt{2\pi}\sigma}{\mu} \exp\left(\frac{\mu^2}{2\sigma^2} - 1\right),$$

as asked. We also denote this as $M_\mu = R(x^\star)$ for the following answer. $\boxed{2, \mathrm{C}}$

(iii) The acceptance rate is given as $1/M_\mu$ as provided in the course. In this case, this gives us

$$\hat{a} = \frac{1}{M_\mu} = \frac{\mu}{\sigma\sqrt{2\pi}} \exp\left(1 - \frac{\mu^2}{2\sigma^2}\right).$$

In order to maximise $\hat{a}$, we need to minimise M_μ, similarly by taking its derivative and setting it to zero. For this, we take again the logarithm

$$\log M_\mu = \log \sqrt{2\pi}\sigma - \log\mu + \frac{\mu^2}{2\sigma^2} - 1.$$

Taking its derivative and setting it to zero

$$\frac{\mathrm{d}\log M_\mu}{\mathrm{d}\mu} = -\frac{1}{\mu} + \frac{\mu}{\sigma^2} = 0$$

implies that

$$\mu_\star = \sigma.$$

since we constrained $\mu > 0$ in part(b)(i) (excluding the negative solution). Plugging this above, we obtain

$$M_{\mu_\star} = \sqrt{2\pi/e}$$

$\boxed{3, \mathrm{D}}$

(iv) Given all components we can write the optimal rejection sampler (RS).

1. Sample

$$X' \sim q_{\mu_\star}(x)$$

2. Compute the acceptance ratio

$$R(X') = \frac{p(x)}{q_{\mu_\star}(x)} = \frac{\sqrt{2\pi}}{\sigma} X' e^{-\frac{X'}{\sigma}} e^{\frac{1}{2}}.$$

3. Sample $U \sim \mathrm{Unif}(0,1)$ and accept if

$$U \leq \frac{1}{M_{\mu_\star}} R(X') = \frac{1}{\sigma} X' e^{-\frac{X'}{\sigma}+1}.$$

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

2. (a) (i) Given the integral $\boxed{\text{sim. seen} \Downarrow}$

$$p(y_0) = \int p(y_0|x)p(x)\mathrm{d}x,$$

the MC estimator is given using the samples $X_1, \ldots, X_N \sim p(x)$ and setting $\varphi(x) = p(y_0|x)$, hence

$$p_{\mathsf{MC}}^N(y_0) = \frac{1}{N}\sum_{i=1}^N p(y_0|X_i).$$

$\boxed{2, \text{A}}$

(ii) This estimator is unbiased, i.e.

$$\mathbb{E}_{p(x)}[p_{\mathsf{MC}}^N(y_0)] = \frac{1}{N}\sum_{i=1}^N \mathbb{E}_{p(x)}[p(y_0|X_i)] = \frac{1}{N}\sum_{i=1}^N \int p(y_0|x_i)p(x_i)\mathrm{d}x_i = p(y_0).$$

$\boxed{2, \text{B}}$

(iii) Let $\varphi(x) = p(y_0|x)$. The variance of the MC estimator is given by

$$\mathsf{var}_p[p_{\mathsf{MC}}^N(y_0)] = \frac{\mathsf{var}_p(\varphi)}{N}.$$

$\boxed{2, \text{A}}$

(iv) This estimate can be inaccurate if $p(y_0)$ is very small. This can occur with very unlikely observations. $\boxed{1, \text{A}}$

(b) (i) Consider a generic proposal $q(x)$. Given the samples $X_1, \ldots, X_N \sim q(x)$, the IS estimator is $\boxed{\text{sim. seen} \Downarrow}$

$$p_{\mathsf{IS}}^N(y_0) = \frac{1}{N}\sum_{i=}^N p(y_0|X_i)\frac{p(X_i)}{q(X_i)}.$$

$\boxed{2, \text{A}}$

(ii) The proof is given as follows:

$$\begin{aligned}\mathbb{E}_{q(x)}[p_{\mathsf{IS}}^N(y_0)] &= \frac{1}{N}\sum_{i=1}^N \mathbb{E}_{X_i\sim q(x)}\left[p(y_0|X_i)\frac{p(X_i)}{q(X_i)}\right], \\ &= \frac{1}{N}\sum_{i=1}^N \int p(y_0|x)\frac{p(x)}{q(x)}q(x)\mathrm{d}x = p(y_0).\end{aligned}$$

$\boxed{2, \text{B}}$

(iii) Let us write $w(x) = p(x)/q(x)$. We would like to compute the variance

$$\mathrm{var}_q[p_{\mathsf{IS}}^N(y = y_0)] = \mathrm{var}_q\left[\frac{1}{N}\sum_{i=1}^N p(y_0|X)w(X)\right] = \frac{1}{N^2}\mathrm{var}_q\left[\sum_{i=1}^N p(y_0|X_i)w(X_i)\right],$$
$$= \frac{1}{N}\mathrm{var}_q\left[p(y_0|X_i)w(X_i)\right] = \frac{1}{N}\left(\mathbb{E}_q[p^2(y_0|X)w^2(X)] - \mathbb{E}_q[w(X)p(y_0|X)]^2\right)$$
$$= \frac{1}{N}\left(\mathbb{E}_q[p^2(y_0|X)w^2(X)] - p(y_0)^2\right).$$

$\boxed{3, \mathrm{C}}$

(iv) For this estimator to be finite, the first integral has to be finite, in particular

$$p^2(y_0|x)w^2(x)q(x) = p^2(y_0|x)\frac{p^2(x)}{q(x)}$$

has to be integrable. $\boxed{1, \mathrm{A}}$

(c) In order to minimise the variance of the estimator given in part(b)-(iii), only relevant quantity is the expectation inside the final expression. In other words, we seek $\boxed{\text{meth seen} \Downarrow}$

$$\mu^\star = \arg\min_\mu \mathbb{E}_{q_\mu}[p^2(y_0|X)w_\mu^2(X)].$$

This minimisation problem should be clearly stated in the answer. We next derive this expectation

$$\mathbb{E}_{q_\mu}[p^2(y_0|X)w_\mu^2(X)] = \int p^2(y_0|x)\frac{p^2(x)}{q_\mu(x)}\mathrm{d}x = \int \frac{\frac{1}{2\pi}\exp\left(-(y_0-x)^2\right)\frac{1}{2\pi}\exp(-x^2)}{\frac{1}{\sqrt{\pi}}\exp\left(-(x-\mu)^2\right)}\mathrm{d}x,$$
$$= \frac{\sqrt{\pi}}{4\pi^2}\int \exp(-y_0^2 + 2y_0x - x^2 - x^2 + x^2 - 2x\mu + \mu^2)\mathrm{d}x$$
$$= \frac{\sqrt{\pi}}{4\pi^2}\int \exp(-y_0^2 + 2(y_0-\mu)x - x^2 + \mu^2)\mathrm{d}x.$$

Let $g(\mu) = \mathbb{E}_{q_\mu}[p^2(y_0|X)w_\mu^2(X)]$ for a shorthand notation. We complete the square here to obtain a Gaussian density:

$$g(\mu) = \frac{\sqrt{\pi}}{4\pi^2}\int \exp\left(-(y_0-\mu)^2\right)\exp\left(2(y_0-\mu)x\right)\exp(-x^2)\exp(2\mu^2)\exp(-2y_0\mu)\mathrm{d}x,$$
$$= \frac{\pi}{4\pi^2}\exp\left(-2y_0\mu\right)\exp\left(2\mu^2\right)\int \mathcal{N}(x; y_0-\mu, 1/2)\mathrm{d}x,$$
$$= \frac{1}{4\pi}\exp\left(-2y_0\mu\right)\exp(2\mu^2).$$

We can easily maximise this by taking the log

$$\log g(\mu) = -\log\frac{1}{4\pi} - 2y_0\mu + 2\mu^2,$$

taking its derivative w.r.t. μ and setting it to zero gives

$$-2y_0 + 4\mu = 0,$$

which implies $\mu^\star = y_0/2$ as asked. This sampler puts mass of the proposal close to the observation point. If y_0 has very small probability under original model, the MC estimator could fail (as mentioned in part(a)), but the optimal IS estimator will sample close to the observation and will provide an accurate estimate. $\boxed{5, \mathrm{D}}$

3. (a) (i) The detailed balance condition for $p(x)$ and a kernel $K(x'|x)$: seen ⇓

$$p(x)K(x'|x) = p(x')K(x|x').$$

2, A

(ii) Given $p(x, y, z)$ and a state of the chain $(x_{n-1}, y_{n-1}, z_{n-1})$, the Gibbs sampler is given by

1. $x_n \sim p(x|y_{n-1}, z_{n-1})$
2. $y_n \sim p(y|x_n, z_{n-1})$
3. $z_n \sim p(z|x_n, y_n)$.

Marks should be given for these updates in any order. 2, A

(iii) If full conditionals themselves are not available and we are given unnormalised full conditionals $\bar{p}(x|y, z)$, $\bar{p}(y|x, z)$, $\bar{p}(z|x, y)$, one can employ *Metropolis-within-Gibbs*. Given the state of the chain $(x_{n-1}, y_{n-1}, z_{n-1})$, the algorithm would consist of running

1. Sample x_n using Metropolis accept/reject step using $\bar{p}(x|y_{n-1}, z_{n-1})$
2. Sample y_n using Metropolis accept/reject step using $\bar{p}(y|x_n, z_{n-1})$
3. Sample z_n using Metropolis accept/reject step using $\bar{p}(z|x_n, y_n)$ 2, B

(iv) A technique to improve sample quality is *thinning* (1 mark). One other statistic that was provided in the course is *effective sample-size*. 2, A

(b) (i) Given an independent proposal, the MH sampler is given as follows. Assume that the chain is at the sample X_{n-1}, then seen ⇓

1. $X' \sim q(x)$
2. Accept with probability

$$\alpha(X_{n-1}, X') = \min\left\{1, \frac{p(X')q(X_{n-1})}{p(X_{n-1})q(X')}\right\}$$

Otherwise set $X_n = X_{n-1}$. 2, A

(ii) Given a symmetric proposal, the MH sampler is given as follows. Assume that the chain is at the sample X_{n-1}, then

1. $X' \sim q(\cdot|X_{n-1})$
2. Accept with probability

$$\alpha(X_{n-1}, X') = \min\left\{1, \frac{p(X')}{p(X_{n-1})}\right\}$$

Otherwise set $X_n = X_{n-1}$. 2, A

(iii) The MALA algorithm requires us to compute the $\nabla \log p(x)$ which is given by

$$\nabla \log p(x) = -\frac{x - \mu}{\sigma}$$

The MALA proposal with a step-size γ is written as

$$X' = X - \gamma\frac{X - \mu}{\sigma} + \sqrt{2\gamma}W$$

where $W \sim \mathcal{N}(0, 1)$.

We get the proposal as

$$q(x'|x) = \mathcal{N}\left(x'; \left(1 - \frac{\gamma}{\sigma}\right) x + \frac{\gamma\mu}{\sigma}, 2\gamma\right).$$

Given the current state X_{n-1}, the algorithm is then given by

1. Sample $X' = X_{n-1} - \gamma \frac{X_{n-1}-\mu}{\sigma} + \sqrt{2\gamma} W_n$, where $W_n \sim \mathcal{N}(0,1)$.
2. Accept with probability

$$\alpha(X_{n-1}, X') = \min\left\{1, \frac{p(X')q(X_{n-1}|X')}{p(X_{n-1})q(X'|X_{n-1})}\right\}$$

   Otherwise set $X_n = X_{n-1}$.

2, A

(iv) If we skip the Metropolis step, we obtain a Markov chain with a different target than p. If we denote this density as p_γ, this density is close to p if γ is small.

1, C

(c) (i) For the Gibbs sampler, we need the full conditionals $p(x|\sigma^2)$ (given in the question) and $p(\sigma^2|x)$ which we derive as follows. The posterior is given as

meth seen ⇓

$$p(\sigma^2|x) = \frac{p(x|\sigma^2)p(\sigma^2)}{p(x)}$$

Using the expressions given in the question, we have

$$\begin{aligned} p(\sigma^2|x) &\propto \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left(-\frac{\beta}{\sigma^2}\right), \\ &\propto \left(\frac{1}{\sigma^2}\right)^{\alpha+\frac{3}{2}} \exp\left(-\frac{\beta + \frac{x^2}{2}}{\sigma^2}\right). \end{aligned}$$

We can recognise this as the unnormalised form of another inverse-Gamma and write

$$p(\sigma^2|x) = \mathcal{IG}\left(\sigma^2; \alpha + \frac{1}{2}, \beta + \frac{x^2}{2}\right).$$

3, C

(ii) Given the state of this chain $(x_{n-1}, \sigma^2_{n-1})$, the Gibbs sampler can be given as follows

1. Sample $x_n \sim p(x|\sigma^2_{n-1}) = \mathcal{N}(x; 0, \sigma^2_{n-1})$
2. Sample $\sigma^2_n \sim p(\sigma^2|x_n) = \mathcal{IG}\left(\sigma^2; \alpha + \frac{1}{2}, \beta + \frac{x_n^2}{2}\right)$

2, A

4. (a) (i) The filtering problem is the estimation of the distribution $p(x_n|y_{1:n})$ sequentially. As we observe $y_{n+1}, y_{n+2}, \ldots$, we are interested in estimating $p(x_{n+1}|y_{1:n+1})$, $p(x_{n+2}|y_{1:n+2})$ and so on. $\boxed{2, \text{M}}$

(ii) The joint distribution can be written as

$$p(y_{1:n}, x_{0:n}) = \mu(x_0) \prod_{k=1}^{n} g(y_n|x_n) f(x_n|x_{n-1}).$$

$\boxed{2, \text{M}}$

(iii) The filtering recursions are given as follows. Given the filtering distribution $p(x_{n-1}|y_{1:n-1})$, we first have prediction

$$p(x_n|y_{1:n-1}) = \int f(x_n|x_{n-1}) p(x_{n-1}|y_{1:n-1}) \mathrm{d}x_{n-1},$$

and then update

$$p(x_n|y_{1:n}) = p(x_n|y_{1:n-1}) \frac{g(y_n|x_n)}{p(y_n|y_{1:n-1})}.$$

$\boxed{3, \text{M}}$

(b) (i) Given the proposal $q(x_t|x_{t-1})$, the SISR method is given as

1. Sample $\bar{x}_t^{(i)} \sim q(x_t|x_{t-1}^{(i)})$ for $i = 1, \ldots, N$
2. Compute weights

$$W_t^{(i)} = \frac{f(\bar{x}_t^{(i)}|x_{t-1}^{(i)}) g(y_t|\bar{x}_t^{(i)})}{q(\bar{x}_t^{(i)}|x_{t-1}^{(i)})}$$

for $i = 1 \ldots, N$ and normalise

$$\mathsf{w}_t^{(i)} = \frac{W_t^{(i)}}{\sum_{i=1}^N W_t^{(i)}}.$$

for $i = 1, \ldots, N$.

3. Resample

$$\begin{aligned} i_k &\sim \text{Discrete}(\mathsf{w}_1, \ldots, \mathsf{w}_N), \\ x_t^{(k)} &= \bar{x}_t^{(i_k)}, \end{aligned}$$

for $k = 1, \ldots, N$. $\boxed{3, \text{M}}$

(ii) If $q = f$, the algorithm is called the Bootstrap Particle Filter. The pseudo-code is the same as above, with the weighting step replaced with

$$W_t^{(i)} = g(y_t|\bar{x}_t^{(i)})$$

$\boxed{2, \text{M}}$

(iii) The answer is to compute is log-unnormalised weights: $\log W_t^{(i)}$ directly in the log domain. Then, we can compute auxiliary log-weights as

$$\log \bar{W}_t^{(i)} = \log W_t^{(i)} - \max_i \log W_t^{(i)}.$$

We then compute the normalisation

$$\mathsf{w}_t^{(i)} = \frac{\exp(\log \bar{W}_t^{(i)})}{\sum_{i=1}^N \exp(\log \bar{W}_t^{(i)})}$$

$\boxed{1, \mathrm{M}}$

(c) (i) The marginal likelihood estimator in a BPF is given as $\boxed{\text{seen} \Downarrow}$

$$p^N(y_{1:t}|\theta) = \prod_{k=1}^t p^N(y_k|y_{1:k-1}, \theta)$$

where

$$p^N(y_k|y_{1:k-1}, \theta) = \frac{1}{N} \sum_{i=1}^N g_\theta(y_k|\bar{x}_k^{(i)}).$$

$\boxed{3, \mathrm{M}}$

(ii) The particle MH is sampling $p(\theta|y_{1:T})$. The algorithm is given as follows. Given the state of the Markov chain θ_{t-1} $\boxed{\text{unseen} \Downarrow}$

1. Sample $\theta' \sim q(\theta|\theta_{t-1})$
2. Accept with probability

$$\alpha(\theta_{t-1}, \theta') = \min\left\{1, \frac{p(\theta')p^N(y_{1:T}|\theta')q(\theta_{t-1}|\theta')}{p(\theta_{t-1})p^N(y_{1:T}|\theta_{t-1})q(\theta'|\theta_{t-1})}\right\}.$$

The sampler works because the BPF provides unbiased estimates $p^N(y_{1:T}|\theta)$ of $p(y_{1:T}|\theta)$ for any θ. $\boxed{4, \mathrm{M}}$

Review of mark distribution:
Total A marks: 24 of 24 marks
Total B marks: 15 of 15 marks
Total C marks: 9 of 9 marks
Total D marks: 12 of 12 marks
Total marks: 80 of 60 marks
Total Mastery marks: 20 of 20 marks

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.		
ExamModuleCode	**QuestionNumber**	**Comments for Students**
MATH60047/70047	1	This question was mostly well done. The algorithms of inversion sampling and rejection sampling were clearly described and most people were able to complete part (b) where optimal proposal needed to be found via an optimization procedure. In general, the marks for this question were quite high.
MATH60047/70047	2	Overall this question was done well by the students, who could in general remember well the definitions of Monte Carlo method and Importance Sampling estimator. Most students proved unbiasedness, and obtained the variance of the estimators. Some students struggled with a.iv., describing only general scenarios on when the MC method might fail, without linking them specifically to the case of y0 fixed. Part (c) appeared to be the most difficult part of the question: many students managed to set up the problem correctly, but then failed to identify a way to obtain the minimiser. The technique of completing the squares was required, but this was often not applied.
MATH60047/70047	3	In general, this was a straightforward question with mostly asking definitions and pseudocode of the algorithms that are given in the class. Students who attempted the questions demonstrated sufficient knowledge to get high marks. However, there were also an important number of students who only attempted parts of this question. As a result, the average marks for this question is lower than the other questions, despite in general the other questions required more computations (integration, optimization etc...).
MATH70047	4	There was some general confusion about the difference between SIS and SISR (the weighting step). Also some students did not attempt the question fully. Some of the students were able to score very high marks.