Avd. Matematisk statistik

KTH Matematik

EXAMINATION IN SF2955 COMPUTER INTENSIVE METHODS IN MATHEMATICAL STATISTICS
1 JUNE, 2021, 14.00–19.00.

Examiner: Jimmy Olsson, tel. 08 790 72 01

All solutions should be well motivated. The exam consists of three parts, Parts I–III, and you need at least 10 points on each of Part I and Part II to pass. The total amount of points is 50.

Begin the solution of each assignment on top of a new paper. Do not use red pencil. Write your full name on each paper.

You are allowed to use a pocket calculator, a table of formulas for a basic course in mathematical statistics, and a reasonable table of mathematical formulas (e.g. TEFYMA or BETA).

Part I

Problem 1

Let X have standard normal distribution; then the conditional distribution of X given $X > \mu$, where $\mu > 0$ is a constant, has density function

$$f(x) \propto e^{-x^2/2}, \quad x > \mu.$$

We aim to simulate from $f(x)$ using rejection sampling with proposals drawn from a *translated exponential distribution* with density function

$$g_\lambda(x) = \lambda e^{-\lambda(x-\mu)}, \quad x > \mu,$$

where $\lambda > \mu$ is a rate parameter.

(a) Show how to simulate from $g_\lambda(x)$ using the inversion method. (3 p)

(b) Show how to simulate from $f(x)$ using rejection sampling with proposal density $g_\lambda(x)$ and find the rate parameter λ for which the expected number of trials before acceptance is minimal. (7 p)

Problem 2

(a) Let $f(x, y)$ be a probability density allowing simulation from the full conditionals $f(x \mid y)$ and $f(y \mid x)$. Describe how to generate samples from $f(x, y)$ using Gibbs sampling and show that the Markov chain produced by the Gibbs sampler has $f(x, y)$ as a stationary distribution. (6 p)

(b) Now consider the following modification of the standard Gibbs sampler in (a). Let $\alpha \in (0, 1)$ be fixed and simulate recursively a Markov chain $(X_n, Y_n)_{n \geq 0}$ as follows: given (X_n, Y_n),

> draw $U \sim \text{Unif}(0, 1)$;
> **if** $U \leq \alpha$ **then**
>   draw $Y_{n+1} \sim f(y \mid x = X_n)$;
>   draw $X_{n+1} \sim f(x \mid y = Y_{n+1})$;
> **else**
>   draw $X_{n+1} \sim f(x \mid y = Y_n)$;
>   draw $Y_{n+1} \sim f(y \mid x = X_{n+1})$;
> **end**
> **return** (X_{n+1}, Y_{n+1})

Show that the Markov chain $(X_n, Y_n)_{n \geq 0}$ generated in this way has $f(x, y)$ as a stationary distribution. You may use the result in (a). (4 p)

Part II

Problem 3

Let $(X_n)_{n \geq 0}$ be a Markov chain with state space $\mathsf{X} = \{1, 2\}$ and transition probabilities $\{q(i; j) : (i, j) \in \mathsf{X}^2\}$ given by the transition matrix

$$\mathbf{Q} = \begin{pmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{pmatrix},$$

where $\mathbf{Q}_{ij} = q(i; j)$ and $p_{11} \in (0, 1)$, $p_{22} \in (0, 1)$. At time zero, the chain is in state 1 or 2 with equal probability. In order to calibrate the model on the basis of an observed realisation $x_{0:n}$ given by

$$11121222112121222222121212112121121112122,$$

with $n = 40$, we take a Bayesian approach where the unknown probabilities p_{11} and p_{22} are assumed to be *a priori* independent with marginal distributions

$$p_{11} \sim \text{Beta}(\alpha_1, \beta_1), \quad p_{22} \sim \text{Beta}(\alpha_2, \beta_2),$$

with α_1, β_1, α_2, and β_2 being known hyperparameters.

(a) Find the conditional distributions of

- p_{11} given p_{22} and $x_{0:n}$
- p_{22} given p_{11} and $x_{0:n}$. (7 p)

(b) How can the distributions in (a) be used for sampling from the observed data posterior, *i.e.*, the joint conditional distribution of p_{11} and p_{22} given $x_{0:n}$, by means of MCMC? Provide a detailed pseudocode. (3 p)

Problem 4

Customers enter into a service system in two different cases, Case 1 and Case 2. The distribution of a customer's service time depends on the case; more precisely, the service time Y_i of the ith costumer is

$$Y_i \mid X_i = x \sim \text{Exp}(\theta_x),$$

where X_i is the case of the customer and $\theta = (\theta_1, \theta_2) \in \mathbb{R}_+^2$ is a parameter. Since both cases are equally frequent, $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = 2) = 1/2$. Cases and observation times are assumed to be independent between customers. Moreover, a customer's case is private, and an external observer has therefore only access to measured service times. In this setting, our aim is to estimate the parameter θ, which is assumed to be unknown, on the basis of a given record $\mathbf{y} = (y_1, \ldots, y_n)$ of n customers' service times. Since $\mathbf{x} = (x_1, \ldots, x_n)$ is unobserved, the problem can be naturally cast into the framework of the EM algorithm.

(a) Compute, up to a constant not depending on θ, the complete data log-likelihood $\ln f_\theta(\mathbf{x}, \mathbf{y})$ of the model. (3 p)

(b) For a given parameter θ, find the conditional distribution of X_i given Y_i. (3 p)

(c) Use (a) and (b) to derive an EM updating formula for θ. (4 p)

Part III

Problem 5

Let $(X_n)_{n \geq 0}$ be a Markov chain on $\mathbb{R}$ evolving according to the autoregressive model

$$X_{n+1} = \alpha X_n + \sigma \varepsilon_{n+1}, \quad n \geq 0,$$

where $(\varepsilon_n)_{n \geq 1}$ are independent and standard normally distributed and $\alpha \in (0, 1)$ and $\sigma > 0$ are known parameters. Denote by q the transition density of the chain. The distribution χ of the initial state X_0 is the zero-mean normal distribution with variance $\sigma^2/(1 - \alpha^2)$. Let $(A_n)_{n \geq 1}$ be a sequence of intervals $A_n = (a_n, b_n) \subset \mathbb{R}$ and define the rectangle sets $A^n = A_1 \times A_2 \times \cdots \times A_n \subset \mathbb{R}^n$, $n \geq 1$. Then for every n, let

$$f_{A^n}(x_0, \ldots, x_n) \propto \chi(x_0) \prod_{m=0}^{n-1} q(x_m; x_{m+1}) \mathbb{1}_{A_{m+1}}(x_{m+1})$$

be the conditional distribution of the states $(X_0, \ldots, X_n)$ given the event $(X_1, \ldots, X_n) \in A^n$.

(a) Provide a detailed pseudo-code for the sequential importance sampling with resampling (SISR) algorithm sampling sequentially from the conditional distributions $f_{A^n}(x_0, \ldots, x_n)$, $n \geq 1$. Use q as proposal kernel, initialise the particles according to χ, and denote by N the number of particles. (3 p)

(b) As usual, denote by Ω_m the sum of the particle weights at time m. What is the expectation of $\frac{1}{N^{n+1}} \prod_{m=0}^{n} \Omega_m$ in this case? (3 p)

(c) A possible improvement of the algorithm in (a) is to mutate the particles using the instrumental transition density

$$g_n(x_n; x_{n+1}) \propto q(x_n; x_{n+1}) \mathbb{1}_{A_{n+1}}(x_{n+1}),$$

i.e., the conditional distribution of X_{n+1} given X_n and $X_{n+1} \in A_{n+1}$, instead of $q(x_n; x_{n+1})$. Show how to simulate from $g_n(x_n; x_{n+1})$ for a given x_n using the inversion method. You may assume that you are able to evaluate the distribution function Φ of the standard normal distribution as well as its inverse Φ^{-1}. (3 p)

(d) Adapt your algorithm in (a) to the modified proposal in (c) and provide a detailed pseudocode. (1 p)

A Some relevant distributions

- The Beta(α, β) distribution, where $\alpha > 0$ and $\beta > 0$, has density function

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad x \in (0,1),$$

  where B is the beta function. It holds that $\mathbb{E}(\ln X) = \psi(\alpha) - \psi(\alpha + \beta)$, where ψ is the *digamma function*.

- The Exp(λ) distribution, where $\lambda > 0$, has density function

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

- The $N(\mu, \sigma^2)$ distribution has density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}, \quad x \in \mathbb{R}.$$

Avd. Matematisk statistik

KTH Matematik

SOLUTIONS TO
EXAM IN SF2955 COMPUTER INTENSIVE METHODS IN MATHEMATICAL STATISTICS
1 JUNE, 2021, 14.00–19.00.

Problem 1

(a) We first calculate the distribution function G_λ of the translated exponential distribution according to

$$G_\lambda(x) = \int_{-\infty}^{x} g_\lambda(z)\,dz = \int_{\mu}^{x} \lambda e^{-\lambda(z-\mu)}\,dz = 1 - e^{-\lambda(x-\mu)}, \quad x > \mu.$$

Next, in order to compute the inverse G_λ^{-1} of G_λ, let $u \in (0,1)$ and write

$$u = G_\lambda(x) \Leftrightarrow u = 1 - e^{-\lambda(x-\mu)} \Leftrightarrow x = \mu - \frac{1}{\lambda}\ln(1-u) = G_\lambda^{-1}(u).$$

Thus, we may simulate X from g_λ by generating $U \sim \mathrm{U}(0,1)$ and letting $X = G_\lambda^{-1}(U) = \mu - \ln(1-U)/\lambda$.

(b) First, we note that $f(x) = e^{-x^2/2}\mathbb{1}_{(\mu,\infty)}(x)/c$, where c is a normalising constant. Next, in order to sample from f using rejection sampling we need to find a constant $K_\lambda > 0$ such that

$$f(x) \leq K_\lambda g_\lambda(x), \quad x > \mu. \tag{1}$$

(Note that since g_λ depends on λ, so does K_λ; thus, we have added λ as a subscript to K_λ.) Then we may generate a draw X from f using the following rejection-sampling algorithm:

set `accepted` $\leftarrow$ `false`;
while `accepted` $=$ `false` **do**
  draw $X^* \sim g_\lambda(x)$;
  draw $U \sim \mathrm{U}(0,1)$;
  if $U \leq \frac{f(X^*)}{K_\lambda g_\lambda(X^*)}$ **then**
    set $X \leftarrow X^*$;
    set `accepted` $\leftarrow$ `true`;
  end
end
return X

Since the expected number of trials before acceptance is K_λ, we should find the smallest possible K_λ for which (1) holds. We proceed in two steps:

1. For a fixed λ, find the smallest K_λ satisfying (1).

2. Hereafter, find the λ minimising this K_λ.

In order to carry through the first step we should let K_λ be the maximum of

$$\rho(x) = \frac{f(x)}{g_\lambda(x)} = \frac{1}{c\lambda} e^{-x^2/2+\lambda x-\lambda\mu}$$

over $x \in (\mu, \infty)$. Note that finding the maximum of ρ corresponds to finding the maximum of the mapping $x \mapsto -x^2/2 + \lambda x$, which is given by $x = \lambda$. This yields

$$K_\lambda = \frac{1}{c\lambda} e^{\lambda^2/2-\lambda\mu}.$$

Next, in order to carry through the second step we should find the parameter $\lambda > \mu$ that minimises the previous K_λ; thus, we solve

$$\frac{d}{d\lambda} K_\lambda = \frac{1}{c\lambda^2}(\lambda^2 - \lambda\mu - 1)e^{\lambda^2/2-\lambda\mu} = 0 \Leftrightarrow \lambda = \frac{\mu \pm \sqrt{\mu^2+4}}{2}.$$

Out of the two solutions only

$$\lambda = \frac{1}{2}(\mu + \sqrt{\mu^2+4}) \tag{2}$$

satisfies $\lambda > \mu$, and it is readily checked (*e.g.* by inspecting the sign of the second derivative) that this point indeed corresponds to a minimum. Thus, (2) provides the final answer to the problem.

Problem 2

(a) See lectures.

(b) Let $q_1(x, y; x', y')$ and $q_2(x, y; x', y')$ denote the transition densities of the standard Gibbs sampler when the y component is updated before the x component and vice versa. In order to determine the transition density $q(x, y; x', y')$ of the Markov chain generated by the algorithm in (b), we pick an arbitrary set A and write, using that U is independent of everything else,

$$\begin{aligned}
&\mathbb{P}((X_{n+1}, Y_{n+1}) \in A \mid X_n = x, Y_n = y) \\
&\quad = \mathbb{P}((X_{n+1}, Y_{n+1}) \in A \mid X_n = x, Y_n = y, U \le \alpha)\mathbb{P}(U \le \alpha) \\
&\qquad + \mathbb{P}((X_{n+1}, Y_{n+1}) \in A \mid X_n = x, Y_n = y, U > \alpha)\mathbb{P}(U > \alpha) \\
&\quad = \alpha \int_A q_1(x, y; x', y')\, dx'\, dy' + (1-\alpha) \int_A q_2(x, y; x', y')\, dx'\, dy',
\end{aligned}$$

which yields

$$q(x, y; x', y') = \alpha q_1(x, y; x', y') + (1-\alpha) q_2(x, y; x', y').$$

Now, by (a) we know that q_1 and q_2 satisfy global balance with respect to the target f, *i.e.*,

$$\int f(x, y) q_1(x, y; x', y')\, dx\, dy = \int f(x, y) q_2(x, y; x', y')\, dx\, dy = f(x', y');$$

thus, we may establish that also q satisfies global balance by proceeding like

$$\begin{aligned}
&\int f(x, y) q(x, y; x', y')\, dx\, dy \\
&= \alpha \int f(x, y) q_1(x, y; x', y')\, dx\, dy + (1-\alpha) \int f(x, y) q_2(x, y; x', y')\, dx\, dy \\
&= \alpha f(x', y') + (1-\alpha) f(x', y') \\
&= f(x', y').
\end{aligned}$$

This answers (b).

Problem 3

(a) We first compute the likelihood function

$$
\begin{aligned}
f(x_{0:n} \mid p_{11}, p_{22}) &= \frac{1}{2} \prod_{k=0}^{n-1} q(x_k; x_{k+1}) \\
&= \frac{1}{2} p_{11} p_{11} (1 - p_{11})(1 - p_{22}) \cdots (1 - p_{11}) p_{22} \\
&= \frac{1}{2} p_{11}^7 (1 - p_{11})^{13} p_{22}^8 (1 - p_{22})^{12},
\end{aligned}
$$

where the exponents are obtained by counting, in the given sequence, the number of transitions of each of the four types. In addition, in the first step we used that X_0 takes on the values 1 and 2 with equal probability 1/2. We may now write the full conditional of p_{11} given p_{22} and $x_{0:n}$ up to a constant of proportionality according to

$$
\begin{aligned}
\pi(p_{11} \mid p_{22}, x_{0:n}) \propto f(x_{0:n} \mid p_{11}, p_{22}) \pi(p_{11}) \propto p_{11}^7 (1 - p_{11})^{13} p_{11}^{\alpha_1 - 1} (1 - p_{11})^{\beta_1 - 1} \\
= p_{11}^{6 + \alpha_1} (1 - p_{11})^{12 + \beta_1},
\end{aligned}
$$

and since densities that are proportional are equal we conclude that

$$
p_{11} \mid p_{22}, x_{0:n} \sim \mathrm{Beta}(7 + \alpha_1, 13 + \beta_1).
$$

Similarly,

$$
\pi(p_{22} \mid p_{11}, x_{0:n}) \propto p_{22}^{7 + \alpha_2} (1 - p_{22})^{11 + \beta_2},
$$

and hence

$$
p_{22} \mid p_{11}, x_{0:n} \sim \mathrm{Beta}(8 + \alpha_2, 12 + \beta_2).
$$

This answers (a).

(b) Since the full conditionals in (a) are known, samplable distributions, we may use the Gibbs sampler to generate recursively a random sequence $(p_{11}^n, p_{22}^n)_{n \geq 0}$ targeting the joint posterior $\pi(p_{11}, p_{22} \mid x_{0:n})$ as follows: given (p_{11}^n, p_{22}^n),

draw $p_{11}^{n+1} \sim \pi(p_{11} \mid p_{22}^n, x_{0:n})$;
draw $p_{22}^{n+1} \sim \pi(p_{22} \mid p_{11}^{n+1}, x_{0:n})$;

However, note that the full conditional of p_{11} derived in (a) does *not* depend on p_{22} and vice versa; this means that the parameters p_{11} and p_{22} "decouple" for this simple Markov model and are also a posteriori independent. As a consequence, *the Gibbs sampler outlined above actually produces independent and identically distributed draws from* $\pi(p_{11}, p_{22} \mid x_{0:n})$.

Problem 4

(a) In the following we assign the index θ to any density that depends on the parameter θ. First, note that

$$f_\theta(y_i \mid x_i) = \left\{ \begin{array}{ll} \theta_1 e^{-\theta_1 y_i} & \text{if } x_i = 1 \\ \theta_2 e^{-\theta_2 y_i} & \text{if } x_i = 2 \end{array} \right\} = (\theta_1 e^{-\theta_1 y_i})^{2-x_i} (\theta_2 e^{-\theta_2 y_i})^{x_i - 1}, \quad x_i \in \{1, 2\}. \tag{3}$$

Using the previous expression we may write the complete-data likelihood up to a constant of proportionality (not depending on θ) according to

$$f_\theta(\mathbf{x}, \mathbf{y}) \propto f_\theta(\mathbf{y} | \mathbf{x}) = \prod_{i=1}^n f_\theta(y_i \mid x_i) = \prod_{i=1}^n (\theta_1 e^{-\theta_1 y_i})^{2-x_i} (\theta_2 e^{-\theta_2 y_i})^{x_i - 1}$$

(where we used that the marginal of $\mathbf{x}$ does not depend on θ). Now, let "$\overset{\text{c.}}{=}$" denote equality up to additive constants not depending on θ; then,

$$\begin{aligned} \ln f_\theta(\mathbf{x}, \mathbf{y}) &\overset{\text{c.}}{=} \sum_{i=1}^n ((2 - x_i)(\ln \theta_1 - \theta_1 y_i) + (x_i - 1)(\ln \theta_2 - \theta_2 y_i)) \\ &= \ln \theta_1 \left(2n - \sum_{i=1}^n x_i \right) - \theta_1 \sum_{i=1}^n (2 - x_i) y_i + \ln \theta_2 \left(\sum_{i=1}^n x_i - n \right) - \theta_2 \sum_{i=1}^n (x_i - 1) y_i, \end{aligned}$$

which answers (a).

(b) For a given parameter θ, we examine the probability function $f_\theta(x_i \mid y_i)$ up to a constant of proportionality (not depending on x_i):

$$f_\theta(x_i \mid y_i) \propto f_\theta(y_i \mid x_i) f(x_i) = \frac{1}{2} (\theta_1 e^{-\theta_1 y_i})^{2-x_i} (\theta_2 e^{-\theta_2 y_i})^{x_i - 1}, \quad x_i \in \{1, 2\},$$

where we used (3), and normalising the previous function yields

$$f_\theta(x_i \mid y_i) = \left(\frac{\theta_1 e^{-\theta_1 y_i}}{\theta_1 e^{-\theta_1 y_i} + \theta_2 e^{-\theta_2 y_i}} \right)^{2-x_i} \left(\frac{\theta_2 e^{-\theta_2 y_i}}{\theta_1 e^{-\theta_1 y_i} + \theta_2 e^{-\theta_2 y_i}} \right)^{x_i - 1} = \alpha_\theta(y_i)^{2-x_i} (1 - \alpha_\theta(y_i))^{x_i - 1}, \quad x_i \in \{1, 2\},$$

where we have introduced the short-hand notation

$$\alpha_\theta(y_i) = \frac{\theta_1 e^{-\theta_1 y_i}}{\theta_1 e^{-\theta_1 y_i} + \theta_2 e^{-\theta_2 y_i}}$$

(so that $\alpha_\theta(y_i) = f_\theta(x_i = 1 \mid y_i)$). This answers (b).

(c) Using the expression of the complete-data log-likelihood obtained in (a), the intermediate quantity of EM can be written as

$$\begin{aligned} \mathcal{Q}_{\theta'}(\theta) &= \mathbb{E}_{\theta'}[\ln f_\theta(\mathbf{X}, \mathbf{Y}) \mid \mathbf{Y} = \mathbf{y}] \\ &\overset{\text{c.}}{=} \ln \theta_1 \left(2n - \sum_{i=1}^n \mathbb{E}_{\theta'}[X_i \mid Y_i = y_i] \right) - \theta_1 \sum_{i=1}^n (2 - \mathbb{E}_{\theta'}[X_i \mid Y_i = y_i]) y_i \\ &\quad + \ln \theta_2 \left(\sum_{i=1}^n \mathbb{E}_{\theta'}[X_i | Y_i = y_i] - n \right) - \theta_2 \sum_{i=1}^n (\mathbb{E}_{\theta'}[X_i | Y_i = y_i] - 1) y_i. \end{aligned} \tag{4}$$

Now, using the conditional distribution obtained in (b),

$$\begin{aligned}\mathbb{E}_{\theta'}[X_i \mid Y_i = y_i] &= 1 \cdot f_{\theta'}(x_i = 1 \mid y_i) + 2 \cdot f_{\theta'}(x_i = 2 \mid y_i) \\ &= \alpha_{\theta'}(y_i) + 2(1 - \alpha_{\theta'}(y_i)) \\ &= 2 - \alpha_{\theta'}(y_i),\end{aligned}$$

and plugging this expression into (4) yields

$$\begin{aligned}\mathcal{Q}_{\theta'}(\theta) &= \ln\theta_1 \left(2n - \sum_{i=1}^{n}(2 - \alpha_{\theta'}(y_i))\right) - \theta_1 \sum_{i=1}^{n}(2 - 2 + \alpha_{\theta'}(y_i))y_i \\ &\quad + \ln\theta_2 \left(\sum_{i=1}^{n}(2 - \alpha_{\theta'}(y_i)) - n\right) - \theta_2 \sum_{i=1}^{n}(2 - \alpha_{\theta'}(y_i) - 1)y_i \\ &= \ln\theta_1 \sum_{i=1}^{n} \alpha_{\theta'}(y_i) - \theta_1 \sum_{i=1}^{n} y_i \alpha_{\theta'}(y_i) + \ln\theta_2 \left(n - \sum_{i=1}^{n} \alpha_{\theta'}(y_i)\right) - \theta_2 \sum_{i=1}^{n} y_i(1 - \alpha_{\theta'}(y_i)).\end{aligned}$$

This is the E-step. In order carry through the M-step, *i.e.*, to maximise $\mathcal{Q}_{\theta'}(\theta)$ with respect to θ, we solve

$$\begin{aligned}\nabla_\theta \mathcal{Q}_{\theta'}(\theta) = \mathbf{0} &\Leftrightarrow \begin{cases} \frac{\partial}{\partial\theta_1}\mathcal{Q}_{\theta'}(\theta) = \theta_1^{-1}\sum_{i=1}^{n}\alpha_{\theta'}(y_i) - \sum_{i=1}^{n} y_i\alpha_{\theta'}(y_i) = 0 \\ \frac{\partial}{\partial\theta_2}\mathcal{Q}_{\theta'}(\theta) = \theta_2^{-1}\left(n - \sum_{i=1}^{n}\alpha_{\theta'}(y_i)\right) - \sum_{i=1}^{n} y_i(1 - \alpha_{\theta'}(y_i)) = 0 \end{cases} \\ &\Leftrightarrow \begin{cases} \theta_1 = \dfrac{\sum_{i=1}^{n}\alpha_{\theta'}(y_i)}{\sum_{i=1}^{n} y_i\alpha_{\theta'}(y_i)} \\ \theta_2 = \dfrac{n - \sum_{i=1}^{n}\alpha_{\theta'}(y_i)}{\sum_{i=1}^{n} y_i(1 - \alpha_{\theta'}(y_i))}. \end{cases}\end{aligned} \tag{5}$$

It is readily checked (*e.g.* by computing the Hessian) that the solution found in (5) provides indeed a global maximum. We hence obtain the following recursive EM-updating formula: given an approximation $\theta^\ell = (\theta_1^\ell, \theta_2^\ell)$ of the maximum-likelihood estimate, an improved approximation $\theta^{\ell+1} = (\theta_1^{\ell+1}, \theta_2^{\ell+1})$ is given by

$$\theta^{\ell+1} = \begin{cases} \theta_1^{\ell+1} = \dfrac{\sum_{i=1}^{n}\alpha_{\theta^\ell}(y_i)}{\sum_{i=1}^{n} y_i p_{\theta^\ell}(y_i)} \\ \theta_2^{\ell+1} = \dfrac{n - \sum_{i=1}^{n}\alpha_{\theta^\ell}(y_i)}{\sum_{i=1}^{n} y_i(1 - \alpha_{\theta^\ell}(y_i))}, \end{cases}$$

where

$$\alpha_{\theta^\ell}(y_i) = \frac{\theta_1^\ell e^{-\theta_1^\ell y_i}}{\theta_1^\ell e^{-\theta_1^\ell y_i} + \theta_2^\ell e^{-\theta_2^\ell y_i}},$$

and the scheme is iterated until convergence. This answers (c).

Problem 5

(a) For $n \geq 1$, let $f_{A^n}(x_{0:n}) = z_n(x_{0:n})/c_n$, where

$$z_n(x_{0:n}) = \chi(x_0) \prod_{m=0}^{n-1} q(x_m; x_{m+1}) \mathbb{1}_{A_{m+1}}(x_{m+1})$$

and c_n is the normalising constant. Note that sampling from χ and $q(x_m; \cdot)$ corresponds to sampling from the $\mathrm{N}(0, \sigma^2/(1-\alpha^2))$ and $\mathrm{N}(\alpha x_m, \sigma^2)$ distributions, respectively. The following pseudocode describes how to sample from $f_{A^n}(x_{0:n})$, $n \geq 1$, by means of the sequential importance sampling with resampling (SISR) algorithm.

for $i \leftarrow 1$ **to** N **do**
  draw $X_0^i \sim \mathrm{N}\left(0, \dfrac{\sigma^2}{1-\alpha^2}\right)$;
  set $\omega_0^i \leftarrow 1$;
end
for $n \leftarrow 0, 1, 2, \ldots$ **do**
  draw, with replacement, new particles $(\tilde{X}_{0:n}^i)_{i=1}^N$ among $(X_{0:n}^i)_{i=1}^N$ according to weights proportional to $(\omega_n^i)_{i=1}^N$;
  for $i \leftarrow 1$ **to** N **do**
    draw $X_{n+1}^i \sim \mathrm{N}(\alpha \tilde{X}_n^i, \sigma^2)$;
    set $X_{0:n+1}^i \leftarrow (\tilde{X}_{0:n}^i, X_{n+1}^i)$;
    set $\omega_{n+1}^i \leftarrow \dfrac{z_{n+1}(X_{0:n+1}^i)}{z_n(\tilde{X}_{0:n}^i) q(\tilde{X}_n^i; X_{n+1}^i)} = \mathbb{1}_{A_{n+1}}(X_{n+1}^i)$;
  end
end

Note that the weights of the updated particles will be either zero or one depending on whether their last components belong to the interval $A_{n+1} = (a_{n+1}, b_{n+1})$.

(b) We know that $\frac{1}{N^{n+1}} \prod_{m=0}^{n} \Omega_m$ is an unbiased estimator of the normalising constant c_n, which remains to be determined. However, since $\chi(x_0) \prod_{m=0}^{n-1} q(x_m; x_{m+1})$ is the density of the joint law of the Markov states $(X_0, \ldots, X_n)$, it holds that

$$\begin{aligned} c_n &= \int \chi(x_0) \prod_{m=0}^{n-1} q(x_m; x_{m+1}) \mathbb{1}_{A_{m+1}}(x_{m+1}) \, dx_{0:n} \\ &= \int_{A^n} \int_{\mathbb{R}} \chi(x_0) \prod_{m=0}^{n-1} q(x_m; x_{m+1}) \, dx_0 \, dx_{1:n} \\ &= \mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n), \end{aligned}$$

i.e., the probability that $X_{1:n}$ belongs to the rectangle A^n. This answers (b).

(c) Recall from (a) that $q(x_n; x_{n+1})$ is the density of the $\mathrm{N}(\alpha x_n, \sigma^2)$ distribution, whose distribution function is given by $x_{n+1} \mapsto \Phi((x_{n+1} - \alpha x_n)/\sigma)$. In order to simulate, for a given x_n, from

$g_n(x_n; x_{n+1})$ using the inversion method we need, first, to calculate the distribution function

$$\begin{aligned} G(x) &= \int_{-\infty}^{x} g_n(x_n; z)\, dz \\ &= \frac{\int_{-\infty}^{x} q(x_n; z)\mathbb{1}_{A_{n+1}}(z)\, dz}{\int_{-\infty}^{\infty} q(x_n; z')\mathbb{1}_{A_{n+1}}(z')\, dz'} \\ &= \frac{\Phi((x - \alpha x_n)/\sigma) - \Phi((a_{n+1} - \alpha x_n)/\sigma)}{\Phi((b_{n+1} - \alpha x_n)/\sigma) - \Phi((a_{n+1} - \alpha x_n)/\sigma)} \end{aligned}$$

(which implicitly depends on x_n). Next, in order to compute the inverse G^{-1} of G, pick $u \in (0, 1)$ and write

$$\begin{aligned} u = G(x) &\Leftrightarrow u = \frac{\Phi((x - \alpha x_n)/\sigma) - \Phi((a_{n+1} - \alpha x_n)/\sigma)}{\Phi((b_{n+1} - \alpha x_n)/\sigma) - \Phi((a_{n+1} - \alpha x_n)/\sigma)} \\ &\Leftrightarrow \Phi\left(\frac{x - \alpha x_n}{\sigma}\right) = u\left(\Phi\left(\frac{b_{n+1} - \alpha x_n}{\sigma}\right) - \Phi\left(\frac{a_{n+1} - \alpha x_n}{\sigma}\right)\right) + \Phi\left(\frac{a_{n+1} - \alpha x_n}{\sigma}\right) \\ &\Leftrightarrow x = G^{-1}(u), \end{aligned}$$

where

$$G^{-1}(u) = \alpha x_n + \sigma \Phi^{-1}\left(u\left[\Phi\left(\frac{b_{n+1} - \alpha x_n}{\sigma}\right) - \Phi\left(\frac{a_{n+1} - \alpha x_n}{\sigma}\right)\right] + \Phi\left(\frac{a_{n+1} - \alpha x_n}{\sigma}\right)\right).$$

Thus, we may simulate X_{n+1} from $g_n(x_n; x_{n+1})$ by drawing first $U \sim \mathrm{U}(0, 1)$ and then letting $X_{n+1} = G^{-1}(U)$.

(d) We modify the SISR algorithm in (a) by changing the proposal q to g_n. This modification will affect not only the mutation step, but also the way the particle weights are computed; more precisely, with the proposal g_n, the weights will be updated according to

$$\begin{aligned} \omega_{n+1}^{i} &= \frac{z_{n+1}(X_{0:n+1}^{i})}{z_n(\tilde{X}_{0:n}^{i}) g_n(\tilde{X}_n^{i}; X_{n+1}^{i})} \\ &= \frac{q(\tilde{X}_n^{i}; X_{n+1}^{i})\mathbb{1}_{A_{n+1}}(X_{n+1}^{i})}{g_n(\tilde{X}_n^{i}; X_{n+1}^{i})} \\ &= \frac{q(\tilde{X}_n^{i}; X_{n+1}^{i})\mathbb{1}_{A_{n+1}}(X_{n+1}^{i})}{q(\tilde{X}_n^{i}; X_{n+1}^{i})\mathbb{1}_{A_{n+1}}(X_{n+1}^{i}) / (\int_{-\infty}^{\infty} q(\tilde{X}_n^{i}; z)\mathbb{1}_{A_{n+1}}(z)\, dz)} \\ &= \Phi\left(\frac{b_{n+1} - \alpha \tilde{X}_n^{i}}{\sigma}\right) - \Phi\left(\frac{a_{n+1} - \alpha \tilde{X}_n^{i}}{\sigma}\right). \end{aligned}$$

In other words: in this case, ω_{n+1}^{i} is the normalising constant of $g_n(\tilde{X}_n^{i}, \cdot)$. The full modified SISR algorithm goes as follows.

for $i \leftarrow 1$ **to** N **do**
  draw $X_0^i \sim \mathrm{N}\left(0, \dfrac{\sigma^2}{1-\alpha^2}\right)$;
  set $\omega_0^i \leftarrow 1$;
end
for $n \leftarrow 0, 1, 2, \ldots$ **do**
  draw, with replacement, new particles $(\tilde{X}_{0:n}^i)_{i=1}^N$ among $(X_{0:n}^i)_{i=1}^N$ according to weights proportional to $(\omega_n^i)_{i=1}^N$;
  for $i \leftarrow 1$ **to** N **do**
    draw $X_{n+1}^i \sim g_n(\tilde{X}_n^i; x_{n+1})$ using the inversion method in (c);
    set $X_{0:n+1}^i \leftarrow (\tilde{X}_{0:n}^i, X_{n+1}^i)$;
    set $\omega_{n+1}^i \leftarrow \Phi\left(\dfrac{b_{n+1} - \alpha \tilde{X}_n^i}{\sigma}\right) - \Phi\left(\dfrac{a_{n+1} - \alpha \tilde{X}_n^i}{\sigma}\right)$;
  end
end
return $(X_{0:n}^i, \omega_n^i)_{i=1}^N$