Mock exam: solutions

U. Vaes

March 2020

Question 1. 1. We first calculate the normalization constant:

$$Z = \int_{-\infty}^{\infty} \mathrm{e}^{-\lambda x}\, \mathbb{1}_{[0,\alpha]}(x)\, \mathrm{d}x = \frac{1 - \mathrm{e}^{-\lambda \alpha}}{\lambda}.$$

The cumulative distribution is given by

$$G_{\alpha,\lambda}(x) = \int_{-\infty}^{x} g_{\alpha,\lambda}(y)\, \mathrm{d}y = \frac{1}{Z} \int_{-\infty}^{x} \mathrm{e}^{-\lambda y}\, \mathbb{1}_{[0,\alpha]}(y)\, \mathrm{d}y = \max\left(0, \min\left(1, \frac{1 - \mathrm{e}^{-\lambda x}}{1 - \mathrm{e}^{-\lambda \alpha}}\right)\right),$$

which is a compact way of writing

$$G_{\alpha,\lambda}(x) \begin{cases} 0, & \text{if } x < 0; \\ \frac{1-\mathrm{e}^{-\lambda x}}{1-\mathrm{e}^{-\lambda\alpha}} & \text{if } 0 \leq x \leq \alpha; \\ 1, & \text{if } x > \alpha. \end{cases}$$

The generalized inverse of $G_{\alpha,\lambda}$ is given by

$$F_{\alpha,\lambda}(u) = \inf\{x : G_{\alpha,\lambda}(x) \geq u\} = -\frac{1}{\lambda} \log\left(1 - u(1 - \mathrm{e}^{-\lambda\alpha})\right)$$

To check our result, we can verify that $F_{\alpha,\lambda}(0) = 0$ and $F_{\alpha,\lambda}(1) = \alpha$. Notice that $F_{\alpha,\lambda}$ coincides with $G_{\alpha,\lambda}^{-1}$ on $(0,1)$, which is how we obtained the expression for $F_{\alpha,\lambda}$: for $u \in (0,1)$,

$$G_{\alpha,\lambda}(x) = u \Leftrightarrow \frac{1 - \mathrm{e}^{-\lambda x}}{1 - \mathrm{e}^{-\lambda\alpha}} = u \Leftrightarrow x = -\frac{1}{\lambda} \log\left(1 - u(1 - \mathrm{e}^{-\lambda\alpha})\right).$$

If $\{U_i\}_{i=1,2,\dots}$ is a stream of independent $U(0,1)$ random variables, then $\{F_{\alpha,\lambda}(U_i)\}_{i=1,2,\dots}$ is a stream of IID samples from $g_{\lambda,\alpha}$.

2. Let us rewrite the expression of $f(x)$ for convenience:

$$f(x) = 32\, \frac{x(1-x)\, \mathrm{e}^{-4(x-1)}}{3 + \mathrm{e}^4}\, \mathbb{1}_{[0,1]}(x).$$

The derivative of $f(x)$ in $(0,1)$ is given by

$$f'(x) = 32\, \frac{(1-2x)\, \mathrm{e}^{-4(x-1)} - 4x(1-x)\, \mathrm{e}^{-4(x-1)}}{3 + \mathrm{e}^4},$$

which vanishes when

$$(1-2x) - 4x(1-x) = 0 \Leftrightarrow 4x^2 - 6x + 1 = 0 \Leftrightarrow \left(2x - \frac{3}{2}\right)^2 - \frac{5}{4} = 0,$$

i.e. when

$$x = \frac{3 \pm \sqrt{5}}{4}.$$

Of the two roots, only $x_1 := \frac{3-\sqrt{5}}{4}$ lies in $[0,1]$. To check that this root corresponds to a maximum, we can calculate the second derivative

$$f''(x) = -64\,\frac{(8x^2 - 16x + 5)\,\mathrm{e}^{-4(x-1)}}{3+\mathrm{e}^4} = -512\,\frac{\left((x-1)^2 - \frac{3}{8}\right)\,\mathrm{e}^{-4(x-1)}}{3+\mathrm{e}^4},$$

which is negative at x_1, so we conclude

$$\operatorname*{arg\,max}_{x\in[0,1]} \frac{f(x)}{h(x)} = \operatorname*{arg\,max}_{x\in[0,1]}\, f(x) = x_1.$$

This implies that the best (i.e. the smallest, since the acceptance probability/rate is given by $\frac{1}{M}$ and we want to maximize this rate) constant for rejection sampling is given by

$$\begin{aligned} M_1 &:= \inf\{M : f(x) \leq M\,h(x) \forall x \in [0,1]\} = \inf\left\{M : \frac{f(x)}{h(x)} \leq M \;\forall x \in [0,1]\right\} \\ &= \inf\left\{M : \frac{f(x_1)}{h(x_1)} \leq M\right\} = \frac{f(x_1)}{h(x_1)} = \frac{4\mathrm{e}\left(-1+\sqrt{5}\right)\mathrm{e}^{\sqrt{5}}}{3+e^4}. \end{aligned}$$

The rejection sampler works as follows:

- Generate $X \sim h$, i.e. here $X \sim U(0,1)$, and $U \sim U(0,1)$.
- If $U \leq \frac{f(x)}{M_1\,h(x)}$, accept X and stop (or return to the first step to generate other samples).
- Else, reject X and return to the first step.

Suppose now that we use $g_{1,4}$ instead of h. We calculate

$$\frac{f(x)}{g_{1,4}(x)} = \frac{32\,\frac{x(1-x)\,\mathrm{e}^{-4(x-1)}}{3+\mathrm{e}^4}}{\frac{4\mathrm{e}^{-4x}}{1-\mathrm{e}^{-4}}} = 8\,\frac{1-\mathrm{e}^{-4}}{3+\mathrm{e}^4}\,\frac{x(1-x)\,\mathrm{e}^{-4(x-1)}}{\mathrm{e}^{-4x}} = 8\,\frac{\mathrm{e}^4-1}{\mathrm{e}^4+3}\,x(1-x),$$

which is maximized at $x_2 = \frac{1}{2}$. Given that

$$M_1 = \frac{f(x_1)}{h(x_1)} > \frac{f(x_2)}{g_{1,4}(x_2)} =: M_2,$$

rejection sampling is more efficient using $g_{1,4}$ – if the constant M in the rejection sampling algorithm is chosen optimally in both cases, using $g_{1,4}$ leads to a higher acceptance probability.

Question 2. 1. The MH algorithm is given in the lecture notes. The proposal density associated with the proposal

$$y = \sqrt{1-\beta^2}\,x + \beta\,w, \qquad w \sim \mathcal{N}(0,1),$$

is given by

$$q(y|x) = \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(\frac{(y - \sqrt{1-\beta^2}\,x)^2}{2\beta^2}\right)$$

The acceptance probability for sampling from π is defined by

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

Assuming that the chain has state X_n at time n, X_{n+1} is obtained as follows:

- Generate $Y \sim q(\cdot|X_n)$ and $u \sim U(0,1)$;
- If $u < \alpha(X_n, Y)$, set $X_{n+1} = Y$, else set $X_{n+1} = X_n$.

Denoting by $\{X_n\}_{n=1,2,\ldots}$ the Markov chain generated by the MH algorithm, we can define an estimator for $I = \mathbb{E}_{Z\sim\pi}(f(Z))$ by

$$\hat{I}_n = \frac{1}{n}\sum_{i=1}^{n} f(X_n).$$

We say that family $\{\hat{I}_n\}_{n\in\mathbb{N}}$ of estimators for I is

- Unbiased if $\mathbb{E}[I_n] = I$ for all $n \in \mathbb{N}$;
- (Weakly) consistent if $I_n \to I$ in probability as $n \to \infty$.

See Definition 2.1 in the lecture notes for more details.

The estimator $\hat{I}_n$ defined above is in general biased, but it can be unbiased if the Markov chain is started at the right probability measure, i.e. if $X_0 \sim \pi$.

To show that $\hat{I}_n$ is strongly consistent, and thus also weakly consistent, we would like to employ Theorem 3.6 from the lecture notes:

- We are told that $\pi(x)$ is a smooth positive density, so it is in particular bounded and positive on every compact domain of $\mathbb{R}$.
- We would like to show that there exist positive numbers δ and ε such that

  $$q(y|x) > \varepsilon \qquad \text{if } |x - y| < \delta,$$

  but this is not the case. Indeed if $\beta \in (0,1)$.

$$q(x|x) = \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{(x - \sqrt{1-\beta^2}\, x)^2}{2\beta^2}\right) \to 0 \qquad \text{as } x \to \infty.$$

It thus seems that we cannot apply Theorem 3.6 directly. Since we haven't seen more advanced ergodicity results in class this year, you are not expected to be able to answer this part of the exercise. Note that there is a typo in the mock exam here, because the question allows $\beta = 0$, in which case the estimator is clearly not consistent.

2. We now consider another MCMC scheme for sampling from π, where a new state y is obtained from a current state x by

$$y = x + \sqrt{\delta}\, w, \qquad w \sim \mathcal{N}(0,1),$$

and the acceptance probability is given by

$$\alpha(x, y) = \frac{\pi(y)}{\pi(x) + \pi(y)}.$$

We first calculate the transition function $p(x,y) =$ "$\mathbb{P}[X_{n+1} = y | X_n = x]$". We add quotation marks here because, in fact, the transition probability measure is not absolutely continuous with respect to the Lebesgue measure, so it does not really make sense to consider its density. Strictly speaking, we should view $p(x,\cdot)$ as a probability measure on $\mathbb{R}$.

Employing the same reasoning as in the lecture notes, we obtain that the probability that a proposal from x is accepted is given by

$$\int_{\mathbb{R}} q(y|x)\,\alpha(x,y)\,\mathrm{d}y.$$

Therefore, we find that for any set $B \in \mathcal{B}(\mathbb{R})$, where $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra on $\mathbb{R}$,

$$p(x,B) = \int_B q(y|x)\,\alpha(x,y)\,\mathrm{d}y + \left(1 - \int_{\mathbb{R}} q(y|x)\,\alpha(x,y)\,\mathrm{d}y\right)\delta_x(B), \tag{1}$$

where δ_x is a Dirac measure. Another suitable notation to write this is

$$p(x,\cdot) = q(\cdot|x)\,\alpha(x,\cdot) + \left(1 - \int_{\mathbb{R}} q(y|x)\,\alpha(x,y)\,\mathrm{d}y\right)\delta_x,$$

For this equation to make sense as an equality of measures, we interpret the first term in the right-hand side as the measure induced by the function $q(\cdot|x)\,\alpha(x,\cdot)$. Indeed, remember that any function $f \in L^1(\mathbb{R})$ (or even L^1_{loc}, but don't worry if you haven't seen this notation before), induces a measure μ_f by

$$\mu_f(B) = \int_B f(x)\,\mathrm{d}x.$$

Let us emphasize that these comments are mostly for your information; as mentioned in the revision class, the MCMC question at the exam will focus on discrete state spaces, as do the lecture notes for the most part.

A distribution π is reversible for the Markov chain if

$$\int_A \pi(x)\,p(x,B)\,\mathrm{d}x = \int_B \pi(y)\,p(y,A)\,\mathrm{d}y \qquad \forall A,B \in \mathcal{B}(\mathbb{R}).$$

Employing the expression of $p(x,\cdot)$ that we found in (1), the left-hand side is

$$\begin{aligned}
LHS &= \int_A\int_B \pi(x)q(y|x)\,\alpha(x,y)\,\mathrm{d}y\,\mathrm{d}x + \int_A \pi(x)\left(1 - \int_{\mathbb{R}} q(y|x)\,\alpha(x,y)\,\mathrm{d}y\right)\delta_x(B)\,\mathrm{d}x, \\
&= \int_A\int_B \pi(x)q(y|x)\,\frac{\pi(y)}{\pi(x)+\pi(y)}\,\mathrm{d}y\,\mathrm{d}x + \int_{A\cap B}\left(1 - \int_{\mathbb{R}} q(y|x)\,\frac{\pi(y)}{\pi(x)+\pi(y)}\,\mathrm{d}y\right)\pi(x)\mathrm{d}x, \\
&= \int_A\int_B q(y|x)\,\frac{\pi(x)\pi(y)}{\pi(x)+\pi(y)}\,\mathrm{d}y\,\mathrm{d}x + \int_{A\cap B}\pi(x)\mathrm{d}x - \int_{A\cap B}\int_{\mathbb{R}} q(y|x)\,\frac{\pi(x)\,\pi(y)}{\pi(x)+\pi(y)}\,\mathrm{d}y\,\mathrm{d}x
\end{aligned}$$

Since $q(x|y) = q(y|x)$, this expression is invariant upon swapping A and B, and so (easy to check) the RHS can be developed simiarly to obtain the same expression.

Question 3. 1. A continuous time Gaussian process is defined in Definition 4.8 of the lecture notes.

2. Strict and weak stationarity are defined in Definitions 4.9 and 4.10, respectively.

3. If $\{X_t\}$ is a Gaussian process with mean function $\mu(t)$ and covariance function $C(s,t)$, then

$$\begin{pmatrix} X_{t_1} \\ \vdots \\ X_{t_N} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} m(t_1) \\ \vdots \\ m(t_N) \end{pmatrix}, \begin{pmatrix} C(t_1,t_1) & \dots & C(t_1,t_N) \\ \vdots & & \vdots \\ C(t_N,t_1) & \dots & C(t_N,t_N) \end{pmatrix} \right) =: \mathcal{N}(m,\Sigma),$$

where m and Σ can be calculated explicitly. We can generate a $\mathcal{N}(m,\Sigma)$ random vector from a vector $Y=(Y_1,\dots,Y_N)^T$ of IID $\mathcal{N}(0,1)$ random variables by using Lemma 2.4 in the lecture notes. If C denotes a solution of $CC^T=\Sigma$, which can be obtained e.g. by Cholesky decomposition, it holds that

$$X = m + CY \sim \mathcal{N}(m,\Sigma).$$

4. Suppose that

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right) =: \mathcal{N}(m,\Sigma),$$

The conditional distribution of Z_2 conditional on Z_1 is given by

$$\begin{aligned} f_{Z_2|Z_1}(z_2|z_1) &= \frac{f_{Z_1,Z_2}(z_1,z_2)}{\int_{\mathbb{R}} f_{Z_1,Z_2}(z_1,z_2)\,\mathrm{d}z_2} \\ &= \frac{1}{Z(z_1)} \exp\left(-\frac{1}{2} \begin{pmatrix} z_1-m_1 \\ z_2-m_2 \end{pmatrix}^T \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}^{-1} \begin{pmatrix} z_1-m_1 \\ z_2-m_2 \end{pmatrix} \right) \end{aligned}$$

where $Z(z_1)$ is the normalization constant. Using the Schur's complement or Cramer's formula, we calculate

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}^{-1} = \frac{1}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} \begin{pmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{pmatrix},$$

so

$$\begin{aligned} (z-m)^T\Sigma(z-m) &= \frac{1}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} \begin{pmatrix} z_1-m_1 \\ z_2-m_2 \end{pmatrix}^T \begin{pmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{pmatrix} \begin{pmatrix} z_1-m_1 \\ z_2-m_2 \end{pmatrix} \\ &= \frac{1}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} \left(\sigma_{11}\, z_2\, z_2 - 2\sigma_{11}\, m_2\, z_2 - 2\sigma_{12} z_2 (z_1-m_1) \right) + C(z_1) \\ &= \frac{\sigma_{11}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} \left(z_2 - \left(m_2 + \frac{\sigma_{12}}{\sigma_{11}}(z_1-m_1) \right) \right)^2 + C(z_1). \end{aligned}$$

where $C(z_1)$ is a (changing) constant independent of z_2. This shows that

$$f_{Z_2|Z_1}(z_2|z_1) \propto \exp\left(-\frac{1}{2\gamma(z_1)} (z_2-\mu(z_1))^2 \right),$$

where $\mu(z_1) = m_2 + \frac{\sigma_{12}}{\sigma_{11}}(z_1-m_1)$ and $\gamma(z_1) = \frac{\sigma_{11}\sigma_{22}-\sigma_{12}^2}{\sigma_{11}}$.

Suppose now that we have already generated $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ and that $\{X_t\}$ is a Markov process. Since $\{X_t\}$ is a Markov process,

$$\mathbb{E}[X_{t_{n+1}}|X_{t_1},\dots,X_{t_n}] = \mathbb{E}[X_{t_{n+1}}|X_{t_n}]$$

and

$$\mathbb{V}[X_{t_{n+1}}|X_{t_1},\dots,X_{t_n}] = \mathbb{V}[X_{t_{n+1}}|X_{t_n}].$$

By the formulas found above, we can generate $X_{t_{n+1}}$ by

$$X_{t_{n+1}} = m(t_{n+1}) + \frac{\sigma_{12}}{\sigma_{11}}(X_{t_n} - m(t_n)) + \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}\,\xi, \qquad \xi \sim \mathcal{N}(0,1).$$

If $m(t) = 0$ and $C(s,t) = \mathrm{e}^{-\frac{\alpha}{2}|t-s|}$, this formula reads

$$X_{t_{n+1}} = m(t_{n+1}) + \mathrm{e}^{-\frac{\alpha}{2}|t_{n+1}-t_n|}\, X_{t_n} + \sqrt{1 - \mathrm{e}^{-\alpha|t_{n+1}-t_n|}}\,\xi, \qquad \xi \sim \mathcal{N}(0,1),$$

as required.

Question 4. We consider the SDE

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \qquad X_0 = x_0.$$

1. (i) Written in integral form, the SDE reads

$$X_t - X_0 = \int_0^t b(X_s)\,\mathrm{d}s + \int_0^t \sigma(X_s)\,\mathrm{d}W_s,$$

so in particular,

$$X_{t_{n+1}} - X_{t_n} = (X_{t_{n+1}} - X_0) - (X_{t_n} - X_0) = \int_{t_n}^{t_{n+1}} b(X_s)\,\mathrm{d}s + \int_{t_n}^{t_{n+1}} \sigma(X_s)\,\mathrm{d}W_s.$$

Approximating the integrals as

$$\int_{t_n}^{t_{n+1}} b(X_s)\,\mathrm{d}s \approx b(X_{t_n})\,\Delta t$$
$$\int_{t_n}^{t_{n+1}} \sigma(X_s)\,\mathrm{d}W_s \approx b(X_{t_n})\,\Delta W_n,$$

and noting that $\Delta W_n \sim \mathcal{N}(0,\Delta t)$ we define the Euler-Maruyama scheme as

$$X_{n+1} = X_n + b(X_n)\,\Delta t + \sigma(X_n)\,\sqrt{\Delta t}\,\xi, \qquad \xi \sim \mathcal{N}(0,1).$$

(ii) See 4.10.1 and 4.10.2 in the lecture notes. In class we saw that the strong and weak errors can be defined in different ways so, should these concepts come up at the exam, we will be very precise about what we expect.

(iii) The EM scheme has strong and weak orders of convergence equal to 1/2 and 1, respectively.

2. See the lecture notes, or exercise 4 in the 2016 exam.

3. We consider the θ-Euler method,

$$X_{n+1} = X_n + [(1-\theta)b(X_n) + \theta\, b(X_{n+1})]\Delta t + \sigma(X_n)\,\sqrt{\Delta t}\,\xi, \qquad \xi \sim \mathcal{N}(0,1).$$

To calculate $\mathbb{E}|X_t|^2$ for scalar geometric Brownian motion, let us use a different approach from the one used in class. Letting $Y_t = |X_t|^2$ and using Itô's formula, we have

$$\begin{aligned}\mathrm{d}Y_t &= \text{``}2X_t\,\mathrm{d}X_t\text{''} + |\sigma X_t|^2\,\mathrm{d}t = 2\lambda|X_t|^2\,\mathrm{d}t + 2\sigma|X_t|^2\,\mathrm{d}W_t + \sigma^2|X_t|^2\,\mathrm{d}t \\ &= (2\lambda+\sigma^2)Y_s\,\mathrm{d}s + 2\sigma\,Y_s\,\mathrm{d}W_s.\end{aligned}$$

In integral form, this is

$$Y_t - Y_0 = \int_0^t (2\lambda+\sigma^2)Y_s\,\mathrm{d}s + \int_0^t 2\sigma\,Y_s\,\mathrm{d}W_s,$$

Letting $f(t) = \mathbb{E}[Y_t]$ and differentiating the previous equation, we obtain a differential equation for f:

$$f'(t) = (2\lambda+\sigma^2)f(t) \Rightarrow f(t) = f(0)\,\mathrm{e}^{(2\lambda+\sigma^2)t}.$$

In order for this function to converge to 0 as $t \to \infty$, it is necessary and sufficient that

$$2\lambda + \sigma^2 < 0.$$

The update formula for the θ Euler method reads, in the case of scalar gBM,

$$(1-\lambda\,\theta\,\Delta t)X_{n+1} = X_n\left(1 + \lambda(1-\theta)\Delta t + \sigma\sqrt{\Delta t}\,\xi\right),$$

that is, assuming $1 - \lambda\,\theta\,\Delta t \neq 0$,

$$X_{n+1} = X_n\left(\frac{1+\lambda(1-\theta)\Delta t + \sigma\sqrt{\Delta t}\,\xi}{1-\lambda\,\theta\,\Delta t}\right).$$

Therefore

$$\begin{aligned}\mathbb{E}|X_{n+1}|^2 &= \mathbb{E}|X_n|^2\mathbb{E}\left(\frac{1+\lambda(1-\theta)\Delta t + \sigma\sqrt{\Delta t}\,\xi}{1-\lambda\,\theta\,\Delta t}\right)^2 \\ &= \mathbb{E}|X_n|^2\,\frac{|1+\lambda(1-\theta)\Delta t|^2 + \sigma^2\,\Delta t}{|1-\lambda\,\theta\,\Delta t|^2}.\end{aligned}$$

The discrete time approximation is mean-square stable if and only if

$$\begin{aligned}&\frac{|1+\lambda(1-\theta)\Delta t|^2 + \sigma^2\,\Delta t}{|1-\lambda\,\theta\,\Delta t|^2} < 1 \\ &\Leftrightarrow |1+\lambda(1-\theta)\Delta t|^2 - |1-\lambda\,\theta\,\Delta t|^2 + \sigma^2\,\Delta t < 0 \\ &\Leftrightarrow \Delta t(2\lambda+\sigma^2+\Delta t(1-2\theta)\lambda^2) < 0 \\ &\Leftrightarrow 2\lambda+\sigma^2+\Delta t(1-2\theta)\lambda^2 < 0.\end{aligned}$$

When $\theta = \frac{1}{2}$, this condition becomes $2\lambda+\sigma^2 < 0$, which is the mean-square stability condition for the underlying equation: the stability region of the numerical solution coincides with that of the continuous solution.