Computational Stochastic Processes
Problem Sheet 2

U. Vaes
with problems from G. Pavliotis, A. Duncan and S. Gomes

March 2020

You are free to return a selection of your work to me for marking. This is entirely optional and the mark will not count for assessment.

1 Stochastic differential equations

Problem 1 (Weak error)**.** Let us consider the weak Euler–Maruyama update defined by

$$X_{n+1}^{\Delta t} = X_n^{\Delta t} + b(X_n^{\Delta t})\,\Delta t + \sigma(X_n^{\Delta t})\,\sqrt{\Delta t}\,\xi_n,$$

where $\{\xi_n\}_{n=0}^{N-1}$ are i.i.d. discrete-valued random variables taking values 1 and -1 with equal probability. Show that the weak error, for geometric Brownian motion and for the observables $f(x) = x$ and $f(x) = x^2$, scales as Δt, i.e. show that

$$\left|\mathbb{E}\left[f(X_{N\Delta t}) - f(X_N^{\Delta t})\right]\right| \leq C\Delta t,$$

for a constant C independent of Δt.

Solution. *We use the following parametrization of the geometric Brownian motion:*

$$\mathrm{d}X_t = \mu X_t\,\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t, \qquad X_0 = x_0, \tag{1}$$

where μ and σ are constants and x_0 is a random initial condition. The solution to this SDE (which we found earlier in the course by applying Itô's formula to the function $\ln X_t$) is

$$X_t = x_0\,\exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right).$$

This can be written as

$$X_T = x_0 \prod_{n=0}^{N-1} \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)\,\Delta t + \sigma\,\Delta W_n\right),$$

where $\Delta W_n := W_{(n+1)\Delta t} - W_{n\Delta t}$. On the other hand, the weak Euler-Maruyama discretization of (1), denoted by $X_n^{\Delta t}$, is given by

$$X_{n+1}^{\Delta t} = X_n^{\Delta t} + \mu\,X_n^{\Delta t}\,\Delta t + \sigma\,X_n^{\Delta t}\,\sqrt{\Delta t}\,\xi_n = (1 + \mu\,\Delta t + \sigma\,\sqrt{\Delta t}\,\xi_n)\,X_n^{\Delta t},$$

It follows that

$$X_N^{\Delta t} = x_0 \prod_{n=0}^{N-1} \left(1 + \mu\,\Delta t + \sigma\,\sqrt{\Delta t}\,\xi_n\right).$$

We will prove that the weak error is bounded from above by $C_m\,\Delta t$ for the observables $f(x) = x^m$, where $m \in \mathbb{N}$ and C denotes any constant independent of Δt. Since all the factors in the expressions of X_t and $X_N^{\Delta t}$ are independent,

$$\begin{aligned}
\mathbb{E}\left[(X_T)^m\right] &= \mathbb{E}\left[(x_0)^m\right] \prod_{n=0}^{N-1} \mathbb{E}\left[\exp\left(m\,\left(\mu - \frac{\sigma^2}{2}\right)\,\Delta t + m\,\sigma\,\Delta W_n\right)\right], \\
\mathbb{E}\left[(X_N^{\Delta t})^m\right] &= \mathbb{E}\left[(x_0)^m\right] \prod_{n=0}^{N-1} \mathbb{E}\left[(1 + \mu\,\Delta t + \sigma\,\sqrt{\Delta t}\,\xi_n)^m\right].
\end{aligned}$$

By the binomial theorem, here applied twice,

$$\begin{aligned}
\mathbb{E}\left[(1 + \mu\,\Delta t + \sigma\,\sqrt{\Delta t}\,\xi_n)^m\right] &= \mathbb{E}\left[\sum_{k=0}^{m} \binom{m}{k} 1^{m-k}\,(\mu\,\Delta t + \sigma\,\sqrt{\Delta t}\,\xi_n)^k\right] \\
&= \mathbb{E}\left[\sum_{k=0}^{m} \binom{m}{k} \sum_{\ell=0}^{k} \binom{k}{\ell} (\mu\,\Delta t)^{k-\ell}\,(\sigma\,\sqrt{\Delta t}\,\xi_n)^{\ell}\right] \\
&= \sum_{k=0}^{m} \binom{m}{k} \sum_{\ell=0}^{k} \binom{k}{\ell} (\mu\,\Delta t)^{k-\ell}\,(\sigma\,\sqrt{\Delta t})^{\ell}\,\mathbb{E}[(\xi_n)^{\ell}] \\
&= \sum_{k=0}^{m} \binom{m}{k} \sum_{\ell=0}^{k} \binom{k}{\ell} (\mu\,\Delta t)^{k-\ell}\,(\sigma\,\sqrt{\Delta t})^{\ell}\,\frac{(-1)^{\ell} + 1^{\ell}}{2} =: A(\Delta t).
\end{aligned}$$

On the other hand, by the properties of the lognormal distribution,

$$\mathbb{E}\left[\exp\left(m\,\left(\mu - \frac{\sigma^2}{2}\right)\,\Delta t + m\,\sigma\,\Delta W_n\right)\right] = \exp\left(\left(m\mu - \frac{m\sigma^2}{2} + \frac{m^2\sigma^2}{2}\right)\,\Delta t\right) =: B(\Delta t).$$

We will break down the rest of the proof in 3 steps.

- ***Step 1**: Show that $|A(\Delta t) - B(\Delta t)| \leq C\,\Delta t^2$. From the definition of $A(\Delta t)$, by calculating explicitly the terms of the sum corresponding to $k = 0, 1, 2$, we obtain that*

$$\begin{aligned}
A(\Delta t) = &1 + m(\mu\Delta t) + \binom{m}{2}\,((\mu\Delta t)^2 + \sigma^2\,\Delta t) \\
&+ \sum_{k=3}^{m} \binom{m}{k} \sum_{\ell=0}^{k} \binom{k}{\ell} (\mu\,\Delta t)^{k-\ell}\,(\sigma\,\sqrt{\Delta t})^{\ell}\,\frac{(-1)^{\ell} + 1^{\ell}}{2} \\
= &1 + m(\mu\Delta t) + \frac{m(m-1)}{2}\,(\sigma^2\,\Delta t) + \mathcal{O}(\Delta t^2) \\
= &1 + \left(m\mu - \frac{m\sigma^2}{2} + \frac{m^2\sigma^2}{2}\right)\,\Delta t + \mathcal{O}(\Delta t^2).
\end{aligned}$$

On the other hand, by a Taylor expansion,

$$B(\Delta t) = 1 + \left(m\mu - \frac{m\sigma^2}{2} + \frac{m^2\sigma^2}{2}\right)\,\Delta t + \mathcal{O}(\Delta t^2),$$

so we conclude that $|A(\Delta t) - B(\Delta t)| = \mathcal{O}(\Delta t^2)$.

- ***Step 2:*** *Show that* $|A(\Delta t)^N - B(\Delta t)^N| = \mathcal{O}(\Delta t)$. *Employing the binomial theorem again, and noticing that the term corresponding to* $k = 0$ *cancels out with* $B(\Delta t)^N$, *we obtain*

$$\begin{aligned}
\left|A(\Delta t)^N - B(\Delta t)^N\right| &= \left|\Big(\big(A(\Delta t) - B(\Delta t)\big) + B(\Delta t)\Big)^N - B(\Delta t)^N\right| \\
&= \left|\sum_{k=1}^{N} \binom{N}{k} \big(A(\Delta t) - B(\Delta t)\big)^k B(\Delta t)^{N-k}\right| \\
&\leq \sum_{k=1}^{N} \binom{N}{k} \left|\big(A(\Delta t) - B(\Delta t)\big)^k B(\Delta t)^{N-k}\right| \\
&\leq \sum_{k=1}^{N} \binom{N}{k} (C\,\Delta t^2)^k \left|B(\Delta t)^{N-k}\right|.
\end{aligned}$$

In order to be able to conclude this step, we note that, for $k = 1, \dots, N$,

$$|B(\Delta t)|^{N-k} \leq \exp\left(\left|m\mu - \frac{m\sigma^2}{2} + \frac{m^2\sigma^2}{2}\right| T\right),$$

by definition of $B(\Delta t)$. *Using the inequality* $\binom{N}{k} \leq N^k$, *we finally obtain*

$$\begin{aligned}
\left|A(\Delta t)^N - B(\Delta t)^N\right| \leq C \sum_{k=1}^{N} N^k\, \Delta t^{2k} &= C \sum_{k=1}^{N} T^k\, \Delta t^k \\
&= C\,T\,\Delta t \left(1 + \sum_{k=1}^{N} (T\,\Delta t)^k\right) = \mathcal{O}(\Delta t),
\end{aligned}$$

because T *is fixed and the term in the round brackets is bounded from above as* $\Delta t \to 0$.

- ***Step 3:*** *Show that* $\left|\mathbb{E}\left[(X_N^{\Delta t})^m - (X_T)^m\right]\right| = \mathcal{O}(\Delta t)$. *This follows from the previous step after noticing that* $\mathbb{E}[(X_N^{\Delta t})^m] = \mathbb{E}[(x_0)^m] A(\Delta t)^N$ *and* $\mathbb{E}[(X_T)^m] = \mathbb{E}[(x_0)^m] B(\Delta t)^N$:

$$\left|\mathbb{E}\left[(X_N^{\Delta t})^m - (X_T)^m\right]\right| \leq \left|\mathbb{E}\left[(x_0)^m\right]\right| \left|A(\Delta t)^N - B(\Delta t)^N\right| = \mathcal{O}(\Delta t).$$

Problem 2 (Variance reduction)**.** Consider the *overdamped Langevin* equation

$$\mathrm{d}X_t = -V'(X_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t, \qquad X_0 = -1, \tag{2}$$

where $V(\cdot)$ is the double well potential:

$$V(x) = \frac{x^4}{4} - \frac{x^2}{2}.$$

1. By using a Monte Carlo simulation wth the Euler–Maruyama method, estimate the probability P defined by

$$P := \mathbb{P}[X_T > 0], \qquad T = 1.$$

Solution. *Let us denote by* $\{X_k^{\Delta t}\}_{k=0}^N$ *the Euler–Maruyama discretization of* (2):

$$X_{k+1}^{\Delta t} = X_k^{\Delta t} - V'(X_k^{\Delta t})\,\Delta t + \sqrt{2\beta^{-1}}\,\Delta W_k, \qquad X_0^{\Delta t} = -1,$$

where, as usual, $\Delta W_k = W_{(k+1)\Delta t} - W_{k\Delta t}$*. To estimate* P*, we will use the approximation*

$$\mathbb{P}[X_T > 0] \approx \mathbb{P}[X_N^{\Delta t} > 0] \tag{3}$$

and estimate the right-hand side by Monte Carlo simulation:

$$I_J = \frac{1}{J}\sum_{j=1}^{J} I_{(0,\infty)}(X_N^{\Delta t,j}),$$

where $X^{\Delta t,j}$*, with* $j = 1, \dots, J$*, are replicas of the Euler–Maruyama solution. See the Jupyter notebook for the numerical estimation.*

2. By using importance sampling, implement an estimator for P with a lower variance.

Solution. *There is not a unique solution for this question. The general idea is to add a drift term in (2),*

$$\mathrm{d}Y_t = -V'(Y_t)\,\mathrm{d}t + b(Y_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t, \qquad X_0 = -1, \tag{4}$$

in such a way such a way that $\mathbb{P}[Y_T > 0]$ *is higher than* $\mathbb{P}[X_T > 0]$*. A simple way to achieve this is to choose* $b(\cdot)$ *to be the function*

$$b(y) = \mathbb{1}_{(-\infty,0)}\, b,$$

for some positive constant b*. Using an Euler–Maruyama discretization for* Y_t*, we define a new discrete-time process* $\{Y_k^{\Delta t}\}_{k=0}^N$*:*

$$Y_{k+1}^{\Delta t} = Y_k^{\Delta t} - V'(Y_k^{\Delta t})\,\Delta t + b(Y_n^{\Delta t})\,\Delta t + \sqrt{2\beta^{-1}}\,\Delta W_k, \qquad Y_0^{\Delta t} = -1. \tag{5}$$

This new process defines a probability measure over $\mathbb{R}^N$*, , the law of* $\{Y_k^{\Delta t}\}_{k=1}^N$*, which can be employed to estimate* $\mathbb{P}[X_N^{\Delta t} > 0]$ *via importance sampling. With this approach, based on employing the approximation (3) from the beginning, we do not need to worry about Girsanov's theorem or measures over an infinite-dimensional state space.*

The probability density functions of $\{X_k^{\Delta t}\}_{k=1}^N$ *and* $\{Y_k^{\Delta t}\}_{k=1}^N$ *are*

$$\rho_X(x_1, \dots, x_N) = \frac{1}{(2\beta^{-1}\Delta t)^{N/2}} \exp\left(-\frac{1}{4\beta^{-1}\Delta t}\sum_{k=0}^{N-1}(x_{k+1} - x_k + V'(x_k)\,\Delta t)^2\right),$$
$$\rho_Y(x_1, \dots, x_N) = \frac{1}{(2\beta^{-1}\Delta t)^{N/2}} \exp\left(-\frac{1}{4\beta^{-1}\Delta t}\sum_{k=0}^{N-1}(x_{k+1} - x_k + V'(x_k)\,\Delta t - b(x_k)\,\Delta t)^2\right),$$

with $x_0 := X_0$*. Their ratio, which coincides with the Radon–Nikodym derivative of* ρ_X *(or, more precisely, of the probability measure with density* ρ_X*) with respect to* ρ_Y*, is given by*

$$\frac{\rho_X(x_1, \dots, x_N)}{\rho_Y(x_1, \dots, x_N)} = \exp\left(-\frac{1}{2\beta^{-1}}\left(\sum_{k=0}^{N-1}(x_{k+1} - x_k + V'(x_k)\,\Delta t)\,b(x_k) - \frac{1}{2}\sum_{k=0}^{N-1}|b(x_k)|^2\,\Delta t\right)\right).$$

Evaluated at $(Y_1^{\Delta t}, \dots, Y_N^{\Delta t})$ *and using (5), this ratio, is equal almost surely to*

$$\frac{\rho_X(Y_1^{\Delta t}, \dots, Y_N^{\Delta t})}{\rho_Y(Y_1^{\Delta t}, \dots, Y_N^{\Delta t})} = \exp\left(-\frac{1}{2\beta^{-1}}\left(\sqrt{2\beta^{-1}}\sum_{k=0}^{N-1} b(Y_k)\Delta W_k + \frac{1}{2}\sum_{k=0}^{N-1}|b(Y_k)|^2\Delta t\right)\right).$$

Now, introducing on $\mathbb{R}^N$ the indicator function $f(\cdot) := I_{\mathbb{R}\times\mathbb{R}\times\cdots\times\mathbb{R}_{>0}}(\cdot)$, we estimate the desired probability as follows:

$$\begin{aligned}
\mathbb{P}[X_T > 0] \approx \mathbb{P}[X_N^{\Delta t} > 0] &= \mathbb{E}[f(X_1^{\Delta t}, \dots, X_N^{\Delta t})] \\
&= \int_{\mathbb{R}^N} f(x_1, \dots, x_N)\, \rho_X(x_1, \dots, x_N)\, \mathrm{d}x_1 \dots \mathrm{d}x_N \\
&= \int_{\mathbb{R}^N} f(x_1, \dots, x_N)\, \frac{\rho_X(x_1, \dots, x_N)}{\rho_Y(x_1, \dots, x_N)}\, \rho_Y(x_1, \dots, x_N)\, \mathrm{d}x_1 \dots \mathrm{d}x_N \\
&= \mathbb{E}\left[f(Y_1^{\Delta t}, \dots, Y_N^{\Delta t})\, \frac{\rho_X(Y_1^{\Delta t}, \dots, Y_N^{\Delta t})}{\rho_Y(Y_1^{\Delta t}, \dots, Y_N^{\Delta t})} \right].
\end{aligned}$$

The last expression can then be approximated numerically by using a standard Monte Carlo method with the modified discrete-time dynamics. An alternative, conceptually more difficult manner of deriving the same estimate is by calculating the likelihood ratio between the laws of the continuous-time *processes using Girsanov's theorem, and by then discretizing in a later step so that it can be estimated numerically.*

By Girsanov's theorem, the law of $\{Y_t\}_{t\in[0,T]}$, viewed as a probability measure over the space of continuous functions, is equivalent (i.e. one is absolutely continuous with respect to the other, and conversely) to that of $\{U_t\}_{t\in[0,T]} := \{\sqrt{2\beta^{-1}} W_t\}_{t\in[0,T]}$, which is itself equivalent to the law of $\{X_t\}_{t\in[0,T]}$. Denoting these probability measures by μ_Y, μ_U and μ_X, the Radon–Nikodym derivatives are given by the following expressions: evaluated at an arbitrary continuous function $\{Z_t\}_{t\in[0,t]}$,

$$\begin{aligned}
\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu_U}(\{Z_t\}_{t\in[0,t]}) &= \exp\left(\frac{1}{2\beta^{-1}} \left(\int_0^T (-V'(Y_s))\, \mathrm{d}Z_s - \frac{1}{2} \int_0^T |-V'(Z_s)|^2\, \mathrm{d}s) \right) \right), \\
\frac{\mathrm{d}\mu_Y}{\mathrm{d}\mu_U}(\{Z_t\}_{t\in[0,t]}) &= \exp\left(\frac{1}{2\beta^{-1}} \left(\int_0^T (-V'(Z_s) + b(Z_s))\, \mathrm{d}Z_s - \frac{1}{2} \int_0^T |-V'(Z_s) + b(Z_s)|^2\, \mathrm{d}s \right) \right),
\end{aligned}$$

To keep the notation concise, we will abbreviate $\{Z_t\} := \{Z_t\}_{t\in[0,t]}$. Here, the Radon–Nikodym derivatives are viewed as functionals over the space of continuous functions: they take continuous functions as arguments and return a real number. We emphasize that, although convenient for formal calculations, these expressions are in fact not rigorous: since Z_t is a deterministic function with potentially infinite variation, the integral with respect to $\mathrm{d}Z_t$ is not well-defined. However, as we shall see below, what we need for importance sampling is only an expression of the Radon–Nikodym derivative evaluated at a given stochastic process, which we will be able to interpret precisely as a random variable.

We recall that from the basic properties of the Radon–Nikodym derivative that, for any three equivalent measures μ, ν and λ, it holds that

$$\frac{\mathrm{d}\mu}{\mathrm{d}\lambda} = \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \frac{\mathrm{d}\nu}{\mathrm{d}\lambda} \qquad (\mu, \nu, \lambda)\text{-almost everywhere.}$$

In particular, taking $\lambda = \mu$,

$$1 = \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \quad \Rightarrow \frac{\mathrm{d}\mu}{\mathrm{d}\nu} = \left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu} \right)^{-1} \qquad (\mu, \nu) \text{ almost everywhere.}$$

Therefore, for functionals $\mathcal{F}$ on the space of continuous functions,

$$\begin{aligned}\int \mathcal{F}(\{Z_t\})\,\mathrm{d}\mu_X(\{Z_t\}) &= \int \mathcal{F}(\{Z_t\})\,\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu_U}(\{Z_t\})\,\mathrm{d}\mu_U(\{Z_t\}),\\ &= \int \mathcal{F}(\{Z_t\})\,\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu_U}(\{Z_t\})\,\frac{\mathrm{d}\mu_U}{\mathrm{d}\mu_Y}(\{Z_t\})\,\mathrm{d}\mu_Y(\{Z_t\})\\ &= \int \mathcal{F}(\{Z_t\})\,\frac{\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu_U}(\{Z_t\})}{\frac{\mathrm{d}\mu_Y}{\mathrm{d}\mu_U}(\{Z_t\})}\,\mathrm{d}\mu_Y(\{Z_t\}).\end{aligned}$$

The fraction in the last equation is equal to

$$\begin{aligned}\frac{\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu_U}(\{Z_t\})}{\frac{\mathrm{d}\mu_Y}{\mathrm{d}\mu_U}(\{Z_t\})} &= \exp\left(-\frac{1}{2\beta^{-1}}\left(\int_0^T b(Z_s)\,\mathrm{d}Z_s - \frac{1}{2}\int_0^T |b(Z_s)|^2 - 2\,b(Z_s)\,V'(Z_s)\,\mathrm{d}s\right)\right)\\ &= \frac{\mathrm{d}\mu_X}{\mathrm{d}\mu_Y}(\{Z_t\}).\end{aligned}$$

(This formula is very similar to the ratio of the finite-dimensional PDFs found above.) Evaluated at $\{Z_t\} = \{Y_t\}$, and taking into account that Y_t solves (2), this can be written as

$$\begin{aligned}\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu_Y}(\{Y_t\}) &= \exp\left(-\frac{1}{2\beta^{-1}}\left(\int_0^T b(Y_s)\,\mathrm{d}Y_s - \frac{1}{2}\int_0^T |b(Y_s)|^2 - 2\,b(Y_s)\,V'(Y_s)\,\mathrm{d}s\right)\right),\\ &= \exp\left(-\frac{1}{2\beta^{-1}}\left(\sqrt{2\beta^{-1}}\int_0^T b(Y_s)\,\mathrm{d}W_s + \frac{1}{2}\int_0^T |b(Y_s)|^2\,\mathrm{d}s\right)\right).\end{aligned}$$

In the second line, the right-hand side now contains an integral with respect to Brownian motion, so we can make sense of it rigorously as a real-valued random variable. Now let us define the indicator functional $\mathcal{F}(\cdot) = I_{\{f\in C([0,T]):f(T)>0\}}(\cdot)$ over the space of continuous functions. We have

$$\begin{aligned}\mathbb{P}[X_T > 0] = \mathbb{E}[\mathcal{F}(\{X_t\})] &= \int \mathcal{F}(\{Z_t\})\,\mathrm{d}\mu_X(\{Z_t\})\\ &= \int \mathcal{F}(\{Z_t\})\,\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu_Y}(\{Z_t\})\,\mathrm{d}\mu_Y(\{Z_t\})\\ &= \mathbb{E}\left[\mathcal{F}(\{Y_t\})\,\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu_Y}(\{Y_t\})\right].\end{aligned}$$

Now let $\{Y_t^{\Delta t}\}_{t\in[0,T]}$ be the piecewise constant continuous-time process obtained from the Euler–Maruyama discretization $\{Y_k^{\Delta t}\}$. For Δt sufficiently small, we expect the law of $\{Y_t^{\Delta t}\}$ to be close to that of $\{Y_t\}$, so in particular

$$\begin{aligned}\mathbb{E}\left[\mathcal{F}(\{Y_t\})\,\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu_Y}(\{Y_t\})\right] &\approx \mathbb{E}\left[\mathcal{F}(\{Y_t^{\Delta t}\})\,\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu_Y}(\{Y_t^{\Delta t}\})\right]\\ &= \mathbb{E}\left[f(Y_1^{\Delta t},\dots,Y_N^{\Delta t})\,\frac{\rho_X(Y_1^{\Delta t},\dots,Y_N^{\Delta t})}{\rho_Y(Y_1^{\Delta t},\dots,Y_N^{\Delta t})}\right],\end{aligned}$$

leading us to the same estimator as the one found before. The choice of the constant b is not evident, and some numerical exploration can help us find a good value. If b is chosen very small, then the probability $\mathbb{P}[Y_T > 0]$ will be small and we don't expect a significant variance reduction. On the other hand, if b is chosen very large, then the probability $\mathbb{P}[Y_T > 0]$ will be close to one, but we expect the likelihood ratio to take a very small value with high *probability, thus also leading to a large variance. Numerically, we found $b = 1.5$ to be a good value, yielding a variance reduction by a factor close to 9.*

The presentation of Girsanov's theorem that we gave in the course and in the Jupyter notebook is based on Section 3.6 of Prof. Pavliotis's book, Stochastic Processes and Applications*. For a more detailed and fully rigorous presentation of Girsanov's theorem, see Section 8.6 of Oksendal's book entitled* Stochastic Differential Equations*.*

Problem 3 (Maximum Likelihood estimator)**.** Consider the SDE

$$\mathrm{d}X_t = (\alpha X_t - \beta X_t^3)\,\mathrm{d}t + \mathrm{d}W_t.$$

Our objective is to derive maximum likelihood estimators for α and β for a given observation of the path X_t, $t \in [0, T]$.

1. Show that the log of the likelihood function is

$$\log L = \alpha B_1 - \beta B_3 - \frac{1}{2}\alpha^2 M_2 - \frac{1}{2}\beta^2 M_6 + \alpha\beta M_4,$$

where

$$M_n\left(\{X_t\}_{t\in[0,T]}\right) = \int_0^T X_t^n\,\mathrm{d}t \quad \text{and} \quad B_n(\{X_t\}_{t\in[0,T]}) := \int_0^T X_t^n\,\mathrm{d}X_t.$$

Solution. *By Girsanov's theorem, the Radon–Nykodym of the law of $X := \{X_t\}_{t\in[0,T]}$ with respect to the law of Brownian motion (called the Wiener measure) is given by*

$$\begin{aligned}
\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\mathbb{P}_W}(X;\alpha,\beta) &= \exp\left(\int_0^T b(X_t;\alpha,\beta)\,\mathrm{d}X_t - \frac{1}{2}\int_0^T b(X_t;\alpha,\beta)^2\,\mathrm{d}t\right) \\
&= \exp\left(\int_0^T b(X_t;\alpha,\beta)\,\mathrm{d}W_t + \frac{1}{2}\int_0^T b(X_t;\alpha,\beta)^2\,\mathrm{d}t\right),
\end{aligned}$$

where $b(x;\alpha,\beta) = \alpha X_t - \beta X_t^3$. Taking the logarithm of the first expression (the second expression is not necessary in this exercise), we obtain

$$\begin{aligned}
\log\left(\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\mathbb{P}_W}(X;\alpha,\beta)\right) &= \int_0^T b(X_t;\alpha,\beta)\,\mathrm{d}X_t - \frac{1}{2}\int_0^T b(X_t;\alpha,\beta)^2\,\mathrm{d}t, \\
&= \alpha\,B_1(X) - \beta\,B_3(X) - \frac{1}{2}\left(\alpha^2\,M_2(X) - 2\alpha\beta\,M_4(X) + \beta^2\,M_6(X)\right) \\
&= \alpha\,B_1(X) - \beta\,B_3(X) - \frac{\alpha^2}{2}\,M_2(X) - \frac{\beta^2}{2}M_6(X) + \alpha\beta\,M_4(X).
\end{aligned}$$

2. Consequently show that the MLE for α and β are given by

$$\hat{\alpha} = \frac{B_1 M_6 - B_3 M_4}{M_2 M_6 - M_4^2} \quad \text{and} \quad \hat{\beta} = \frac{B_1 M_4 - B_3 M_2}{M_2 M_6 - M_4^2}.$$

Solution. *Equaling to zero the derivatives with respect to α and β, we obtain*

$$\begin{aligned}
0 &= B_1(X) - \alpha\,M_2(X) + \beta\,M_4(X), \\
0 &= -B_3(X) - \beta\,M_6(x) + \alpha\,M_4(X),
\end{aligned}$$

which in matrix form is

$$\begin{pmatrix} M_2(X) & -M_4(X) \\ -M_4(x) & M_6(x) \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} B_1(X) \\ -B_3(X) \end{pmatrix}.$$

Using Cramer's rule, and omitting the dependence on X for convenience, we obtain

$$\alpha = \frac{B_1\, M_6 - B_3\, M_4}{M_2\, M_6 - |M_4|^2}, \qquad \beta = \frac{B_1\, M_4 - B_3\, M_2}{M_2\, M_6 - |M_4|^2}.$$

Problem 4 (Nonlinear SDEs in population dynamics)**.** The following SDE appears in population dynamics:

$$\mathrm{d}X_t = -\mu X_t(1 - X_t)\,\mathrm{d}t - \sigma X_t(1 - X_t)\,\mathrm{d}W_t \tag{6}$$

1. Show that $X_t = 1$ is a fixed point for (6) and that linearizing about this fixed point we obtain the SDE for geometric Brownian motion.

   Solution. *If we substitute $X_t = 1$ in the SDE (6), we obtain*

   $$\mathrm{d}X_t = -\mu 1(1-1)\,\mathrm{d}t - \sigma 1(1-1)\,\mathrm{d}W_t = 0\,\mathrm{d}t + 0\,\mathrm{d}W_t,$$

   so X_t = 1 is a fixed point. We now write $X_t = 1 + \varepsilon Y_t$ and substitute in (6), which leads to

   $$\mathrm{d}(1 + \varepsilon Y_t) = -\mu(1 + \varepsilon Y_t)(1 - (1 + \varepsilon Y_t)\,\mathrm{d}t - \sigma(1 + \varepsilon Y_t)(1 - (1 + \varepsilon Y_t))\,\mathrm{d}W_t$$

   or

   $$\varepsilon\,\mathrm{d}Y_t = \varepsilon(\mu Y_t\ \mathrm{d}t + \sigma Y_t\ \mathrm{d}W_t) + \varepsilon^2(\mu Y_t^2\,\mathrm{d}t + \sigma Y_t^2\,\mathrm{d}W_t).$$

   We finally get

   $$\mathrm{d}Y_t = \mu Y_t\,\mathrm{d}t + \sigma Y_t\,\mathrm{d}W_t + \varepsilon(\mu Y_t^2\,\mathrm{d}t + \sigma Y_t^2\,\mathrm{d}W_t).$$

   Disregarding the term multiplying ε, we obtain the equation for geometric Brownian motion.

2. Solve (6) numerically using the explicit Euler scheme for $\mu = -1$, $X_0 = 1.1$ and for $\sigma =$ .5, .6, .7, .8, .9. Calculate numerically $\mathbb{E}|X_t - 1|^2$ and comment on the mean square stability of the explicit Euler scheme for the nonlinear SDE (6).

   Solution. *Note first that, during the lectures, we defined mean-square stability only for the geometric Brownian motion. It is therefore appropriate to clarify what is meant by mean-square stability in this context. For $\sigma = 0$ and $\mu < 0$, (6) admits two equilibrium points: a unstable one at $X_t = 0$ and a stable one at $X_t = 1$. In the context of this exercise, we will say that the equation is mean-square stable if*

   $$\lim_{t\to\infty} \mathbb{E}|X_t - 1|^2 = 0,$$

   and similarly for the numerical approximation. A first question to ask oneself is the following: what are the values of the parameters μ and σ such that the (continuous-time) solution to (6) is mean-square stable, regardless of the initial condition? Unfortunately, this turns out to be a difficult question. Take for example the deterministic case $\sigma = 0$. The linearized equation suggests that the nonlinear equation (6) is mean-square stable provided that $\mu < 0$, but this *is true only for a subset of initial conditions: the exact solution to*

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = -\mu X_t(1 - X_t)$$

is given by

$$X_t = \frac{1}{1 - (1 - X_0^{-1})\,\mathrm{e}^{\mu t}}.$$

When $X_0 < 0$, the solutions blow up in finite time; the basin of attraction of the stable equilibrium (still when $\sigma = 0$ and $\mu < 0$) is only $(0, +\infty)$.

The condition for mean-square stability of the Euler–Maruyama scheme, when applied to geometric Brownian motion, is the following:

$$\Delta t < -\left(\frac{2}{\mu} + \frac{\sigma^2}{\mu^2}\right)$$

If $\mu = -1$, this condition becomes

$$\Delta t < 2 - \sigma^2. \tag{7}$$

Assuming that the behavior of the nonlinear equation around the equilibrium $X_t = 1$ can be well approximated by that of the linearized equation, i.e. by geometric Brownian motion, we expect the Euler–Maruyama scheme to be mean-square stable for the nonlinear equation (6) for all the values of σ given, provided that Δt is very small. This is indeed what is observed in the Jupyter notebook. When Δt is taken to be larger, however, stability issues arise, even for values of Δt such that (7) is satisfied.

3. Solve (6) using the θ-Euler scheme with $\theta = \frac{1}{2}$. Investigate the mean square stability of this numerical scheme when applied to (6).

   Solution. *The θ-Euler scheme is*

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

   This can be rewritten as

   $$X_{n+1} - \theta b(X_{n+1})\Delta t = X_n + (1-\theta) b(X_n)\Delta t + \sigma(X_n)\Delta W_n.$$

   This is a **nonlinear** *equation for X_{n+1} which can be written as:*

   $$F(X_{n+1}) = G(X_n) \quad \Leftrightarrow \quad F(X_{n+1}) - G(X_n) = 0$$

   where $F(x) = x - \theta\, b(x)\, \Delta t$ and $G(x) = x + (1-\theta)\, \Delta t\, b(x) + \sigma(x)\Delta W$. The method is implemented in the Jupyter notebook, and it seems to be more stable that the explicit Euler–Maruyama method for the nonlinear equation.

2 Markov chain Monte Carlo

Problem 5. Read Section 3.3 in the lecture notes, and show that π_{st} and π_{pt} are reversible for the Markov chains generated by the *simulated tempering* and *parallel tempering* algorithms, respectively. For the case of *parallel tempering*, consider for simplicity the case where $N = 2$. In both cases, assume that the MCMC schemes employed with probability α_0, in the notations of the lecture notes, are such that the associated transition distributions $p_i(x, y)$ satisfy detailed balance:

$$\pi_i(x)\, p_i(x, y) = \pi_i(y)\, p_i(y, x), \qquad \pi_i \propto \exp\left(-\frac{H(x)}{T_i}\right). \tag{8}$$

Here T_i denote positive constants, called *temperatures* by analogy with physical systems, and $H(x)$ denotes a smooth confining potential – a function such that $\lim_{|x|\to\infty} H(x) = +\infty$ and $\mathrm{e}^{-H(x)/T} \in L^1(X)$ for all $T > 0$. (This second condition guarantees that $\mathrm{e}^{-H(x)/T}$ defines a probability measure, up to the normalization constant.)

Solution (Simulated tempering)**.** *We will assume in this question that X is a continuous state space, such as $\mathbb{R}$. In the case of* simulated tempering *we consider the distribution*

$$\pi_{st}(x, i) \propto c_i\, \pi_i(x), \qquad (x, i) \in X \times I,$$

where $I = \{1, \dots, N\}$ and the coefficients c_i are positive. Note that, since X is a continuous state space and I is a discrete state space, π_{st} is neither a probability distribution function *nor a* probability mass function*, but a mixture of both. We need to show* detailed balance*:*

$$\pi_{st}(x, i)\, P\big((x, i), (y, j)\big) = \pi_{st}(y, j)\, P\big((y, j), (x, i)\big) \qquad \forall (x, i), (y, j) \in (X \times I)^2,$$

where $P\big((x, i), (y, j)\big)$ denotes the transition distribution from (x, i) to (y, j). To this end, we will proceed by exhaustion of the different cases:

- *If $i = j$ and $x = y$, the condition is trivially satisfied.*
- *If $i \neq j$ and $x \neq y$, then the move is impossible: $P\big((x, i), (y, j)\big) = 0$. In this case both sides of the equality are zero and the condition is satisfied.*
- *If $i = j$ and $x \neq y$, then the move corresponds to an acceptance event $u \leq \alpha_0$: the transition density is $P\big((x, i), (y, j)\big) = \alpha_0\, p_i(x, y)$. In this case, the detailed balance condition reads*

  $$c_i\, \pi_i(x)\, \alpha_0\, p_i(x, y) =^? c_i\, \pi_i(y)\, \alpha_0\, p_i(y, x),$$

  which is satisfied by the assumption (8).
- *If $i \neq j$ and $x = y$, then the move corresponds to a rejection event $u > \alpha_0$, followed by an acceptance of a proposal j drawn from $\alpha(i, \cdot)$. The probability of this acceptance is*

  $$\beta(i, j; x) = \min\left(1, \frac{c_j\, \pi_j(x)\, \alpha(j, i)}{c_i\, \pi_i(x)\, \alpha(i, j)}\right),$$

  so we deduce

  $$P\big((x, i), (y, j)\big) = (1 - \alpha_0)\, \alpha(i, j)\, \beta(i, j; x).$$

  Substituting this in the detailed balance condition, we obtain

  $$c_i\, \pi_i(x)\, (1 - \alpha_0)\, \alpha(i, j)\, \beta(i, j; x) =^? c_j\, \pi_j(x)\, (1 - \alpha_0)\, \alpha(j, i)\, \beta(j, i; x),$$
  $$\Leftrightarrow c_i\, \pi_i(x)\, \alpha(i, j)\, \beta(i, j; x) =^? c_j\, \pi_j(x)\, \alpha(j, i)\, \beta(j, i; x).$$

Employing the expression of $\beta(\cdot,\cdot;x)$, it is easily seen that both sides are equal to

$$\min\left(c_i\,\pi_i(x)\,\alpha(i,j), c_j\,\pi_j(x)\,\alpha(j,i)\right),$$

so detailed balance is satisfied.

Solution (Parallel tempering)**.** *In this case we consider the distribution*

$$\pi_{pt}(x_1,x_2) \propto \pi_1(x_1)\,\pi_2(x_2), \qquad (x_1,x_2)\in X\times X,$$

We need to show detailed balance*:*

$$\pi_{pt}(x_1,x_2)\,P\big((x_1,x_2),(y_1,y_2)\big) = \pi_{pt}(y_1,y_2)\,P\big((y_1,y_2),(x_1,x_2)\big) \qquad \forall (x_1,x_2),(y_1,y_2)\in (X\times X)^2,$$

where $P\big((x_1,x_2),(y_1,y_2)\big)$ denotes the transition density from (x_1,x_2) to (y_1,y_2). To this end, we will again proceed by exhaustion of the different cases:

- *If $x_1 = x_2$ and $y_1 = y_2$, the condition is trivially satisfied.*
- *If $x_1 \neq y_2$ or $x_2 \neq y_1$, then the move does not correspond to a swap, so $P\big((x_1,x_2),(y_1,y_2)\big) = \alpha_0\,p(x_1,y_1)\,p(x_2\,y_2)$. In this case the detailed balance condition reads:*

  $$\pi_1(x_1)\,\pi_2(x_2)\,\alpha_0\,p(x_1,y_1)\,p(x_2,\,y_2) =^? \pi_1(y_1)\,\pi_2(y_2)\,\alpha_0\,p(y_1,x_1)\,p(y_2,\,x_2),$$

  which is satisfied by the assumption (8).
- *If $x_1 = y_2$ and $x_2 = y_1$ and $x_1 \neq x_2$ (if $x_1 = y_2$ and $x_2 = y_1$ and $x_1 = x_2$, then we are in the first case addressed above) then the move might or might not correspond to a swap. In this case*

  $$P\big((x_1,x_2),(y_1,y_2)\big) = \alpha_0\,p(x_1,y_1)\,p(x_2\,x_2) + (1-\alpha_0)\beta(x_1,x_2).$$

  The first term corresponds to the case where $u \leq \alpha_0$, and the second term is the probability of an accepted swap. Here we defined

  $$\beta(x_1,x_2) = \min\left(1, \frac{\pi_1(x_2)\,\pi_2(x_1)}{\pi_1(x_1)\,\pi_2(x_2)}\right).$$

  Substituting in the detailed balance condition, we obtain

  $$\begin{aligned}&\pi_1(x_1)\,\pi_2(x_2)\,p(x_1,y_1)\,\alpha_0\,p(x_2,\,y_2) + \pi_1(x_1)\,\pi_2(x_2)\,(1-\alpha_0)\beta(x_1,x_2)\\ =^?\;&\pi_1(y_1)\,\pi_2(y_2)\,p(y_1,x_1)\,\alpha_0\,p(y_2,\,x_2) + \pi_1(x_2)\,\pi_2(x_1)\,(1-\alpha_0)\beta(x_2,x_1).\end{aligned}$$

  The first term on the left-hand side coincides with the first term on the right-hand side, by the assumption (8). Detailed balance is therefore satisfied provided that

  $$\pi_1(x_1)\,\pi_2(x_2)\,\beta(x_1,x_2) = \pi_1(x_2)\,\pi_2(x_1)\,\beta(x_2,x_1).$$

  Employing the definition of $\beta(\cdot,\cdot)$, we see that both sides are equal to

  $$\min\left(\pi_1(x_1)\,\pi_2(x_2), \pi_1(x_2)\,\pi_2(x_1)\right),$$

  which concludes the proof.

Problem 6 (Metropolis–Hastings)**.** In this question we explore the Metropolis–Hastings algorithm in a discrete state space.

1. Suppose we wish to sample from the binomial distribution

$$p_k = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}, \quad k \in \{0, 1, \ldots, n\},$$

with parameters $n \in \mathbb{N}$ and $p \in (0,1)$. Derive an independence sampler using a uniform distribution on $0, \ldots, n$ as proposal distribution.

Solution. *Note that if we are sampling a random variable Y uniformly in the finite set $\{1, \ldots, n\}$, then the probability mass function of Y is $\mathbb{P}(Y = y) = \frac{1}{n}$, for any $y \in \{1, \ldots, n\}$. We will denote the probability mass function at y by g_y. The algorithm is the following: Given a state X_n,*

(a) Sample $Y \sim \mathcal{U}(\{1, \ldots, n\})$ and $u \sim \mathcal{U}(0,1)$.

(b) If $u < \alpha(X_n, Y)$, set $X_{n+1} = Y$, where

$$\begin{aligned}
\alpha(x,y) &= \min\left\{1, \frac{p_y\, g_x}{p_x\, g_y}\right\} = \min\left\{1, \frac{p_y}{p_x}\right\} \\
&= \min\left\{1, \frac{\frac{n!}{y!(n-y)!} p^y (1-p)^{n-y}}{\frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}}\right\} \\
&= \min\left\{1, \frac{x!(n-x)!}{y!(n-y)!} p^{y-x} (1-p)^{x-y}\right\}
\end{aligned}$$

(c) Otherwise set $X_{n+1} = X_n$.

2. The geometric probability distribution is

$$p_k = p(1-p)^{k-1}, \quad k \in \{1, 2, 3, \ldots\},$$

with parameter $p \in (0,1)$. Derive a simple symmetric random walk Metropolis Hastings algorithm to sample from this distribution.

Solution. *We will use a uniform distribution on a set of integers, centered at X_n and with width 2δ for some integer δ. This means that*

$$q(y|x) = \frac{1}{2\delta + 1} \mathbb{1}(|y - x| \leq \delta)$$

The algorithm is the following: Given a state X_n,

(a) Sample $Z \sim \mathcal{U}(\{-\delta, \ldots, \delta\})$ and propose $Y = X_n + Z$.

(b) Sample $U \sim \mathcal{U}(0,1)$ and set $X_{n+1} = Y$ if $U < \alpha(X_n, y)$, where

$$\alpha(x,y) = \min\left\{1, \frac{p_y}{p_x}\right\} \tag{9}$$

$$= \min\left\{1, \frac{p(1-p)^{y-1}}{p(1-p)^{x-1}}\right\} \tag{10}$$

$$= \min\left\{1, (1-p)^{y-x}\right\} \tag{11}$$

(c) *Otherwise reject y and set $X_{n+1} = X_n$.*

In both cases implement the samplers in a programming language of your choice (with your chosen values of p and n), and confirm that they work by comparing the estimated means and variances with the known theoretical means and variances of these distributions.

Solution. *See the Jupyter notebook.*

Problem 7 (Metropolis-Hastings using deterministic transformations)**.** Suppose we wish to sample from a distribution $\pi(x)$. We consider sampling from this distribution using a Metropolis-Hastings algorithm in which the proposal distribution $q(y\,|x)$ is an equal mixture of two uniform distributions, as follows:

$$q(\cdot\,|\,x) = \frac{1}{2}\mathcal{U}((a-\varepsilon)x, (a+\varepsilon)x) + \frac{1}{2}\mathcal{U}(x/(a+\varepsilon), x/(a-\varepsilon),$$

where a is a constant greater than one, and $0 < \varepsilon < a-1$, for $x \geq 0$, and analogously (i.e. with bounds flipped) for $x < 0$.

1. Derive an expression for the MH acceptance probability for this proposal distribution.

   Solution. *We consider $x > 0$ (the $x < 0$ case is anlogous). The interval $I_1(x) = [(a-\varepsilon)x, (a+\varepsilon)x]$ has length $2\varepsilon x$, and the interval $I_2(x) = \left[\frac{x}{a+\varepsilon}, \frac{x}{a-\varepsilon}\right]$ has length $\frac{2\varepsilon x}{a^2-\varepsilon^2}$. So*

$$q(y|x) = \frac{1}{4\varepsilon x}\left(\mathbb{1}_{I_1(x)}(y) + (a^2-\varepsilon^2)\mathbb{1}_{I_2(x)}(y)\right)$$

   Therefore

$$\begin{aligned}
\alpha(x,y) &= \min\left\{1, \frac{\pi(y)q(x|y)}{\pi(x)q(y|x)}\right\} \\
&= \min\left\{1, \frac{\pi(y)\frac{1}{4\varepsilon x}\left(\mathbb{1}_{I_1(x)}(y) + (a^2-\varepsilon^2)\mathbb{1}_{I_2(x)}(y)\right)}{\pi(x)\frac{1}{4\varepsilon y}\left(\mathbb{1}_{I_1(y)}(x) + (a^2-\varepsilon^2)\mathbb{1}_{I_2(y)}(x)\right)}\right\} \\
&= \min\left\{1, \frac{y\pi(y)\left(\mathbb{1}_{I_1(x)}(y) + (a^2-\varepsilon^2)\mathbb{1}_{I_2(x)}(y)\right)}{x\pi(x)\left(\mathbb{1}_{I_1(y)}(x) + (a^2-\varepsilon^2)\mathbb{1}_{I_2(y)}(x)\right)}\right\}
\end{aligned}$$

   Now notice that

$$\begin{aligned}
y \in I_1(x) &\Leftrightarrow y \in [(a-\varepsilon)x, (a+\varepsilon)x] \Leftrightarrow \frac{y}{x} \in [(a-\varepsilon), (a+\varepsilon)] \\
&\Leftrightarrow \frac{x}{y} \in \left[\frac{1}{a+\varepsilon}, \frac{1}{a-\varepsilon}\right] \Leftrightarrow x \in \left[\frac{y}{a+\varepsilon}, \frac{y}{a-\varepsilon}\right] \Leftrightarrow x \in I_2(y),
\end{aligned}$$

   so $\mathbb{1}_{I_1(x)}(y) = \mathbb{1}_{I_2(y)}(x)$. Similarly $\mathbb{1}_{I_2(x)}(y) = \mathbb{1}_{I_1(y)}(x)$, and we deduce

$$\alpha(x,y) = \min\left\{1, \frac{y\pi(y)\left(\mathbb{1}_{I_1(x)}(y) + (a^2-\varepsilon^2)\mathbb{1}_{I_2(x)}(y)\right)}{x\pi(x)\left(\mathbb{1}_{I_2(x)}(y) + (a^2-\varepsilon^2)\mathbb{1}_{I_1(x)}(y)\right)}\right\} \tag{12}$$

2. Consider the limit of this algorithm as $\varepsilon \to 0$. Describe the resulting algorithm in this limit.

   Solution. *In the limit $\varepsilon \to 0$, the proposal $q(\,\cdot\,|x)$ converges in the sense of distributions to*

$$q_0(y|x) := \frac{1}{2}\left(\delta_{ax}(y) + \delta_{x/a}(y)\right),$$

and the algorithm proposes ax or $\frac{x}{a}$ with probability $1/2$ each. The resulting Markov chain can be viewed as a chain on the discrete state space $\dots \frac{1}{a^2} X_0, \frac{1}{a} X_0, X_0, a\, X_0, a^2\, X_0 \dots$. The probability of acceptance is

$$\alpha(x, ax) = \min\left\{1, \frac{\pi(ax)}{\pi(x)}\right\}, \qquad \alpha(x, x/a) = \min\left\{1, \frac{\pi(x/a)}{\pi(x)}\right\}.$$

3. For the particular case where π is the following distribution

$$\pi(x) = \frac{2}{\pi} \frac{1}{(1+x^2)^2},$$

would either of the schemes proposed be efficient for sampling from π?

Solution. *Note that if $x > 0$, the algorithms propose only positive values for y (and vice versa). It follows that the resulting Markov chains are not ergodic, and the schemes would therefore not be suitable to sample from π. Additionally, the second scheme corresponding to the $\varepsilon \to 0$ limit explores only a discrete subset of $\mathbb{R}$, making it even less efficient.*

Problem 8 (Alternative acceptance probabilities)**.** While the Metropolis-Hastings acceptance probability is by far the most widely used acceptance probability, there are several other choices. One alternative rule is the *Barker rule*:

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

1. Show that the Metropolis Hastings scheme using this acceptance rule is reversible with respect to π, in the case of a continuous state space.

Solution. *We need to show that $\pi(x)p(x,y) = \pi(y)p(y,x)$. We can use the expression for $p(x,y)$ we derived in the lectures*

$$p(x,y) = q(y|x)\alpha(x,y) + \delta_x(y) \int_S (1 - \alpha(x,z))\, q(z|x)\, \mathrm{d}z.$$

As before, if $x = y$ there is nothing to prove. So let $x \neq y$. We have

$$\begin{aligned}
\pi(x)p(x,y) &= \pi(x)q(y|x)\left(1 + \frac{\pi(x)q(y|x)}{q(x|y)\pi(y)}\right)^{-1} \\
&= \pi(x)q(y|x)\left(\frac{q(x|y)\pi(y) + \pi(x)q(y|x)}{q(x|y)\pi(y)}\right)^{-1} \\
&= \frac{\pi(x)q(y|x)q(x|y)\pi(y)}{q(x|y)\pi(y) + \pi(x)q(y|x)} \\
&= \pi(y)q(x|y)\frac{\pi(x)q(y|x)}{q(x|y)\pi(y) + \pi(x)q(y|x)} \\
&= \pi(y)q(x|y)\left(\frac{q(x|y)\pi(y) + \pi(x)q(y|x)}{\pi(x)q(y|x)}\right)^{-1} \\
&= \pi(y)q(x|y)\left(1 + \frac{q(x|y)\pi(y)}{\pi(x)q(y|x)}\right)^{-1} \\
&= \pi(y)q(x|y)\alpha(y,x) = \pi(y)p(y,x)
\end{aligned}$$

as required.

2. Using a proposal $q(\cdot|x) \sim \mathcal{N}(x, \delta^2)$, implement the Barker-rule based scheme, as well as a standard RWMH for a standard Gaussian target distribution π. Plotting the acceptance rate averaged over time, how do they compare?

   Solution. *The probability that a move from a position x is accepted is given by*

$$P_B(x) = \mathbb{E}_{Z\sim q(\cdot|x)}\, \alpha(x,Z) = \int_S q(z|x)\, \alpha(x,z)\, \mathrm{d}z$$
$$= \int_S q(z|x) \left(1 + \frac{\pi(x)q(z|x)}{q(x|z)\pi(z)}\right)^{-1} \mathrm{d}z$$
$$= \int_S q(z|x) \left(\frac{q(x|z)\pi(z)}{q(x|z)\pi(z) + \pi(x)q(z|x)}\right) \mathrm{d}z.$$

   If the proposal is such that $q(z|x) = q(x|z)$, then

$$P_B(x) = \int_S q(z|x) \left(\frac{\pi(z)}{\pi(z) + \pi(x)}\right) \mathrm{d}z.$$

   In contrast, with the usual Metropolis–Hastings algorithm, the probability that a move from a position x is accepted is given by

$$P_{MH}(x) = \mathbb{E}_{Z\sim q(\cdot|x)}\, \alpha(x,Z) = \int_S q(z|x)\, \alpha(x,z)\, \mathrm{d}z$$
$$= \int_S q(z|x) \min\left(1, \frac{q(x|z)\pi(z)}{\pi(x)q(z|x)}\right) \mathrm{d}z.$$

   If the proposal is such that $q(z|x) = q(x|z)$, then

$$P_{MH}(x) = \int_S q(z|x) \min\left(1, \frac{\pi(z)}{\pi(x)}\right) \mathrm{d}z.$$

   Since

$$\frac{\pi(z)}{\pi(z)+\pi(x)} = \min\left(1, \frac{\pi(z)}{\pi(z)+\pi(x)}\right) \leq \min\left(1, \frac{\pi(z)}{\pi(x)}\right),$$

   we deduce that the probability of an acceptance, from any state $x \in S$, is higher for the standard Metropolis–Hastings algorithm. See the Jupyter notebook for the numerics.

3. Compare the performance in terms of effective sample size.

   Solution. *The effective sample size is given by*

$$ESS(N) = \frac{N}{1 + 2\sum_{k=1}^{N-1} \rho_k}$$

   where $\rho_k = \frac{\gamma_k}{\gamma_0}$ and the γ_k are estimated by

$$\gamma_k = \frac{1}{N-k} \sum_{i=1}^{N-k} (X_{i+k} - \bar{X})(X_i - \bar{X}),$$

   with $\bar{X}$ the mean of the sample. Since the acceptance rate of Metropolis–Hastings is higher *than that associated with Barker's acceptance probability, we expect the effective sample size to be larger for Metropolis–Hastings. This is confirmed numerically in the Jupyter notebook.*

Problem 9. Consider the independence sampler, i.e. of the Metropolis–Hastings algorithm with proposal $q(\cdot|x) = g(\cdot)$. Show that, if $\pi(x) \leq M\, g(x)$ for some constant M, then the probability of an acceptance from state x is bounded from below by $\frac{1}{M}$.

Solution. *The probability that a proposal from state x is accepted is given by*

$$
\begin{aligned}
P(x) = \mathbb{E}_{Z \sim q(\cdot|x)}\, \alpha(x, Z) &= \int_S q(z|x)\, \alpha(x,z)\, \mathrm{d}z \\
&= \int_S g(z)\, \min\left(1, \frac{g(x)\pi(z)}{\pi(x)g(z)}\right) \mathrm{d}z \\
&= \int_S \min\left(g(z), \frac{g(x)\pi(z)}{\pi(x)}\right) \mathrm{d}z.
\end{aligned}
$$

In the lecture, I swapped the integral in the minimum, which was not correct! The conclusion, however, was correct: under the assumption that $\pi(x) \leq M\, g(x)$,

$$
\begin{aligned}
P(x) &= \int_S \min\left(\frac{g(z)}{\pi(z)}, \frac{g(x)}{\pi(x)}\right) \pi(z)\, \mathrm{d}z \\
&\geq \int_S \min\left(\frac{1}{M}, \frac{1}{M}\right) \pi(z)\, \mathrm{d}z = \frac{1}{M}.
\end{aligned}
$$

Problem 10 (MALA algorithm and preconditioning)**.** Let $\pi(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$ be a probability density function and suppose that we want to calculate the expectation

$$
\mathbb{E}_\pi f = \int_{\mathbb{R}^d} f(\mathbf{x})\pi(\mathbf{x})\, \mathrm{d}\mathbf{x}, \tag{13}
$$

where $f(\mathbf{x})$ is an arbitrary function such that $\mathbb{E}_\pi f < +\infty$.

1. Explain how you can use a diffusion process of the form

$$
d\mathbf{X}_t = \nabla \log \pi(\mathbf{X}_t)\, dt + \sqrt{2}\, d\mathbf{W}_t, \qquad \mathbf{X}_0 \sim \rho_0, \tag{14}
$$

where $\mathbf{W}_t$ denotes standard Brownian motion on $\mathbb{R}^d$ in order to calculate $\mathbb{E}_\pi f$.

Solution. *The solution to (14) is a continuous-time Markov process with ergodic measure π. For information purposes (= not examinable), we will include a formal that the law of $\mathbf{X}_t$ converges to π as $t \to \infty$, in a sense that will be made precise below. To this end, we recall that the law of $\mathbf{X}_t$, denoted by $\rho(\cdot, t)$, is governed by the Fokker–Planck (or forward Kolmogorov) equation:*

$$
\partial_t \rho = \nabla \cdot \left(-\nabla(\log \pi)\, \rho + \nabla \rho\right), \qquad \rho(\cdot, 0) = \rho_0.
$$

A first observation is that this equation can be rewritten as

$$
\partial_t \rho = \nabla \cdot \left(\pi\, \nabla \left(\frac{\rho}{\pi}\right)\right), \qquad \rho(\cdot, 0) = \rho_0. \tag{15}
$$

It is thus clear that $\rho = \pi$ is a steady state. Let us denote by $L^2(\pi)$ the space of measurable *functions such that*

$$\int_{\mathbb{R}^2} |f(\mathbf{x})|^2 \, \pi(\mathbf{x}) \, \mathrm{d}\mathbf{x} < \infty.$$

We will assume that π satisfies a Poincaré inequality:

$$\int_{\mathbb{R}^2} \left(f(\mathbf{x}) - \int_{\mathbb{R}^2} f(\mathbf{y}) \, \pi(\mathbf{y}) \, \mathrm{d}\mathbf{y} \right)^2 \pi(\mathbf{x}) \, \mathrm{d}\mathbf{x} \leq C \int_{\mathbb{R}^2} |\nabla f(\mathbf{x})|^2 \, \pi(\mathbf{x}) \, \mathrm{d}x$$

for all f such that $f, |\nabla f| \in L^2(\pi)$, and for some constant C independent of f.

Let us now introduce $\tilde{\rho} = \rho - \pi$ and $u = \tilde{\rho}/\pi$, and notice that

$$\int_{\mathbb{R}^2} u(\mathbf{y}, t) \, \pi(\mathbf{y}) \, \mathrm{d}\mathbf{y} = \int_{\mathbb{R}^2} \rho(\mathbf{y}, t) - \pi(\mathbf{y}) \, \mathrm{d}\mathbf{y} = 1 - 1 = 0.$$

Applying the Poincaré inequality to the function u, we therefore obtain

$$\int_{\mathbb{R}^2} |u(\mathbf{x}, t)|^2 \, \pi(\mathbf{x}) \, \mathrm{d}\mathbf{x} \leq C \int_{\mathbb{R}^2} |\nabla u(\mathbf{x}, t)|^2 \, \pi(\mathbf{x}) \, \mathrm{d}x \tag{16}$$

Now, by (15), $\tilde{\rho}$ satisfies

$$\partial_t \tilde{\rho} = \nabla \cdot \left(\pi \, \nabla \left(\frac{\tilde{\rho}}{\pi} \right) \right), \qquad \tilde{\rho}(\cdot, 0) = \rho_0 - \pi,$$

so u solves

$$\pi \, \partial_t u = \nabla \cdot (\pi \, \nabla u) \,, \qquad \tilde{\rho}(\cdot, 0) = \rho_0 - \pi.$$

We deduce from this that

$$\begin{aligned}
\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \left(\int_{\mathbb{R}^2} |u(\mathbf{x}, t)|^2 \, \pi(\mathbf{x}) \, \mathrm{d}\mathbf{x} \right) &= \int_{\mathbb{R}^2} u(\mathbf{x}, t) \, \partial_t u(\mathbf{x}, t) \, \pi(\mathbf{x}) \, \mathrm{d}\mathbf{x} \\
&= \int_{\mathbb{R}^2} \nabla \cdot (\pi(\mathbf{x}) \, \nabla u(\mathbf{x}, t)) \; u(\mathbf{x}, t) \, \mathrm{d}\mathbf{x} \\
&= - \int_{\mathbb{R}^2} |\nabla u(\mathbf{x}, t)|^2 \, \pi(\mathbf{x}) \, \mathrm{d}\mathbf{x},
\end{aligned}$$

where we used integration by part, assuming that the boundary terms are zero, which is reasonable because we expect that $\pi(\mathbf{x}) \to 0$ as $|\mathbf{x}| \to \infty$. Now, using (16), we deduce

$$\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \left(\int_{\mathbb{R}^2} |u(\mathbf{x})|^2 \, \pi(\mathbf{x}) \, \mathrm{d}\mathbf{x} \right) \leq -C \int_{\mathbb{R}^2} |u(\mathbf{x})|^2 \, \pi(\mathbf{x}) \, \mathrm{d}\mathbf{x}.$$

Using Grönwall's lemma, we conclude

$$\int_{\mathbb{R}^2} |u(\mathbf{x}, t)|^2 \, \pi(\mathbf{x}) \, \mathrm{d}\mathbf{x} \leq \mathrm{e}^{-2Ct} \int_{\mathbb{R}^2} |u(\mathbf{x}, 0)|^2 \, \pi(\mathbf{x}) \, \mathrm{d}\mathbf{x},$$

which implies that $u(\cdot, t) \to 0$ exponentially in $L^2(\pi)$ as $t \to \infty$. The left-hand side can be expressed in terms of ρ as

$$\int_{\mathbb{R}^2} |\rho(\mathbf{x}, t) - \pi(\mathbf{x})|^2 \, \pi(\mathbf{x})^{-1} \, \mathrm{d}\mathbf{x},$$

so we conclude that $\rho(\cdot, t) \to \pi$ exponentially in $L^2(\pi^{-1})$ as $t \to \infty$. To give a complete answer to the question, we would need to prove additionally an ergodicity statement, but this is beyond the scope of this course.

2. Let $\pi(x)$ be a bivariate normal distribution $\pi \sim \mathcal{N}(\mu, \Sigma)$, where

$$\mu = \begin{bmatrix} 3 \\ 6 \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 2 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

(a) Write down $\pi(x)$, $\log \pi(x)$ and $\nabla \log \pi(x)$.

Solution. *To alleviate the notations, from here on we no longer use a bold font for vectors. We have*

$$\pi(x) = \frac{1}{2\pi \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}\langle \Sigma^{-1}(x-\mu), (x-\mu)\rangle\right)$$

so

$$\log(\pi(x)) = \log(-2\pi \det(\Sigma)^{-1/2}) - \frac{1}{2}\langle \Sigma^{-1}(x-\mu), (x-\mu)\rangle$$

and

$$\nabla \log(\pi(x)) = -\Sigma^{-1}(x-\mu)$$

because Σ is symmetric.

(b) Use the above calculations to sample from π using the MALA distribution.

Solution. *The MALA proposes Y by stepping in the SDE*

$$dX_t = \nabla \log \pi(X_t)\ dt + \sqrt{2}\ dW_t$$

using the Euler-Maruyama scheme with time step δ. So we propose

$$Y = X_n - \Sigma^{-1}(X_n - \mu)\delta + \sqrt{2\delta}\,\xi$$

where $\xi \sim \mathcal{N}(0,1)$. The acceptance probability is

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

and, as we saw in lectures,

$$q(y|x) \propto \exp\left(-\frac{\|y - x + \delta\Sigma^{-1}(x-\mu)\|^2}{4\delta}\right).$$

For the implementation see the Jupyter notebook.

(c) Compute an estimator for $I = \mathbb{E}(f(\mathbf{X}))$, where $\mathbf{X} = (X, Y) \sim \pi$ and $f(\mathbf{x}) = x^3 + y^2$.

Solution. *See the Jupyter notebook.*

(d) Track the acceptance rate for different time steps δ.

Solution. *See the Jupyter notebook.*