These notes are loosely based on Section 5.3 of the lecture notes from 2016.

Inference for Stochastic Differential Equations

Once a stochastic model for a given physical system has been derived, we must choose the parameters such that the output of the stochastic model agrees with the observed data. In this section, we present some simples techniques for estimating the diffusion coefficient and parameters in SDEs. As usual, we shall focus on the one-dimensional case. We shall consider the following one dimensional Itô SDE of the form:

$$\mathrm{d}Xt = b(X_t;\theta)\,\mathrm{d}t + \sigma(X_t;\theta)\,\mathrm{d}W_t, \qquad X_0 = x_0, \tag{1}$$

where $\theta \in \Theta \subset \mathbb{R}^N$ is a finite set of parameters that we want to estimate from the observations. The initial conditions can be taken to be either deterministic or random. We assume that we are provided with observations of the path of the process. This can be either be:

1. Discrete observations $X_{t_0}, X_{t_1}, \ldots, X_{t_N}$, or
2. The entire path X_t, $t \in [0, T]$.

Some simple examples

- The Ornstein-Uhlenbeck process with unknown drift coefficient α:

$$dX_t = -\alpha X_t\, dt + dW_t.$$

- Brownian motion in a bistable potential, with unknown parameters A, B:

$$dX_t = (AX_t - BX_t^3)\, dt + dW_t.$$

Inferring the diffusion coefficient

In order to estimate parameters in the diffusion coefficient, it is natural to use the *quadratic variation* over the interval $[0, T]$ of the solution X_t of the SDE (1), which is defined as

$$[X]_T := \lim_{\Delta t \to 0} \sum_{k=0}^{N-1} \left| X_{t_{k+1}} - X_{t_k} \right|^2, \tag{2}$$

where the limit is in probability. We will show in Proposition 1 that this limit is well defined for the solution to (1) when the diffusion term is constant, $\sigma(X_t, \theta) = \sigma$, and that

$$[X]_T = \int_0^T \sigma^2\, \mathrm{d}s = T\,\sigma^2. \tag{3}$$

For more general Itô processes, such as the solution to (1) with non-constant diffusion, it is possible to show that the limit in the definition of the quadratic variation also converges in probability and that $[X]_T$ admits the explicit expression

$$[X]_T = \int_0^T \sigma^2(X_s;\theta)\, \mathrm{d}s,$$

but we will not do this here. Equation (3) and the limit in (2) suggest that we define the following estimator for σ^2:

$$\hat{\sigma}_N^2 = \frac{1}{T} \sum_{k=0}^{N-1} \left| X_{t_{k+1}} - X_{t_k} \right|^2, \qquad t_k = k\,(T/N). \tag{4}$$

Notice that this is an estimator for σ^2 and not for σ. This is because the law of the solution to (1) (and in particular all the finite-dimensional distributions of the solution) is exactly the same when σ is replaced by $-\sigma$, which is a consequence of the fact that $W_t = -W_t$ in law. In other words, the sign of σ cannot be deduced from observations.

In class we showed that the bias of the estimator (4) decreases to 0 as $\Delta t^{1/2}$. Here we provide a more complete result for information purposes, but you are expected to know the proof of only the simpler result shown in class.

Proposition 1 (The mean square error scales as Δt)**.** *Assume that the drift coefficient $b(\cdot)$ is a bounded function and let $\{X_{t_i}\}_{i=0}^N$ be a sequence of equidistant observations of the solution to*

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

with timestep $\Delta t = T/N$ and T fixed. Let also $\hat{\sigma}_N^2$ be as defined in (4)*. Then the mean-square error converges to zero in the limit as $N \to \infty$. More precisely,*

$$MSE(\hat{\sigma}_N^2) = \mathrm{var}[\hat{\sigma}_N^2] + |\mathbb{E}|\hat{\sigma}_N^2| - \sigma^2|^2 \leq C(\Delta t + \Delta t^2).$$

Remark 1. This result implies that $\hat{\sigma}_N^2$ is asymptotically unbiased. By Chebyshev's inequality, it also implies, that $\hat{\sigma}_N^2$ is weakly consistent: $\hat{\sigma}_N^2 \to \sigma^2$ in probability in the limit as $N \to \infty$. ⊘

Proof of Proposition 1. In this proof, C denotes any constant that is independent of Δt; it can change from occurrence to occurrence. We have that

$$X_{t_{i+1}} - X_{t_i} = \int_{t_i}^{t_{i+1}} b(X_s)\,\mathrm{d}s + \sigma\,\Delta W_i =: I_i + M_i.$$

where $\Delta W_i := W_{t_{i+1}} - W_{t_i} \sim \mathcal{N}(0, \Delta t)$. We substitute this into (4) to obtain

$$e := T(\hat{\sigma}_N^2 - \sigma^2) = \sum_{i=0}^{N-1} I_i^2 + 2 \sum_{i=0}^{N-1} I_i\,M_i + \sigma^2 \sum_{i=0}^{N-1} \underbrace{(|\Delta W_i|^2 - \Delta t)}_{=:Z_i}. \tag{5}$$

By definition, the mean-square error is equal to $\mathbb{E}[e^2]/T^2$. Since $b(\cdot)$ is bounded by assumption, we obtain using the Cauchy-Schwarz inequality

$$|I_i|^2 \leq \langle b(X_s), 1\rangle_{L^2(t_i,t_{i+1})}^2 \leq \Delta t\,\|b(X_s)\|_{L^2(t_i,t_{i+1})}^2 = \Delta t \int_{t_i}^{t_{i+1}} b(X_s)^2\,\mathrm{d}s \leq C\,\Delta t^2. \tag{6}$$

Therefore, employing the fact that the Z_i are i.i.d. and $\mathbb{E}[Z_i] = 0$, together with the standard inequality $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$, we deduce

$$\mathbb{E}[e^2] \leq C\Delta t^2 + 6\,\mathbb{E}\left|\sum_{i=0}^{N-1} I_i\,M_i\right|^2 + 3\,\sigma^4\,N\,\mathbb{E}|Z_0^2|.$$

Using Cauchy-Schwarz' inequality, Young's inequality with ε, and (6), we obtain for any $\varepsilon > 0$

$$\left|\sum_{i=0}^{N-1} I_i\, M_i\right|^2 \leq N \sum_{i=0}^{N-1} |I_i|^2\, |M_i|^2 \leq \frac{N}{2} \sum_{i=0}^{N-1} \left(\frac{|I_i|^4}{\varepsilon} + \varepsilon\, |M_i|^4 \right) \leq C\, N \sum_{i=0}^{N-1} \left(\frac{\Delta t^4}{\varepsilon} + \varepsilon |M_i|^4 \right).$$

Using the fact that $\mathbb{E}|M_i|^4 = \mathbb{E}|\sigma\, \Delta W_i|^4 = C\, \Delta t^2$, and choosing $\varepsilon = \Delta t$ in order to balance the terms (and thereby obtain the best possible bound), we obtain

$$\mathbb{E} \left|\sum_{i=0}^{N-1} I_i\, M_i\right|^2 \leq C \Delta t.$$

Using the fact that $\mathbb{E}[Z_0^2] = C \Delta t^2$ and going back to (5), we deduce

$$\mathbb{E}[e^2] \leq C(\Delta t^2 + \Delta t).$$

which concludes the proof. □

Proposition 2 (Simpler result shown in class - examinable)**.** *With the same notations and assumptions as in Proposition 1, it holds that*

$$|\mathbb{E}|\hat{\sigma}_N^2| - \sigma^2| \leq C(\sqrt{\Delta t} + \Delta t).$$

Proof. We have that

$$X_{t_{i+1}} - X_{t_i} = \int_{t_i}^{t_{i+1}} b(X_s)\, \mathrm{d}s + \sigma\, \Delta W_i =: I_i + M_i.$$

where $\Delta W_i := W_{t_{i+1}} - W_{t_i} \sim \mathcal{N}(0, \Delta t)$. Employing (4), we obtain

$$\hat{\sigma}_N^2 - \sigma^2 = \frac{1}{T} \left(\sum_{i=0}^{N-1} I_i^2 + 2 \sum_{i=0}^{N-1} I_i\, M_i + \sigma^2 \sum_{i=0}^{N-1} (|\Delta W_i|^2 - \Delta t) \right),$$

so

$$\mathbb{E}[\hat{\sigma}_N^2 - \sigma^2] = \frac{1}{T} \left(\sum_{i=0}^{N-1} \mathbb{E}[I_i^2] + 2 \sum_{i=0}^{N-1} \mathbb{E}[I_i\, M_i] \right). \tag{7}$$

Since $b(\cdot)$ is bounded by assumption, an application of the Cauchy-Schwarz inequality gives

$$|I_i|^2 \leq \langle b(X_s), 1\rangle_{L^2(t_i,t_{i+1})}^2 \leq \Delta t\, \|b(X_s)\|_{L^2(t_i,t_{i+1})}^2 = \Delta t \int_{t_i}^{t_{i+1}} b(X_s)^2\, \mathrm{d}s \leq C\, \Delta t^2. \tag{8}$$

On the other hand, by Young's inequality with ε,

$$|I_i\, M_i| \leq \frac{1}{2\varepsilon} |I_i|^2 + \frac{\varepsilon}{2} |M_i|^2 \qquad \forall \varepsilon > 0,$$

so, using (8) and the fact that $\mathbb{E}|M_i|^2 = \mathbb{E}|\sigma\, \Delta W_i|^2 = \sigma^2\, \Delta t$, we obtain

$$\mathbb{E}|I_i\, M_i| \leq \frac{C}{2\varepsilon} \Delta t^2 + \frac{\varepsilon}{2} (\sigma^2\, \Delta t).$$

Taking $\varepsilon = \sqrt{\Delta t}$, we deduce $\mathbb{E}|I_i\, M_i| \leq C\,\Delta t^{3/2}$, and going back to (7) we conclude

$$\mathbb{E}[\hat{\sigma}_N^2 - \sigma^2] \leq \frac{C}{T}\left(N\Delta t^2 + N\Delta t^{3/2}\right) = C\left(\Delta t + \sqrt{\Delta t}\right). \tag{9}$$

(Note that, since $\Delta t \leq T$ and T is fixed, we have in fact $\mathbb{E}[\hat{\sigma}_N^2 - \sigma^2] \leq C\sqrt{\Delta t}$ for another constant C independent of Δt.) □

Proposition 1 informs us that, provided that Δt is small enough, our estimator $\hat{\sigma}_N^2$ is close to σ^2 with high probability. With a little bit more effort, it is possible to show $\hat{\sigma}_N^2$ is "almost strongly consistent", in the sense that suitably chosen subsequences $\{\hat{\sigma}_{N_k}^2\}_{k=1}^{\infty}$ converge almost surely to σ^2.

Corollary 3 (Almost sure convergence - not examinable)**.** *Under the assumptions of Proposition 1, the estimator defined by* $\tilde{\sigma}_N^2 = \hat{\sigma}_{2^N}^2$ *is strongly consistent:* $\lim_{N\to\infty} \tilde{\sigma}_N^2 = \sigma^2$ *almost surely.*

Proof. By Markov's inequality (which simply follows from $I_{[\varepsilon,\infty)}(x) \leq x/\varepsilon$ any $x > 0$)

$$\mathbb{P}(|\tilde{\sigma}_N^2 - \sigma^2|^2 \geq \varepsilon) = \mathbb{E}\left[I_{[\varepsilon,\infty)}(|\tilde{\sigma}_N^2 - \sigma^2|^2)\right] \leq \frac{1}{\varepsilon}\,\mathbb{E}\left[|\tilde{\sigma}_N^2 - \sigma^2|^2\right] \leq C\,2^{-N},$$

for any $\varepsilon > 0$, and where C is independent of N. Therefore

$$\sum_{N=1}^{\infty} \mathbb{P}(|\tilde{\sigma}_N^2 - \sigma^2|^2 \geq \varepsilon) < \infty.$$

Let us now denote by E_N be the event that $|\tilde{\sigma}_N^2 - \sigma^2|^2 \geq \varepsilon$. By the Borel–Cantelli lemma,

$$\mathbb{P}\left[\limsup_{N\to\infty} E_N\right] = \mathbb{P}\left[\bigcap_{N=1}^{\infty}\bigcup_{k=N}^{\infty} E_k\right] = 0.$$

Now notice that if $\omega \in \Omega$ is such that $\omega \notin \bigcap_{N=1}^{\infty}\bigcup_{k=N}^{\infty} E_k$, then $\omega \in E_N$ for only finitely many N. Therefore $L := \limsup_{N\to\infty} |\tilde{\sigma}_N^2 - \sigma^2|^2 < \varepsilon$ almost surely. Since this is true for any $\varepsilon > 0$, we conclude by employing subadditivity:

$$\mathbb{P}\left[L > 0\right] = \mathbb{P}\left[\bigcup_{M=1}^{\infty} \{\omega \in \Omega : L(\omega) > 1/M\}\right] \leq \sum_{M=1}^{\infty} \mathbb{P}[L > 1/M] = 0.$$

□

Remark 2 (Showing Itô's formula in a particular case – not examinable)*.* Note that Proposition 1 implies, in the particular case where X_t is a Brownian motion ($\mathrm{d}X_t = \mathrm{d}W_t$), that

$$\sum_{k=0}^{N-1} \left|W_{t_{k+1}} - W_{t_k}\right|^2 \to T \text{ in } L^2(\Omega) \text{ as } N \to \infty.$$

As a byproduct of this result, we can now prove

$$I = \int_0^T W_s\,\mathrm{d}W_s = \frac{W_T^2}{2} - \frac{T}{2}.$$

using only the rigorous definition of the Itô integral, i.e. without using Itô's formula. To this end, let us define a piecewise constant approximation of the Brownian motion:

$$W_s^N = W_{n_s \Delta t}^N, \qquad n_s = \left\lfloor \frac{s}{\Delta t} \right\rfloor, \qquad \Delta t = \frac{T}{N}.$$

Clearly,

$$\int_0^T \mathbb{E}|W_s - W_s^N|^2 \mathrm{d}s = \int_0^T \mathbb{E}|W_s - W_{n_s\Delta t}|^2 \mathrm{d}s = \int_0^T (s - n_s \Delta t)\, \mathrm{d}s = \frac{\Delta t}{2}.$$

Therefore $W_s^N \to W_s$ in $L^2(\Omega \times [0,T])$ as $N \to \infty$. Consequently, by Itô's isometry

$$I_N := \int_0^T W_s^N \, \mathrm{d}s \to I \text{ in } L^2(\Omega) \text{ as } N \to \infty.$$

Now notice that, using the usual notation $t_k = k\Delta t$,

$$\begin{aligned}
I_N &= \sum_{k=0}^{N-1} W_{t_k} \left(W_{t_{k+1}} - W_{t_k}\right) \\
&= \frac{1}{2} \sum_{k=0}^{N-1} (W_{t_k} + W_{t_{k+1}}) \left(W_{t_{k+1}} - W_{t_k}\right) - \frac{1}{2} \sum_{k=0}^{N-1} (W_{t_{k+1}} - W_{t_k}) \left(W_{t_{k+1}} - W_{t_k}\right) \\
&= \frac{1}{2} W_T^2 - \frac{1}{2} \sum_{k=0}^{N-1} |W_{t_{k+1}} - W_{t_k}|^2 \to \frac{W_T^2}{2} - \frac{T}{2} \quad \text{in } L^2(\Omega).
\end{aligned}$$

$\oslash$

Estimating the drift coefficient

From now on, we assume that we have already estimated the diffusion coefficient. Thus, we will set $\sigma = 1$, in which case (1) becomes

$$\mathrm{d}X_t = b(X_t; \theta)\, \mathrm{d}t + \mathrm{d}W_t, \qquad X_0 \sim \rho_0. \tag{10}$$

We take the initial condition to be random in order to illustrate a case different from the one seen in class. Our objective is to estimate the unknown parameters in the drift $\theta \in \Theta$ from a time-series of observations. To this end, we will use the *maximum likelihood estimator* (MLE). We will begin by describing the general intuition of the MLE in the case of observations $\{X^{(j)}\}_{j=1}^J$ that live in a finite dimensional state space, but the approach carries over *mutatis mutandis* to the case of function-valued observations. The only difference in that case is that the law of Brownian motion is used as a reference measure, instead of the Lebesgue measure: the Radon–Nikodym derivative with respect to the law of Brownian motion, usually obtained by Girsanov's theorem, is used in place of the *probability distribution function* (PDF).

General introduction to the maximum likelihood estimator

Suppose we have a number J of i.i.d. observations, denoted by $X^{(1)}, \dots, X^{(J)}$, of a random variable X with PDF $f(x; \theta_0)$. Since the observations are independent, their joint PDF (i.e.

density of the law with respect to the Lebesgue measure) is given by

$$L_J(\mathbf{x};\theta_0) = \prod_{j=1}^{J} f(x^{(j)};\theta_0), \qquad \mathbf{x} = \{x^{(j)}\}_{j=1}^{J}.$$

The function $L_J(\mathbf{X};\theta)$, where $\mathbf{X} = \{X^{(j)}\}_{j=1}^{N}$, is called the *likelihood function*, and Θ is a set of admissible parameters. It is viewed as a function of θ given the data $\mathbf{X}$. The maximum likelihood estimator (MLE) is then

$$\hat{\theta} = \operatorname*{arg\,max}_{\theta\in\Theta} L_J(\mathbf{X};\theta) = \operatorname*{arg\,max}_{\theta\in\Theta} \frac{1}{J}\sum_{j=1}^{J} \ln f(X^{(j)};\theta),$$

where the second equality is justified because $x \mapsto \ln x$ is an increasing function. Let us use the notation $l_J(\theta) := \frac{1}{J}\sum_{j=1}^{J} \ln f(X^{(j)};\theta)$. By the strong law of large numbers,

$$\lim_{J\to\infty} l_J(\theta) = \mathbb{E}_{X\sim f(x;\theta_0)}\left[\ln f(X;\theta)\right] = \int \ln(f(x;\theta))\, f(X;\theta_0)\,\mathrm{d}x =: l(\theta) \qquad \text{almost surely.}$$

If the $l(\cdot)$ admits a unique global maximum, it is therefore reasonable to expect $\hat{\theta}$ to converge to the maximizer of $l(\cdot)$ as $J\to\infty$. It is possible to prove this rigorously under appropriate conditions using the uniform law of large numbers, but we will not present this here. We only show that, if $l(\cdot)$ admits a unique maximizer, this maximizer is necessarily θ_0.

Lemma 4. *It holds that $l(\theta) \leq l(\theta_0)$ for all $\theta\in\Theta$.*

Proof. Since $\ln(x) \leq x-1$ for all $x>0$,

$$\begin{aligned} l(\theta) - l(\theta_0) &= \int \left[\ln(f(x;\theta)) - \ln(f(x;\theta_0))\right] f(x;\theta_0)\,\mathrm{d}x = \int \ln\left(\frac{f(x;\theta)}{f(x;\theta_0)}\right) f(x;\theta_0)\,\mathrm{d}x \\ &\leq \int \left(\frac{f(x;\theta)}{f(x;\theta_0)} - 1\right) f(x;\theta_0)\,\mathrm{d}x = \int \left(f(x;\theta) - f(x;\theta_0)\right)\mathrm{d}x = 1-1 = 0, \end{aligned}$$

which concludes the proof. □

Example 1. Suppose that $\mathbf{X} = \{X^{(j)}\}_{j=1}^{J}$ are i.i.d. samples from a Gaussian $\mathcal{N}(\mu,\sigma^2)$ with unknown parameters μ and σ^2. The likelihood function takes the form

$$L(\mathbf{x};\mu,\sigma) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{\sum_{j=1}^{J}(x^{(j)}-\mu)^2}{2\sigma^2}\right), \qquad \mathbf{x}\in\mathbb{R}^J.$$

The maximum likelihood estimator $(\hat{\mu},\hat{\sigma}^2)$ for (μ,σ^2) is given by

$$(\hat{\mu},\widehat{\sigma^2}) = \operatorname*{arg\,max}_{\mu,\sigma^2}\, L(\mathbf{X};\mu,\sigma^2)$$

which gives

$$\hat{\mu} = \frac{1}{J}\sum_{j=1}^{J} X^{(j)}, \qquad \hat{\sigma}^2 = \frac{1}{J}\sum_{j=1}^{J}(X^{(j)}-\hat{\mu})^2.$$

Notice that this estimator for the variance is biased (but asymptotically unbiased). ⊘

Example 2. Suppose that $W_t^{(j)}$ are independent Brownian motions on $[0,T]$ and that the observations $\mathbf{X} = \{X^{(j)}\}_{j=1}^J$ are obtained by $X^{(j)} = (\sigma W_{t_1}, \dots, \sigma W_{t_N})$, where $t_k = k(T/N)$ for some $N \in \mathbb{N}_{>0}$. Our aim in this example is to obtain the maximum likelihood estimator for σ^2. In class we examined only the case where $J = 1$, but it is only slightly more difficult to consider the more general case, which we consider here. The PDF of an individual observation is

$$f(x^{(j)};\sigma) = \left|\frac{1}{2\pi\,\sigma^2\,\Delta t}\right|^{N/2} \exp\left(-\frac{1}{2\,\sigma^2\,\Delta t}\sum_{k=0}^{N-1}|x_{k+1}^{(j)} - x_k^{(j)}|^2\right), \qquad x^{(j)} \in \mathbb{R}^N, \qquad x_0^{(j)} := 0.$$

The logarithm of the likelihood is thus given by

$$L_J(\mathbf{X};\sigma) = -\frac{J\,N}{2}\ln(\pi\,\sigma^2\,\Delta t) - \frac{1}{2\,\sigma^2\,\Delta t}\sum_{j=1}^{J}\sum_{k=0}^{N-1}|X_{k+1}^{(j)} - X_k^{(j)}|^2.$$

Equaling the derivative with respect to σ^2 to 0, we obtain

$$-\frac{J\,N}{2\,\hat\sigma^2} + \frac{1}{2\,\hat\sigma^4\,\Delta t}\sum_{j=1}^{J}\sum_{k=0}^{N-1}|X_{k+1}^{(j)} - X_k^{(j)}|^2 = 0,$$

which gives, taking into account that $N\Delta t = T$,

$$\hat\sigma^2 = \frac{1}{J}\sum_{j=1}^{J}\left(\frac{1}{T}\sum_{k=0}^{N-1}|X_{k+1}^{(j)} - X_k^{(j)}|^2\right).$$

In the case where $J = 1$ (i.e. when we are inferring the diffusion coefficient from only one replica of the process σW_t, which you can view as the solution to $\mathrm{d}X_t = \sigma\,\mathrm{d}W_t$ with $X_0 = 0$), this estimator coincides with the one obtained in (4) above. $\oslash$

Application to the drift coefficient

We now want to derive maximum likelihood estimators for the parameters in the drift of (10). As a first step, we will focus on the easier problem of estimating the drift in the stochastic difference equation that is obtained after Euler–Maruyama discretization of (10):

$$X_{n+1} - X_n = b(X_n;\theta)\,\Delta t + \xi_n\sqrt{\Delta t}, \qquad \xi_n \sim \mathcal{N}(0,1), \qquad n = 0,\dots,N-1, \tag{11}$$

with the initial condition $X_0 \sim \rho_0$. To this end, let us assume that we have J discrete-time trajectories $\mathbf{X} = \{X^{(j)}\}_{j=1}^J$ of the discrete dynamics (11). The PDF of $X^{(j)}$ is

$$f(x^{(j)};\sigma) = \left|\frac{1}{2\pi\Delta t}\right|^{N/2}\rho_0(x_0^{(j)})\exp\left(-\frac{1}{2\,\Delta t}\sum_{k=0}^{N-1}\left|x_{k+1}^{(j)} - x_k^{(j)} - b(x_k^{(j)};\theta)\,\Delta t\right|^2\right),$$

with $x^{(j)} \in \mathbb{R}^{N+1}$. The joint PDF of $\{X^{(j)}\}_{j=1}^J$ can then be obtained by simply taking the tensor product of J times this PDF, because the replicas $\{X^{(j)}\}_{j=1}^J$ are i.i.d. To fix ideas, we will now consider the particular case $b(x;\theta) = -\theta\,b(x)$. Since we are in a finite-dimensional setting, the likelihood can be defined by using the Lebesgue measure as reference measure, as we did above for the diffusion coefficient. Some authors define the likelihood in this case as the function obtained after dividing by the PDF of discretized Brownian motions, but the two approaches are equivalent because that PDF does not depend on the parameter θ. Taking the logarithm of the likelihood, we obtain

$$L_J(\mathbf{X};\sigma) = -\frac{J\,N}{2}\ln(2\,\pi\,\Delta t) + \sum_{j=1}^{J}\ln\rho_0(X_0^{(j)}) - \frac{1}{2\,\Delta t}\sum_{j=1}^{J}\sum_{k=0}^{N-1}\left|X_{k+1}^{(j)} - X_k^{(j)} + \theta\,b(X_k^{(j)})\,\Delta t\right|^2.$$

Equaling the derivative with respect to θ to 0, we obtain

$$\sum_{j=1}^{J}\sum_{k=0}^{N-1}\left(X_{k+1}^{(j)} - X_k^{(j)} + \hat\theta\,b(X_k^{(j)})\,\Delta t\right)\,b(X_k^{(j)}) = 0,$$

which gives

$$\hat\theta = -\frac{\sum_{j=1}^{J}\sum_{k=0}^{N-1} b(X_k^{(j)})\,\left(X_{k+1}^{(j)} - X_k^{(j)}\right)}{\sum_{j=1}^{J}\sum_{k=0}^{N-1}\,|b(X_k^{(j)})|^2\,\Delta t}.$$

- In the case where the drift coefficient is constant: $b(\cdot) = 1$, the estimator simplifies to

$$\hat\theta = -\frac{1}{J}\sum_{j=1}^{J}\frac{X_N^{(j)} - X_0^{(j)}}{N\Delta t}.$$

- If only one replica of (11) is available, the maximum likelihood estimator is

$$\hat\theta = -\frac{\sum_{k=0}^{N-1} b(X_k)\,(X_{k+1} - X_k)}{\sum_{k=0}^{N-1}\,|b(X_k)|^2\,\Delta t}. \tag{12}$$

Let us now discuss how this methodology can be adapted to estimate the drift coefficient based on J continuous-time solutions of (10). For simplicity, and since we saw that considering several replicas does not generally pose serious difficulties, we will consider that $J = 1$: we are estimating the drift coefficient based on only one, possibly long trajectory of (10). The main additional obstacle in this case is that, because there is no analogue of the Lebesgue measure in infinite dimensional Banach spaces, we need to write the density of the law of $X^{(0)}$ with respect to a different reference measure. The usual choice is to define the likelihood function via the density with respect to the law of Brownian motion, which is given by Girsanov's theorem:

$$\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\mathbb{P}_W}(X;\theta) = \exp\left(\int_0^T b(X_t;\theta)\,\mathrm{d}X_t - \frac{1}{2}\int_0^T b(X_t,\theta)^2\,\mathrm{d}t\right).$$

Taking the logarithm of the likelihood and the derivative with respect to θ, and dropping the superscript from $X^{(0)}$ for simplicity, we obtain the following expression for the MLE in the particular case where $b(X_t;\theta) = -b(X_t)\,\theta$,

$$\hat\theta = -\frac{\int_0^T b(X_t)\,\mathrm{d}X_t}{\int_0^T |b(X_t)|^2\,\mathrm{d}t},$$

which has a structure similar to that in (12). Of course, we wouldn't be able to evaluate this estimator excactly. Given a set of discrete equidistant observations $\{X_k\}_{k=0}^N$, we could approximate the integrals:

$$\hat{\theta} \approx -\frac{\sum_{k=0}^{N-1} b(X_k)\ (X_{k+1} - X_k)}{\sum_{k=0}^{N-1}\ |b(X_k)|^2\,\Delta t}. \tag{13}$$

Example 3 (MLE for the stationary Ornstein-Uhlenbeck process)**.** Consider the stationary Ornstein-Uhlenbeck process

$$\mathrm{d}X_t = -\alpha X_t\,\mathrm{d}t + \mathrm{d}W_t, \qquad X_0 \sim \mathcal{N}\left(0, \frac{1}{2\alpha}\right).$$

The MLE estimator for α in this case is

$$\hat{\alpha} = -\frac{\int_0^T X_t\,dX_t}{\int_0^T X_t^2\,dt}.$$

It is possible to show that that this estimator becomes asymptotically unbiased in the *large sample limit* $N \to +\infty$, for Δt fixed. ⊘

Exercise 1 (Maximum Likelihood estimator for a bistable SDE)**.** 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 estimatorrs 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

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

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

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