Bayesian inference of multiscale diffusion processes

Assyr Abdulle* Giacomo Garegnani* Grigoris Pavlitois †

Let $(\Omega, \mathcal{A}, P)$ be a probability space and consider the scalar Ornstein–Uhlenbeck process

$$\begin{aligned} \mathrm{d}X_t(\omega, a) &= -aX_t(\omega, a)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t(\omega), \\ X_0(\omega, a) &= x \in \mathbb{R}, \end{aligned} \tag{1}$$

where a is an unknown parameter with true value $a^\dagger \in \mathbb{R}^+$ and $\Sigma \in \mathbb{R}$ is fixed. Moreover, we consider the initial condition $x \in \mathbb{R}$ to be deterministic and independent of a. Let us define the forward mapping $\mathcal{G} \colon \Omega \times \mathbb{R} \to \mathbb{R}^N$ as

$$\mathcal{G}(\omega, a) \mapsto \mathcal{X}_N = \{X_{t_j}(\omega, a)\}_{j=1}^N, \tag{2}$$

for a sequence of time instants t_j defined as $t_j = jh$, for $h > 0$ and for $j = 1, \dots, N$, $t_N = T$. We are given a set of observations $\mathcal{Y}_N \in \mathbb{R}^n$ of the process X_t as

$$\mathcal{Y}_N = \mathcal{G}(\omega^\dagger, a^\dagger) + \eta,$$

where η is an additive Gaussian source of noise distributed as $\eta \sim \mathcal{N}(0, \sigma_\eta^2 I_N)$, where I_d is the identity in d dimensions, and $\omega^\dagger \in \Omega$ is a fixed event. As in (2), we consider $\mathcal{Y}_N = \{Y_{t_j}(\omega, a)\}_{j=1}^N$. In the following, we denote by $\mathcal{N}(x; \mu, \sigma^2)$ the density of a Gaussian random variable of mean μ and variance σ^2 evaluated at a point x, i.e.,

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

Fixed a Gaussian prior $p(a) = \mathcal{N}(a; 0, \sigma_a^2)$ on the drift coefficient, our goal is to compute the posterior distribution $p(a \mid \mathcal{Y}_N)$ of the parameter given the observations in a Bayesian framework. First, we apply Bayes' rule to obtain

$$\begin{aligned} p(a, \mathcal{X}_N \mid \mathcal{Y}_N) &\propto p(a, \mathcal{X}_N)\, p(\mathcal{Y}_N \mid a, \mathcal{X}_N) \\ &\propto p(a)\, p(\mathcal{X}_N \mid a)\, p(\mathcal{Y}_N \mid a, \mathcal{X}_N), \end{aligned}$$

which, thanks to the observation model and to the Markov property of X_t, can be written as

$$\begin{aligned} p(a, \mathcal{X}_N \mid \mathcal{Y}_N) &\propto p(a)\, p(\mathcal{X}_N \mid a) \prod_{j=1}^N p(Y_{t_j} \mid X_{t_j}, a) \\ &= p(a) \prod_{j=1}^N p(X_{t_j} \mid X_{t_{j-1}}, a) \prod_{j=1}^N p(Y_{t_j} \mid X_{t_j}, a) \end{aligned} \tag{3}$$

Then, in order to obtain the marginal posterior of the parameter, one has to integrate out $\mathcal{X}_N$ and get the posterior distribution of the parameter

$$p(a \mid \mathcal{Y}_N) \propto \int p(a, \mathcal{X}_N \mid \mathcal{Y}_N)\,\mathrm{d}\mathcal{X}_N$$

*Institute of Mathematics, École Polytechnique Fédérale de Lausanne ({assyr.abdulle, giacomo.garegnani}@epfl.ch)
†Department of Mathematics, Imperial College London

Let us compute the terms in (3) separately. First, the prior is given by $p(a) = \mathcal{N}(a; 0, \sigma_a^2)$. Then, for the likelihood term we have $p(Y_{t_j} \mid X_{t_j}, a) = \mathcal{N}(X_{t_j}; Y_{t_j}, \sigma_\eta^2)$. Finally, due to the distribution of the OU process we get

$$p(X_{t_j} \mid X_{t_{j-1}}, a) = \mathcal{N}\big(X_{t_j}; e^{-ah} X_{t_{j-1}}, a^{-1}\Sigma(1 - e^{-2ah})\big).$$

For h small enough, we can consider the transition probability of the Euler–Maruyama method applied to (1), which reads

$$p(X_{t_j} \mid X_{t_{j-1}}, a) = \mathcal{N}\big(X_{t_j}; (1 - ah)X_{t_{j-1}}, 2\Sigma h\big),$$

which is a good approximation of the true transition probability and allows for a simpler treatment.

Analytical posterior for the drift coefficient

Noiseless observations

We consider

$$\begin{aligned} \mathrm{d}X_t &= -a\nabla V(X_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t(\omega), \\ X_0 &= x \in \mathbb{R}, \end{aligned}$$

and an EM sequence $\mathcal{X} = \{X_n\}_{n=0}^N$ with time step h. The approximated MLE $\hat{a}_{N,h}$ for a is given by

$$\hat{a}_{N,h} = -\frac{\sum_{n=0}^{N-1} \langle \nabla V(X_n), X_{n+1} - X_n \rangle}{\sum_{n=0}^{N-1} |\nabla V(X_n)|^2\, h}.$$

In the Bayesian framework, if no noise is added to the observations $\mathcal{X}$ and for a prior $a \sim \mu_0$ with density $P(a)$ we have

$$P(a \mid \mathcal{X}) \propto P(\mathcal{X} \mid a)\, P(a).$$

Let us consider $\mu_0 = \mathcal{U}(A)$, where $A = \{\text{admissible values for } a\}$. Then, maximising $P(a \mid \mathcal{X})$ is equivalent to maximising $P(\mathcal{X} \mid a)$ and therefore the MAP estimator $\tilde{a}_{N,h}$ is equal to the MLE estimator $\hat{a}_{N,h}$. For a Gaussian prior $\mu_0 = \mathcal{N}(\mu_a, \sigma_a^2)$, we get

$$\tilde{a}_{N,h} = -\frac{\sigma_a^2 \beta \sum_{n=0}^{N-1} \langle \nabla V(X_n), X_{n+1} - X_n \rangle + \mu_a}{\sigma_a^2 \beta \sum_{n=0}^{N-1} |\nabla V(X_n)|^2\, h - 1}.$$

Hence, in the limit $N \to \infty$, we have, independently of μ_a and σ_a,

$$\lim_{N\to\infty} \tilde{a}_{N,h} = \lim_{N\to\infty} \hat{a}_{N,h}.$$

Noisy observations

We first state a known result on Gaussian densities we will employ extensively in the computations below.

Lemma 1. *Let $p(x) = \mathcal{N}(x; \mu_p, \sigma_p^2)$ and $q(x) = \mathcal{N}(x; \mu_q, \sigma_q^2)$. Then*

$$p(x)q(x) = Z\,\mathcal{N}(x; \mu_{pq}, \sigma_{pq}^2),$$

where

$$\mu_{pq} = \frac{\mu_p \sigma_q^2 + \mu_q \sigma^2}{\sigma_p^2 + \sigma_q^2}, \quad \sigma_{pq}^2 = \frac{\sigma_p^2 \sigma_q^2}{\sigma_p^2 + \sigma_q^2},$$

and where

$$Z = \mathcal{N}(\mu_p; \mu_q, \sigma_p^2 + \sigma_q^2).$$

Proof. Find a reliable reference. □

Moreover, we will use the relation valid for any $c \in \mathbb{R}$

$$\mathcal{N}(cx; \mu, \sigma^2) = \frac{1}{c}\mathcal{N}(x; c^{-1}\mu, c^{-2}\sigma^2). \tag{4}$$

In the following, we omit all constants which are independent of $\mathcal{X}_N$ and a. Let us compute the posterior for one single observation Y_{t_1}, where $t_1 = h$. In this case, the transition probability is given by $p(X_{t_1} \mid X_{t_0}, a) = \mathcal{N}(X_{t_1}; (1-ah)X_{t_0}, 2\Sigma h)$, and the observation likelihood is $p(Y_{t_1} \mid X_{t_1}, a) = \mathcal{N}(Y_{t_1}; X_{t_1}, 2\sigma_\eta^2)$, which by symmetry can be rewritten as $p(Y_{t_1} \mid X_{t_1}, a) = \mathcal{N}(X_{t_1}; Y_{t_1}, 2\sigma_\eta^2)$. Thanks to Lemma 1 we have

$$\mathcal{N}(X_{t_1}; (1-ah)X_{t_0}, 2\Sigma h)\, \mathcal{N}(X_{t_1}; Y_{t_1}, \sigma_\eta^2) = Z_0(a)\, \mathcal{N}(X_{t_1}; \mu_1, \sigma_1^2),$$

where μ_1 and σ_1^2 are given by

$$\begin{aligned} \mu_1 &= \frac{(1-ah)X_{t_0}\sigma_\eta^2 + 2Y_{t_1}\Sigma h}{2\Sigma h + \sigma_\eta^2}, \\ \sigma_1^2 &= \frac{2\Sigma h \sigma_\eta^2}{2\Sigma h + \sigma_\eta^2}, \end{aligned}$$

and where $Z_0(a)$ reads

$$Z_0(a) = \mathcal{N}((1-ah)X_{t_0}; Y_{t_1}, \sigma_\eta^2 + 2\Sigma h),$$

which can be rewritten using (4) as

$$Z_0(a) = \frac{1}{1-ah}\mathcal{N}\big(X_{t_0}; (1-ah)^{-1}Y_{t_1}, (1-ah)^{-2}(\sigma_\eta^2 + 2\Sigma h)\big) \tag{5}$$

In general, in the following we will write

$$\begin{aligned} \mu_j &= \frac{(1-ah)X_{t_{j-1}}\sigma_\eta^2 + 2Y_{t_j}\Sigma h}{2\Sigma h + \sigma_\eta^2}, \\ \sigma_j^2 &= \frac{2\Sigma h \sigma_\eta^2}{2\Sigma h + \sigma_\eta^2}, \end{aligned}$$

and

$$\begin{aligned} Z_j(a) &= \mathcal{N}((1-ah)X_{t_j}; Y_{t_{j+1}}, \sigma_\eta^2 + 2\Sigma h) \\ &= \frac{1}{1-ah}\mathcal{N}\big(X_{t_j}; (1-ah)^{-1}Y_{t_{j+1}}, (1-ah)^{-2}(\sigma_\eta^2 + 2\Sigma h)\big). \end{aligned}$$

Finally, we get

$$\begin{aligned} p(a \mid \mathcal{Y}_1) &= \int p(a, \mathcal{X}_N \mid \mathcal{Y}_N)\, \mathrm{d}\mathcal{X}_N \\ &\propto \int p(a)\, p(X_{t_1} \mid X_{t_0}, a)\, p(Y_{t_1} \mid X_{t_0}, a)\, \mathrm{d}X_{t_1} \\ &= p(a)\, Z_0(a) \int \mathcal{N}(X_{t_1}; \mu_1, \sigma_1^2)\, \mathrm{d}X_{t_1} = p(a)\, Z_0(a), \end{aligned}$$

and the marginal posterior on the drift coefficient is therefore given in closed form. Let us now consider a second observation Y_{t_2} at time $t_2 = 2h$. In this case and thanks to the computations above, we have

$$p(a \mid \mathcal{Y}_2) \propto p(a)\, Z_0(a) \int \mathcal{N}(X_{t_1}; \mu_1, \sigma_1^2)\, Z_1(a)\, \mathcal{N}(X_{t_2}; \mu_2, \sigma_2^2)\, \mathrm{d}X_{t_2}\mathrm{d}X_{t_1}. \tag{6}$$

Let us compute the product $\mathcal{N}(X_{t_1}; \mu_1, \sigma_1^2)\, Z_1(a)$. Thanks to the relation (5), we have

$$\mathcal{N}(X_{t_1}; \mu_1, \sigma_1^2)\, Z_1(a) \propto \frac{1}{1-ah}\mathcal{N}(X_{t_1}; \mu_1, \sigma_1^2)\, \mathcal{N}\big(X_{t_1}; (1-ah)^{-1}Y_{t_2}, (1-ah)^{-2}(\sigma_\eta^2 + 2\Sigma h)\big),$$

which gives thanks to Lemma 1

$$\mathcal{N}(X_{t_1};\mu_1,\sigma_1^2)\,Z_1(a) \propto \widetilde{Z}_1(a)\,\mathcal{N}(X_{t_1};\tilde{\mu}_1,\tilde{\sigma}_1^2),$$

where

$$\tilde{\mu}_1 = \frac{\mu_1(1-ah)^{-2}(\sigma_\eta^2+2\Sigma h)+(1-ah)^{-1}Y_{t_2}\sigma_1^2}{\sigma_1^2+(1-ah)^{-2}(\sigma_\eta^2+2\Sigma h)},$$
$$\tilde{\sigma}_1^2 = \frac{\sigma_1^2(1-ah)^{-2}(\sigma_\eta^2+2\Sigma h)}{\sigma_1^2+(1-ah)^{-2}(\sigma_\eta^2+2\Sigma h)},$$

and where

$$\widetilde{Z}_1(a) = \frac{1}{1-ah}\mathcal{N}(\mu_1;(1-ah)^{-1}Y_{t_2},\sigma_1^2+(1-ah)^{-2}(\sigma_\eta^2+2\Sigma h))$$

Replacing back into (6), one gets

$$\begin{aligned} p(a \mid \mathcal{Y}_2) &\propto p(a)\,Z_0(a)\,\widetilde{Z}_1(a)\int \mathcal{N}(X_{t_1};\tilde{\mu}_1,\tilde{\sigma}_1^2)\int \mathcal{N}(X_{t_2};\mu_2,\sigma_2^2)\,\mathrm{d}X_{t_2}\,\mathrm{d}X_{t_1} \\ &= p(a)\,Z_0(a)\,\widetilde{Z}_1(a). \end{aligned}$$

A recursive argument gives now the following result.

Theorem 1. Maybe *The posterior distribution on the drift coefficient corresponding to N observations is given by*

$$p(a \mid \mathcal{Y}_N) \propto p(a)\,Z_0(a)\prod_{j=1}^{N-1}\widetilde{Z}_j(a).$$

NO!

Proof. The result is obtained through an inductive argument. In particular, we first show

$$p(a,\mathcal{X}_N \mid \mathcal{Y}_N) \propto p(a)\,Z_0(a)\Big(\prod_{j=1}^{N-1}\widetilde{Z}_j(a)\Big)\mathcal{N}(X_{t_N};\mu_N,\sigma_N^2)\prod_{j=1}^{N-1}\mathcal{N}(X_{t_j};\tilde{\mu}_j,\tilde{\sigma}_j^2). \tag{7}$$

The base cases $N=1$ and $N=2$ are treated above. Let us suppose the equality above to be true for $\mathcal{X}_{N-1}$ and $\mathcal{Y}_{N-1}$. Then

$$\begin{aligned} p(a,\mathcal{X}_N \mid \mathcal{Y}_N) &\propto p(a)\prod_{j=1}^{N} p(X_{t_j} \mid X_{t_{j-1}},a)\prod_{j=1}^{N} p(Y_{t_j} \mid X_{t_j},a) \\ &= p(a,\mathcal{X}_{N-1} \mid \mathcal{Y}_{N-1})\,p(X_{t_N} \mid X_{t_{N-1}})\,p(Y_{t_N} \mid X_{t_N},a) \\ &\propto p(a)\,Z_0(a)\Big(\prod_{j=1}^{N-2}\widetilde{Z}_j(a)\Big)\mathcal{N}(X_{t_{N-1}};\mu_{N-1},\sigma_{N-1}^2)\prod_{j=1}^{N-2}\mathcal{N}(X_{t_j};\tilde{\mu}_j,\tilde{\sigma}_j^2) \\ &\quad\times Z_{N-1}(a)\,\mathcal{N}(X_{t_N};\mu_N,\sigma_N^2). \end{aligned}$$

Computations similar to the case $N=2$ lead to

$$Z_{N-1}(a)\,\mathcal{N}(X_{t_{N-1}};\mu_{N-1},\sigma_{N-1}^2) = \widetilde{Z}_{N-1}(a)\,\mathcal{N}(X_{t_{N-1}};\tilde{\mu}_{N-1},\tilde{\sigma}_{N-1}^2),$$

which shows (7). Let us now remark that for each j, $\tilde{Z}_j$ depends only on μ_j, which in turn depends only on $X_{t_{j-1}}$. Hence,

$$
\begin{aligned}
\frac{p(a \mid \mathcal{Y}_N)}{p(a)\, Z_0(a)} &\propto \int \Big(\prod_{j=1}^{N-1} \widetilde{Z}_j(a) \Big) \mathcal{N}(X_{t_N}; \mu_N, \sigma_N^2) \prod_{j=1}^{N-1} \mathcal{N}(X_{t_1}; \tilde{\mu}_j, \tilde{\sigma}_j^2)\, \mathrm{d}\mathcal{X}_N \\
&= \int \cdots \int \Big(\prod_{j=1}^{N-1} \widetilde{Z}_j(a) \Big) \prod_{j=1}^{N-1} \mathcal{N}(X_{t_j}; \tilde{\mu}_j, \tilde{\sigma}_j^2) \int \mathcal{N}(X_{t_N}; \mu_N, \sigma_N^2)\, \mathrm{d}X_{t_N}\, \mathrm{d}\mathcal{X}_{N-1} \\
&= \int \cdots \int \Big(\prod_{j=1}^{N-1} \widetilde{Z}_j(a) \Big) \prod_{j=1}^{N-2} \mathcal{N}(X_{t_j}; \tilde{\mu}_j, \tilde{\sigma}_j^2) \int \mathcal{N}(X_{t_{N-1}}; \tilde{\mu}_{N-1}, \tilde{\sigma}_{N-1}^2)\, \mathrm{d}X_{t_{N-1}}\, \mathrm{d}\mathcal{X}_{N-2} \\
&= \int \cdots \int \Big(\prod_{j=1}^{N-2} \widetilde{Z}_j(a) \Big) \prod_{j=1}^{N-3} \mathcal{N}(X_{t_j}; \tilde{\mu}_j, \tilde{\sigma}_j^2) \int \widetilde{Z}_{N-1}(a)\, \mathcal{N}(X_{t_{N-2}}; \tilde{\mu}_{N-2}, \tilde{\sigma}_{N-1}^2)\, \mathrm{d}X_{t_{N-2}}\, \mathrm{d}\mathcal{X}_{N-3}
\end{aligned}
$$

The recursion looks worse than I thought. $\square$

Monte Carlo approach

Explain strategy (Particle Filter).

First experiment. Parameters: $T = 10$, $\sigma_\eta = 0.1$, $M = 400$, $h =$

[**?**, Chapter 1]