Posterior distribution

Jan van Waaij

April 30, 2021

1 Distribution of the posterior of a finite basis expansion with Gaussian coefficients

Lemma 1. *Let $X^T = (X_t : t \in [0, T])$ be an observation of*

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t,$$

where b is equipped with the prior distribution defined by

$$b = \sum_{j=1}^{k} \theta_j \phi_j,$$

where $\{\phi_1, \ldots, \phi_k\}$ is a linearly independent basis, and $\theta = (\theta_1, \ldots, \theta_k)^t$ has multivariate normal distribution $N(\mu, \Sigma)$, with mean vector μ and positive definite matrix Σ, and $\sigma : \mathbb{R} \to \mathbb{R}_{>0}$ is a measurable function. Then the posterior distribution of θ given X^T is $N(\hat{\mu}, \hat{\Sigma})$, where

$$\hat{\mu} = (S + \Sigma^{-1})^{-1}(m + \Sigma^{-1}\mu), \quad \hat{\Sigma} = (S + \Sigma^{-1})^{-1}$$

and the vector $m = (m_1, \ldots, m_k)^t$ is defined by

$$m_l = \int_0^T \frac{\phi_l(X_t)}{\sigma(X_t)^2} dX_t, \quad l = 1, \ldots, k,$$

and the symmetric $k \times k$-matrix S is given by

$$S_{l,l'} = \int_0^T \frac{\phi_l(X_t)\phi_{l'}(X_t)}{\sigma^2(X_t)} dt, \quad l, l' = 1, \ldots, k, \tag{1}$$

provided $S + \Sigma^{-1}$ is invertible. Moreover, the marginal likelihood is given by

$$\int p(X^T \mid \theta)p(\theta)d\theta = |\Sigma^{-1}\hat{\Sigma}|^{1/2} e^{-\frac{1}{2}\mu^t\Sigma^{-1}\mu} e^{\frac{1}{2}\hat{\mu}^t\hat{\Sigma}^{-1}\hat{\mu}}.$$

Proof. Almost surely we have by Girsanov's theorem (e.g. Steele, 2001, chapter 13 or Chung and Williams, 1990 reprint 2014, section 9.4)

$$p(X^T \mid \theta) = \exp\left(\int_0^T \frac{b(X_t)}{\sigma(X_t)^2} dX_t - \frac{1}{2} \int_0^T \left(\frac{b(X_t)}{\sigma(X_t)} \right)^2 dt \right), \tag{2}$$

with respect to the Wiener measure. So

$$\log p(X^T \mid b) = \theta^t m - \frac{1}{2}\theta^t S \theta \tag{3}$$

and the log of the distribution of θ with respect to the Lebesgue measure on $\mathbb{R}^k$ is given by

$$\begin{aligned}\log p(\theta) =& -\frac{k}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(\theta-\mu)^t\Sigma^{-1}(\theta-\mu)\\ =& C_1 - \frac{1}{2}\theta\Sigma^{-1}\theta + \theta^t\Sigma^{-1}\mu,\end{aligned}$$

with

$$C_1 = -\frac{k}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}\mu^t\Sigma^{-1}\mu.$$

So,

$$\begin{aligned}\log(p(X^T \mid \theta)p(\theta)) =& C_1 + \theta^t m - \frac{1}{2}\theta^t S\theta - \frac{1}{2}\theta\Sigma^{-1}\theta + \theta^t\Sigma^{-1}\mu\\ =& C_1 + \theta^t(m + \Sigma^{-1}\mu) - \frac{1}{2}\theta^t(S+\Sigma^{-1})\theta\\ =& C_1 + \theta^t(S+\Sigma^{-1})\Big((S+\Sigma^{-1})^{-1}(m+\Sigma^{-1}\mu)\Big)\\ & - \frac{1}{2}\theta^t(S+\Sigma^{-1})\theta.\end{aligned}$$

By the Bayes formula, the posterior density of θ is proportional to $p(X^T \mid \theta)p(\theta)$. It follows that $\theta \mid X^T$ is normally distributed with mean

$$\hat{\mu} := (S+\Sigma^{-1})^{-1}(m+\Sigma^{-1}\mu).$$

and covariance matrix

$$\hat{\Sigma} := (S+\Sigma^{-1})^{-1},$$

provided $S+\Sigma^{-1}$ is invertible. Moreover

$$\begin{aligned}&\int p(X^T \mid \theta)p(\theta)d\theta\\ =&\int e^{C_1}e^{\theta^t\hat{\Sigma}^{-1}\hat{\mu}}e^{-\frac{1}{2}\theta^t\hat{\Sigma}^{-1}\theta}d\theta\\ =&(2\pi)^{k/2}|\hat{\Sigma}|^{1/2}e^{\frac{1}{2}\hat{\mu}^t\hat{\Sigma}^{-1}\hat{\mu}}e^{C_1}\\ &\times\int(2\pi)^{-k/2}|\hat{\Sigma}|^{-1/2}e^{\theta^t\hat{\Sigma}^{-1}\hat{\mu}}e^{-\frac{1}{2}\theta^t\hat{\Sigma}^{-1}\theta}e^{-\frac{1}{2}\hat{\mu}^t\hat{\Sigma}^{-1}\hat{\mu}}d\theta\\ =&(2\pi)^{k/2}|\hat{\Sigma}|^{1/2}e^{\frac{1}{2}\hat{\mu}^t\hat{\Sigma}^{-1}\hat{\mu}}e^{C_1}\\ =&|\Sigma^{-1}\hat{\Sigma}|^{1/2}e^{-\frac{1}{2}\mu^t\Sigma^{-1}\mu}e^{\frac{1}{2}\hat{\mu}^t\hat{\Sigma}^{-1}\hat{\mu}},\end{aligned}$$

using that the integrant in the third last line is the density of a multivariate normal distribution and therefore integrates to one. □

2 The marginal maximum likelihood estimator

Lemma 2. *Let $\lambda > 0$, $\mu \in \mathbb{R}^k$ and let Σ be a positive definite $k \times k$-matrix. Consider the prior $\theta \sim N(\mu, \Sigma_\lambda)$, where $\Sigma_\lambda = \lambda^2\Sigma$ and denote its density by p_λ. Then*

$$\begin{aligned}&\log\int p_\lambda(X^T \mid \theta)p_\lambda(\theta)d\theta\\ =& -\frac{1}{2}\log|\lambda^2\Sigma S + \mathbb{I}_k| - \frac{1}{2}\mu^t\Sigma^{-1}\mu + \frac{1}{2}(m+\lambda^{-2}\Sigma^{-1}\mu)^t(S+\lambda^{-2}\Sigma^{-1})^{-1}(m+\lambda^{-2}\Sigma^{-1}\mu).\end{aligned}\tag{4}$$

Proof. It follows from lemma 1 that

$$\Sigma_\lambda \hat{\Sigma}_\lambda^{-1} = \Sigma_\lambda (S + \Sigma_\lambda^{-1}) = \Sigma_\lambda S + \mathbb{I}_k = \lambda^2 \Sigma S + \mathbb{I}_k$$

and

$$\begin{aligned}\hat{\mu}^t \hat{\Sigma}_\lambda^{-1} \hat{\mu} =& (m + \Sigma_\lambda^{-1}\mu)^t (S + \Sigma_\lambda^{-1})^{-1} (S + \Sigma_\lambda^{-1}) (S + \Sigma_\lambda^{-1})^{-1} (m + \Sigma_\lambda^{-1}\mu) \\ =& (m + \lambda^{-2}\Sigma^{-1}\mu)^t (S + \lambda^{-2}\Sigma^{-1})^{-1} (m + \lambda^{-2}\Sigma^{-1}\mu).\end{aligned}$$

So it follows from the same lemma that

$$\begin{aligned}&\log \int p_\lambda(X^T \mid \theta) p_\lambda(\theta) d\theta \\ =& -\frac{1}{2}\log|\lambda^2 \Sigma S + \mathbb{I}_k| - \frac{1}{2}\mu^t \Sigma^{-1} \mu + \frac{1}{2}(m + \lambda^{-2}\Sigma^{-1}\mu)^t (S + \lambda^{-2}\Sigma^{-1})^{-1} (m + \lambda^{-2}\Sigma^{-1}\mu).\end{aligned}$$

□

3 Random scaling

Lemma 3. *Let $X^T = (X_t : t \in [0, T])$ be an observation of*

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t,$$

where b is equipped with the prior distribution defined by

$$\begin{aligned}\lambda^2 \sim& \textit{Inverse Gamma}(A, B) = IG(A, B) \\ \theta \mid \lambda \sim& N(\mu, \lambda^2 \Sigma) \\ b \mid \theta =& \sum_{j=1}^{k} \theta_j \phi_j,\end{aligned}$$

where $\{\phi_1, \ldots, \phi_k\}$ is a linearly independent basis. Then

$$\lambda^2 \mid \theta, X^T \sim IG\left(A + k/2, B + \frac{1}{2}(\theta - \mu)^t \Sigma^{-1} (\theta - \mu)\right).$$

Proof. Recall eq. (3), $\log p(X^T \mid b) = \theta^t m - \frac{1}{2}\theta^t S \theta$. The logarithm of the distribution of θ given λ with respect to the Lebesgue measure on $\mathbb{R}^k$ is given by (proportionality w.r.t. λ),

$$\log p(\theta \mid \lambda) = C_1 - k \log \lambda - \frac{1}{2}\lambda^{-2} (\theta - \mu)^t \Sigma^{-1} (\theta - \mu).$$

for some real constant C_1, depending on θ, but not on λ.

In the following, $\propto$ means equal up to a multiplicative constant depending on θ and X^T, but not on λ. By the Bayes formula,

$$p(\lambda^2 \mid \theta, X^T) \propto p(X^T \mid \lambda^2, \theta) p(\lambda^2 \mid \theta)$$

and

$$p(\lambda^2 \mid \theta) \propto p(\theta \mid \lambda^2) p(\lambda^2)$$

so

$$p(\lambda^2 \mid \theta, X^T) \propto p(X^T \mid \lambda^2, \theta) p(\theta \mid \lambda^2) p(\lambda^2).$$

It follows that for some real constants $C, \tilde{C}$ depending on θ and X^T, but not on λ, we have

$$\begin{aligned}
&\log p(\lambda^2 \mid \theta, X^T) \\
=&C + \theta^t m - \frac{1}{2}\theta^t S\theta \\
&- k\log\lambda - \frac{1}{2}\lambda^{-2}(\theta-\mu)^t\Sigma^{-1}(\theta-\mu) \\
&- (A+1)\log(\lambda^2) - \frac{B}{\lambda^2} \\
=&\tilde{C} - (A+k/2+1)\log(\lambda^2) - \frac{B + \frac{1}{2}(\theta-\mu)^t\Sigma^{-1}(\theta-\mu)}{\lambda^2},
\end{aligned}$$

which is up to an additive constant the logarithm of the density of the inverse gamma distribution with shape parameter $A + k/2$ and scale parameter $B + \frac{1}{2}(\theta-\mu)^t\Sigma^{-1}(\theta-\mu)$. □

Lemma 4. *We have*

$$\begin{aligned}
&\log p(X^T \mid j, \lambda^2) \\
=& -\frac{1}{2}\log|\lambda^2\Sigma S + \mathbb{I}_k| - \frac{1}{2}\mu^t\Sigma^{-1}\mu + \frac{1}{2}(m + \lambda^{-2}\Sigma^{-1}\mu)^t(S + \lambda^{-2}\Sigma^{-1})^{-1}(m + \lambda^{-2}\Sigma^{-1}\mu).
\end{aligned}$$

Proof. This follows from

$$p(X^T \mid j, \lambda^2) = \int p(X^T \mid j, \theta^j, \lambda^2)p(\theta^j \mid j, \lambda)d\theta^j$$

and lemma 2. □

4 The sparsity of the Girsanov matrix with Faber-Schauder functions

The Faber-Schauder basis functions $\psi_0, \psi_{j,k}$ are defined as follows:

$$\psi_0(x) = \begin{cases} 1-2x & \text{when } x \in [0,1/2), \\ 2x-1 & \text{when } x \in [1/2,1], \\ 0 & \text{otherwise,} \end{cases}$$

$$\Lambda(x) = \begin{cases} 2x & \text{when } x \in [0,1/2), \\ 2(1-x) & \text{when } x \in [1/2,1], \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\psi_{j,k}(x) = \Lambda(2^j x - k + 1), \quad j = 0,1,\dots, k = 1,\dots,2^j,$$

see van der Meulen, Schauer, and van Waaij, 2018, p. 607. We say that ψ_0 and $\psi_{0,1}$ are of level zero, and the basis functions $\psi_{j,1},\dots,\psi_{j,2^j}$ are said to be of level j. The Girsanov matrix S defined in eq. (1) with all basis function up to and including level J is denoted by S^J. Note that S^J has $2 + \sum_{j=1}^J 2^j = 2^{J+1}$ rows and columns, and 2^{2J+2} entries.

Definition 5. Let M^n be an $n \times n$-matrix, and let $nz(M^n)$ the number of non-zero entries of M^n. The level of sparsity of M^n is the fraction of nonzero entries, $\frac{nz(M^n)}{n^2}$.

The definition of a sparse matrix is vague. Usually, we mean that the number of nonzero entries grows at most linear with the number of rows. We will establish that for S^n, the number of nonzero entries grows at most like $r \log r$ with r the number of rows.

Recall the definition of $S_{l,l'}$ in lemma 2. Note that $S_{l,l'} = 0$ when $\mathrm{SUPP}(\psi_l) \cap \mathrm{SUPP}(\psi_{l'})$ has Lebesgue measure zero. We say that ψ_l and $\psi_{l'}$ have non-overlapping support when their supports are either disjoint or only share a boundary point; otherwise, we say they have overlapping support.

Note that both functions of level zero, ψ_1 and $\psi_{0,1}$, have the same support $[0,1]$.

When $j \geq 0, d \geq 0$ and $d + j \geq 1$, there are 2^d Faber functions of level $j + d$ that have overlapping support with $\psi_{j,k}$, $j \geq 0$. These are

$$\psi_{j+d,(k-1)2^d+1}, \psi_{j+d,(k-1)2^d+2}, \ldots, \psi_{j+d,k2^d}$$

For level 0, there are exactly two, and for level $1, \ldots, j-1$ there is precisely one basis function with overlapping support with $\psi_{j,k}$.

So for ψ_0 and $\psi_{0,1}$ there are

$$2 + \sum_{d=1}^{J} 2^d = 2^{J+1}$$

basis functions $\psi_0, \psi_{j',k'}, j' \leq J$ with overlapping support. For $\psi_{j,k}$, $j \geq 1$, there are

$$2 + j - 1 + \sum_{d=0}^{J-j} 2^d = j + 2^{J-j+1}$$

basis functions $\psi_0, \psi_{j',k'}, j' \leq J$, with overlapping support. When we make use of lemma 6, we see that S^n has at most

$$\begin{aligned}
&2 \cdot 2^{J+1} + \sum_{j=1}^{J} 2^j \left(j + 2^{J-j+1}\right) \\
=&2 \cdot 2^{J+1} + (J-1)2^{J+1} + 2 + J2^{J+1} \\
=&(2J+1)2^{J+1} + 2
\end{aligned}$$

nonzero entries.

So the number of nonzero entries of S^n grows at most like $r \log r$ with r the number of rows. It has level of sparsity at most

$$\frac{(2J+1)2^{J+1} + 2}{2^{2J+2}} = (2J+1)2^{-J-1} + 2^{-2J-1},$$

which is of the order $\frac{\log r}{r}$.

A Lemma

Lemma 6. *For each $J \in \mathbb{N}$,*

$$\sum_{j=1}^{J} j2^j = (J-1)2^{J+1} + 2.$$

Proof. Note that

$$\begin{aligned}\sum_{j=1}^{J} j2^j &= \sum_{j=1}^{J}\sum_{k=j}^{J} 2^k \\ &= \sum_{j=1}^{J} 2^j \sum_{k=0}^{J-j} 2^k \\ &= \sum_{j=1}^{J} 2^j (2^{J-j+1} - 1) \\ &= J2^{J+1} - (2^{J+1} - 2) \\ &= (J-1)2^{J+1} + 2.\end{aligned}$$

$\square$

References

Chung, K.L. and R.J. Williams (1990 reprint 2014). *Introduction to Stochastic Integration*. Modern Birkhäuser Classics. Springer New York. ISBN: 978-1-4614-9587-1. DOI: 10.1007/978-1-4614-9587-1.

Steele, J.M. (2001). *Stochastic Calculus and Financial Applications*. Applications of mathematics : stochastic modelling and applied probability. Springer. ISBN: 9780387950167. DOI: 10.1007/978-1-4684-9305-4.

van der Meulen, F.H., M. Schauer, and J. van Waaij (2018). “Adaptive nonparametric drift estimation for diffusion processes using Faber–Schauder expansions”. In: *Statistical Inference for Stochastic Processes* 21.3, pp. 603–628. DOI: 10.1007/s11203-017-9163-7.