Posterior distribution

Jan van Waaij

April 27, 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 $\{\phi_1, \dots, \phi_k\}$ is a linearly independent basis, and $\theta = (\theta_1, \dots, \theta_k)^t$ 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 $m = (m_1, \dots, m_k)^t$ is defined by

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

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

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

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. $\square$

2 The marginal maximum likelihood estimator

Suppose we have prior $\theta \sim N(\mu, \Sigma_\lambda)$, where $\Sigma_\lambda = \lambda^2\Sigma$. Note 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

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

So

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

Lemma 2. *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 $\{\phi_1,\dots,\phi_k\}$ 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 3. *We have*

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

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 eq. (4). □

3 Number of dependent Faber-Schauder functions with higher or equal index

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

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

see **meulenschauerwaaij2017**. In this question we establish that the Girsanov matrix S, defined by eq. (1) with the Faber-Schauder matrix is sparse.

Definition 4. Let for each n, M^n a $n \times n$ matrix. The sequence $(M^n)_{n\geq 1}$ is sparse when the number of nonzero entries of M^n divided by n^2 converges to zero.

Informally this means that almost all entries of M^n are zero.

Note that $S_{l,l'} = 0$ when $\mathrm{SUPP}(\Psi_l) \cap \mathrm{SUPP}(\Psi_{l'})$ has Lebesgue measure zero. Hier verder!!Note that for level $j \geq 1$, $\psi_{j,k}$ and $\psi_{j,l}$ are only dependent when $k = l$ (obviously, then they are equal).

Note that ψ_1 and $\psi_{0,1}$ are dependent, both of level 0.

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

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

Every Faber-Schauder function is obviously dependent with itself.

Indexing with $i = 2^j + k$, when $\psi_{j,k}$ has index (j, k) (excluding $i = 1$), we see that, when $j \geq 0$, ψ_i is dependent with $2^{j'-j}$, functions $\psi_{j',k'}$, $i' = 2^{j'} + k' \geq i$ of level $j' \geq j$ (including itself, when $j' = j$).

So if J is the higest level, ψ_i is dependent with

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

Faber-Schauder functions $\psi_{i'}$ with index $i' \geq i$. Hence summing over all levels $0, \ldots, J$ and indices within a level, the number of combinations of functions $(\psi_{j,k}, \psi_{j',k'})$, $0 \leq j, j' \leq J$ and $i = 2^j + k \leq 2^{j'} + k' = i'$ which are dependent is

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

The Faber-Schauder function ψ_1 is dependent with every Faber-Schauder function (including itself) up to and including level J, which counts for 2^{J+1} Faber-Schauder functions with a higher index or equal index, up to level J.

In total we have

$$J2^{J+1} + 1 + 2^{J+1} = (J+1)2^{J+1} + 1.$$

Faber-Schauder functions up to level J dependent with a Faber-Schauder function with equal (itself) or higher index.

If we only consider dependent pairs $(\psi_i, \psi_{i'})$ with $i' > i$, then we have

$$J2^{J+1} + 1$$

of such pairs (minus all 2^{J+1} diagonal pairs (ψ_i, ψ_i)).

Hence, by symmetry, there are in total $J2^{J+1} + 1 + J2^{J+1} + 1 + 2^{J+1} = (2J+1)2^{J+1} + 2$ pairs $(\psi_i, \psi_{i'})$ that are dependent.

Lemma 5. *The Girsanov covariantie matrix is sparse.*

Proof. At most $(2J+1)2^{J+1} + 2$ entries of the $2^{J+1} \times 2^{J+1}$-matrix (2^{2J+2} entries) are nonzero. The fraction of nonzero elements is at most

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

which converges to zero. □

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.