Nonparametric Bayesian inference for diffusions

Harry van Zanten (TUE $\rightarrow$ UvA)

(based on work with Yvo Pokern, Andrew Stuart,
Moritz Schauer and Frank van der Meulen)

PEDS II, June 2012, Eindhoven

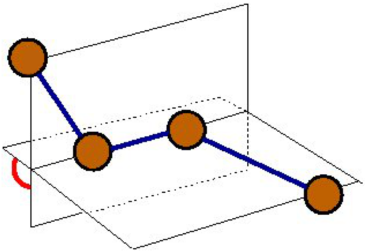
Overview

- Currently available methodology for nonparametric Bayes analysis for SDE's
- Theoretical performance of methods

Data and simulation examples

I: molecular dynamics

Papaspiliopoulos, Pokern, Roberts, Stuart (*Biometrika*, to appear): model dynamics of an angle in a large molecule by an SDE.



Model:

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

Drift function b is 2π-periodic.

Bayesian approach: have to put a prior on, for instance, $L^2(\mathbb{T})$.

A Gaussian random element W in $L^2(\mathbb{T})$ is determined by its mean and its covariance operator $\mathcal{C}_0$. We have

$$\mathrm{Cov}\Big(\int Wf \int Wg\Big) = \int f(\mathcal{C}_0 g).$$

Always: $\mathcal{C}_0$ is trace-class, self-adjoint.

PPRS (2011) take as prior the law of a centered Gaussian in $L^2(\mathbb{T})$ with precision operator

$$\mathcal{C}_0^{-1} = \eta((-\Delta)^p + \kappa I),$$

with $p = 2$, $\eta = .02$ and $\kappa = 0$.

Data: essentially viewed as a continuous-time signal.

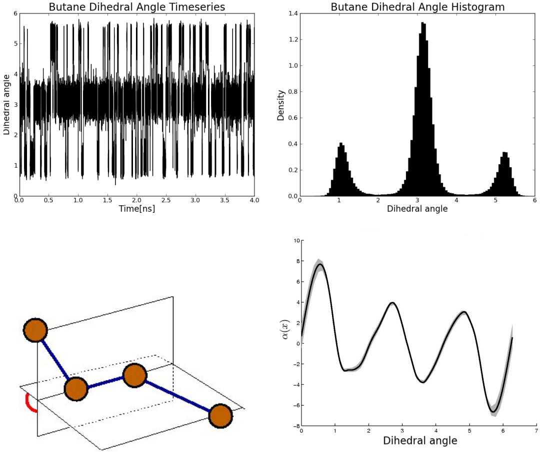
It can be shown that the posterior is again Gaussian.

Posterior mean solves an ordinary differential equation.

Posterior precision (inverse covariance) is again a differential operator.

Can use numerical differential equations methods to sample posterior.

Example of posterior computation:



(Figures from PPRS (2011))

II: hierarchical series priors

The prior of PPRS (2011) is computationally convenient, but also has some disadvantages.

In particular: using fixed values for hyper parameters increases the risk that the prior does not properly reflect the properties of the true drift.

Series representation the prior of PPRS (2011):

Distribution of

$$W(x) = \sum_{i=1}^{\infty} \sqrt{\lambda_i} Z_i \psi_i(x),$$

where the Z_i are independent, standard normal variables,

$$\lambda_i = \left(\eta\left(4\pi^2\left\lceil\frac{i}{2}\right\rceil^2\right)^p + \eta\kappa\right)^{-1}$$

and (ψ_i) is the standard Fourier basis of $L^2(\mathbb{T})$.

Alternative approach (vdMSvZ (in preparation)): truncate the sum at a random point and use a random scaling constant.

Hierarchical prior specification:

$$\boxed{\begin{aligned} j &\sim p(j), \\ s^2 &\sim IG(a,b), \\ \theta^j \,|\, j, s^2 &\sim N_j(0, s^2 \Xi^j), \\ b \,|\, j, s^2, \theta^j &\sim \sum_{i=1}^{j} \theta_i^j \psi_i. \end{aligned}}$$

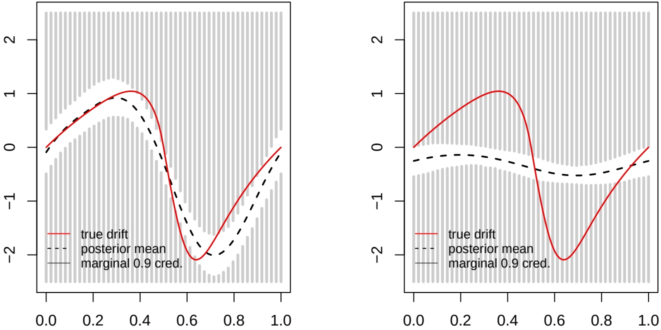
Explicit posterior computations now no longer possible.

Can however devise an efficient MCMC algorithm to sample from the posterior.

Ingredients in the algorithm:

- Within models (fixed j) have partial conjugacy and do Gibbs sampling.
- Use reversible jump MCMC to jump between models.
- In case of low-frequency data add Metropolis-Hastings step for sampling diffusion bridges to achieve data augmentation.

Example: effect of a prior on the scale.



Simulated diffusion data on $[0, 100]$, using the red curve as drift function. Left: hierarchical, right: fixed scale.

Theoretical results for the differential equations approach

I: Computation of the posterior

Recall: model is $dX_t = b(X_t)\,dt + dW_t$, with b 1-periodic. Prior is the centered Gaussian distribution on $L^2(\mathbb{T})$ with inverse covariance operator

$$\mathcal{C}_0^{-1} = \eta((-\Delta)^p + \kappa I),$$

Suppose that we observe the diffusion on the time interval $[0, T]$.

Questions:

- What is the posterior distribution? Is it Gaussian?
- How does the posterior behave as $T \to \infty$? Consistency? Rates?

Thanks to the existence of local time, we can do explicit posterior computations.

In particular: do not need MCMC methods to compute the posterior numerically.

Intermezzo: local time

Let $(L_t(x;X) : t \geq 0, x \in \mathbb{R})$ be the semimartingale local time of X:

$$\int_0^T f(X_u)\,du = \int_{\mathbb{R}} f(x) L_T(x;X)\,dx.$$

In view of periodicity, define the periodic local time

$$L_T^\circ(x;X) = \sum_{k\in\mathbb{Z}} L_T(x+k;X), \quad x \in \mathbb{T}.$$

Occupation times formula:

$$\int_0^T f(X_u)\,du = \int_0^1 f(x) L_T^\circ(x;X)\,dx$$

for 1-periodic f.

Heuristic posterior computation

Likelihood:

$$p(X^T \mid b) = e^{\int_0^T b(X_t)\,dX_t - \frac{1}{2}\int_0^T b^2(X_t)\,dt} = e^{-\Phi_T(b;X)},$$

where

$$\Phi_T(b;X) = \frac{1}{2}\int \Big((b^2 + b')L_T^\circ - 2\chi_T^\circ b\Big).$$

Prior: has "density"

$$p(b) \propto e^{-\frac{1}{2}\int b\mathcal{C}_0^{-1}b}.$$

Heuristic posterior computation

Hence the posterior has "density"

$$
\begin{aligned}
p(b \mid X^T) &\propto p(b)p(X^T \mid b) \\
&\propto \exp\Big(-\frac{1}{2}\int b(\mathcal{C}_0^{-1} + L_T^\circ I)b + \int b(\frac{1}{2}(L_T^\circ)' + \chi_T^\circ)\Big).
\end{aligned}
$$

This suggests that the posterior is Gaussian again, with precision operator

$$
\mathcal{C}_T^{-1} = \mathcal{C}_0^{-1} + L_T^\circ I
$$

and mean $\hat{b}_T$ satisfying

$$
\mathcal{C}_T^{-1}\hat{b}_T = \frac{1}{2}(L_T^\circ)' + \chi_T^\circ.
$$

Weak formulation of posterior mean equation

For $u = \hat{b}_T$, have equation

$$\eta(-1)^p u^{(2p)} + \kappa u + L_T^\circ u = \frac{1}{2}(L_T^\circ)' + \chi_T^\circ$$

Multiplication with a smooth test function v and integrating by parts gives the weak formulation

$$a(u, v; X) = r(v; X), \qquad \forall\, v \in \dot{H}^p(\mathbb{T}), \tag{1}$$

where

$$a(u, v; X) = \eta \int u^{(p)} v^{(p)} + \int (\kappa + L_T^\circ) uv,$$
$$r(v; X) = -\frac{1}{2} \int L_T^\circ v' + \int \chi_T^\circ v.$$

Fact: (1) has a unique solution in $\dot{H}^p(\mathbb{T})$.

Posterior: precise result

Theorem.
A.s., the posterior is the Gaussian measure on $L^2(\mathbb{T})$ with covariance operator $\mathcal{C}_T$ and mean $\hat{b}_T$ which is the unique solution of (1).

Elements of the proof.

- First consider finite-dimensional approximation of posterior by projecting on span of first n eigenfunctions of $\mathcal{C}_0$.
- Using variational analysis ideas to show that the covariance operator and mean of the finite-dim approximations convergence in an appropriate way.

II: Consistency and rates

Bayesian asymptotics for diffusion models

Interested in frequentist asymptotics for Bayes procedures.

There exists all kinds of general results on consistency and convergence rates for nonparametric Bayes procedures.

They avoid explicit posterior computations, but typically involve conditions on the small ball probabilities of the prior and on the metric entropy of (appropriate subsets of) the support of the prior.

For diffusion models: see Van der Meulen, Van der Vaart and vZ ('06), Panzar and vZ ('09).

In this case we can circumvent the general approach and exploit the explicit characterization of the posterior.

Result for the posterior mean

Let b_0 be the true drift function (1-periodic, mean zero, C^1). Let ρ_0 be the probability density on $[0,1]$ given by

$$\rho_0(x) \propto e^{2\int_0^x b_0(y)\,dy}.$$

Theorem.
If $b_0 \in \dot{H}^p(\mathbb{T})$, then for all $s \in (0, 1/2)$, with probability tending to 1,

$$\|\hat{b}_T - b_0\|_{L^2} \lesssim O_{P_0}\left(\frac{1}{\sqrt{T}}\right) + T^{-\frac{p-(1-s)}{2p}} \|\mathbb{G}_T\|_{H^s},$$

where

$$\mathbb{G}_T(x) = \sqrt{T}\left(\frac{L_T^\circ(x;X)}{T} - \rho_0(x)\right).$$

Intermezzo: local time again

Bolthausen ('79) for Brownian motion on the circle:

$$\mathbb{G}_T = \sqrt{T}\Big(\frac{L_T^\circ(\cdot; X)}{T} - 1\Big) \Rightarrow \textit{Gaussian limit}$$

in $C[0,1]$ as $T \to \infty$.

We prove:

Theorem.
If $b_0 \in \dot{C}(\mathbb{T})$, then

- $\|L_T^\circ/T - \rho_0\|_\infty \stackrel{\text{as}}{\to} 0.$
- $\mathbb{G}_T$ is asymptotically tight in $H^s(\mathbb{T})$ for all $s \in (0, 1/2)$.

We use different arguments than Bolthausen.

Proof ingredients:

- Write i.i.d. representation for local time up to winding times.
- Use LLN's and CLT's for Banach space-valued random variables.
- Deal with the bit after the last winding time separately.

Back to posterior consistency and rates

For the posterior mean:

Theorem.
If $b_0 \in \dot{H}^p(\mathbb{T})$, then for all $\delta > 0$

$$\|\hat{b}_T - b_0\|_{L^2} = O_{P_0}\left(T^{-\frac{p-1/2}{2p}+\delta}\right).$$

For the whole posterior:

Theorem.
If $b_0 \in \dot{H}^p(\mathbb{T})$, then for all $\delta > 0$

$$\Pi(b : \|b - b_0\|_{L^2} \geq M_n T^{-\frac{p-1/2}{2p}+\delta} \,|\, X^T) \xrightarrow{P_0} 0$$

for all $M_n \to \infty$.

Remarks

- The prior essentially has Sobolev regularity $p - 1/2$. The rate $T^{-\frac{p-1/2}{2p}}$ is exactly the usual $T^{-\beta/(1+2\beta)}$ for $\beta = p - 1/2$.
- We expect/hope the rate result to be true under the weaker condition $b_0 \in \dot{H}^{p-1/2}(\mathbb{T})$.
- We can not derive the result under this minimal assumption using our direct differential equations approach.
- We believe that using this Gaussian prior, an optimal convergence rate can only be attained if the regularity of the true b_0 matches the regularity of the prior $(p - 1/2)$.

Conclusions and future directions

Summary

- Have working methods for doing BNP for one-dimensional SDE models.
- Have some theory on performance.
- Gaussian priors: consistency, but can have sub-optimal rates.

Further questions

- Find minimal conditions for optimal convergence rates.
- Study asymptotics for hierarchical priors. Adaptation?
- More generally: find numerically feasible methods with good/optimal theoretical properties (adaptation, quality of credible sets).

- Deal with unknown diffusion functions effectively.
- Asymptotics for low-frequency observations.
- Develop methods and theory for multi-dimensional diffusions.

Further questions

- Find minimal conditions for optimal convergence rates.
- Study asymptotics for hierarchical priors. Adaptation?
- More generally: find numerically feasible methods with good/optimal theoretical properties (adaptation, quality of credible sets).

- Deal with unknown diffusion functions effectively.
- Asymptotics for low-frequency observations.
- Develop methods and theory for multi-dimensional diffusions.

THANKS!