Non-parametric Bayesian estimation of a diffusion coefficient

joint work with Shota Gugushvili (Leiden),
Frank van der Meulen (Delft) and Moritz Schauer (Leiden)

Peter Spreij

Universiteit van Amsterdam

Radboud Universiteit

Mathematics of Risk, MATRIX, Creswick, 1 December 2017

Outline

Inference for SDEs

- Model equation for SDE:

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

- Statistical inference for SDEs: long studied and very active field of research, far from saturation.
- Parametric methods: e.g., Kutoyants (2004), Iacus (2008) and references therein.
- Non-parametric methods (for diffusion coefficient): frequentist approaches e.g. in Genon-Catalot et. al (1992), Hoffmann (1997) and Soulier (1998); Bayesian approaches in Gugushvili and Spreij (2014, 2016) and Nickl and Söhl (2015).

Inference for SDEs

- Model equation for SDE:

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

- Statistical inference for SDEs: long studied and very active field of research, far from saturation.
- Parametric methods: e.g., Kutoyants (2004), Iacus (2008) and references therein.
- Non-parametric methods (for diffusion coefficient): frequentist approaches e.g. in Genon-Catalot et. al (1992), Hoffmann (1997) and Soulier (1998); Bayesian approaches in Gugushvili and Spreij (2014, 2016) and Nickl and Söhl (2015).

Inference for SDEs

- Model equation for SDE:

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

- Statistical inference for SDEs: long studied and very active field of research, far from saturation.
- Parametric methods: e.g., Kutoyants (2004), Iacus (2008) and references therein.
- Non-parametric methods (for diffusion coefficient): frequentist approaches e.g. in Genon-Catalot et. al (1992), Hoffmann (1997) and Soulier (1998); Bayesian approaches in Gugushvili and Spreij (2014, 2016) and Nickl and Söhl (2015).

Inference for SDEs

- Parametric approaches specify parametric forms for the drift and diffusion coefficients of SDEs. When these specifications use correct functional forms, such methods attain a higher statistical efficiency over the nonparametric ones.
- On the other hand, nonparametric approaches, where one only assumes qualitative features of the drift and diffusion coefficients, guard one against model misspecification, which may have dramatically negative consequences for valid inference.
- Also, nonparametric techniques may suggest plausible parametric models in those situations where these models cannot be derived from the first principles.

Inference for SDEs

- Parametric approaches specify parametric forms for the drift and diffusion coefficients of SDEs. When these specifications use correct functional forms, such methods attain a higher statistical efficiency over the nonparametric ones.
- On the other hand, nonparametric approaches, where one only assumes qualitative features of the drift and diffusion coefficients, guard one against model misspecification, which may have dramatically negative consequences for valid inference.
- Also, nonparametric techniques may suggest plausible parametric models in those situations where these models cannot be derived from the first principles.

Inference for SDEs

- Parametric approaches specify parametric forms for the drift and diffusion coefficients of SDEs. When these specifications use correct functional forms, such methods attain a higher statistical efficiency over the nonparametric ones.
- On the other hand, nonparametric approaches, where one only assumes qualitative features of the drift and diffusion coefficients, guard one against model misspecification, which may have dramatically negative consequences for valid inference.
- Also, nonparametric techniques may suggest plausible parametric models in those situations where these models cannot be derived from the first principles.

Bayesian methods

- Except Gugushvili and Spreij (2014) and Gugushvili and Spreij (2016), there is hardly any other work available on estimation of the dispersion coefficient (or diffusion coefficient) from the nonparametric Bayesian point of view.
- We can mention only a theoretical contribution Nickl and Söhl (2015) and a practically oriented paper Batz et al. (2017), but the models considered there, as well as the sampling scheme, are different from ours, and the theory developed in Nickl and Söhl (2015) does not cover the approach in Batz et al. (2017).

Bayesian methods

- Except Gugushvili and Spreij (2014) and Gugushvili and Spreij (2016), there is hardly any other work available on estimation of the dispersion coefficient (or diffusion coefficient) from the nonparametric Bayesian point of view.
- We can mention only a theoretical contribution Nickl and Söhl (2015) and a practically oriented paper Batz et al. (2017), but the models considered there, as well as the sampling scheme, are different from ours, and the theory developed in Nickl and Söhl (2015) does not cover the approach in Batz et al. (2017).

Why Bayesian methods?

- Recent theoretical and practical advances made in nonparametric Bayesian estimation of the drift coefficient, see, e.g., van der Meulen et al. (2014), Papaspiliopoulos et al. (2012), Pokern et al. (2013), van Waaij and van Zanten (2016), van der Meulen and Schauer (2017) and van Zanten (2013), suggest that comparable results can be obtained for estimation of the diffusion coefficient too.
- On a general level, apart from a philosophical appeal for Bayesians, advantages of a Bayesian approach include automatic quantification of uncertainty in parameter estimates through Bayesian *credible sets*, and the fact that it is a fundamentally *likelihood-based* method.

Why Bayesian methods?

- Recent theoretical and practical advances made in nonparametric Bayesian estimation of the drift coefficient, see, e.g., van der Meulen et al. (2014), Papaspiliopoulos et al. (2012), Pokern et al. (2013), van Waaij and van Zanten (2016), van der Meulen and Schauer (2017) and van Zanten (2013), suggest that comparable results can be obtained for estimation of the diffusion coefficient too.
- On a general level, apart from a philosophical appeal for Bayesians, advantages of a Bayesian approach include automatic quantification of uncertainty in parameter estimates through Bayesian *credible sets*, and the fact that it is a fundamentally *likelihood-based* method.

Outline

SDE model

- Our model is

$$\mathrm{d}X_t = b_0(t, X_t)\,\mathrm{d}t + s_0(t)\,\mathrm{d}W_t, \quad X_0 = x, \quad t \in [0, 1],$$

  with (true) drift b_0, deterministic diffusion coefficient s_0^2.
  smallskip
  Note: we allow nonlinear drift.

- A more general SDE

$$\mathrm{d}X_t = \widetilde{b}_0(t, X_t)\,\mathrm{d}t + s_0(t) f_0(X_t)\,\mathrm{d}W_t,$$

  with s_0 unknown and f_0 known, can be reduced to our model through a simple, known transformation of X (via 'Itô').

- Examples: Ornstein-Uhlenbeck process, geometric Brownian motion with time-dependent coefficients, and many others.

SDE model

- Our model is

$$\mathrm{d}X_t = b_0(t, X_t)\,\mathrm{d}t + s_0(t)\,\mathrm{d}W_t, \quad X_0 = x, \quad t \in [0,1],$$

  with (true) drift b_0, deterministic diffusion coefficient s_0^2.
  smallskip
  Note: we allow nonlinear drift.
- A more general SDE

$$\mathrm{d}X_t = \widetilde{b}_0(t, X_t)\,\mathrm{d}t + s_0(t) f_0(X_t)\,\mathrm{d}W_t,$$

  with s_0 unknown and f_0 known, can be reduced to our model through a simple, known transformation of X (via 'Itô').
- Examples: Ornstein-Uhlenbeck process, geometric Brownian motion with time-dependent coefficients, and many others.

SDE model

- Our model is

$$\mathrm{d}X_t = b_0(t, X_t)\,\mathrm{d}t + s_0(t)\,\mathrm{d}W_t, \quad X_0 = x, \quad t \in [0,1],$$

  with (true) drift b_0, deterministic diffusion coefficient s_0^2.
  smallskip
  Note: we allow nonlinear drift.
- A more general SDE

$$\mathrm{d}X_t = \widetilde{b}_0(t, X_t)\,\mathrm{d}t + s_0(t) f_0(X_t)\,\mathrm{d}W_t,$$

  with s_0 unknown and f_0 known, can be reduced to our model through a simple, known transformation of X (via 'Itô').
- Examples: Ornstein-Uhlenbeck process, geometric Brownian motion with time-dependent coefficients, and many others.

Goal and first ideas

- Assume a discrete-time sample $\{X_{t_{i,n}}\}$ is available, where $t_{i,n} = i/n$, $i = 1, \ldots, n$. We also define $Y_{i,n} = X_{t_{i,n}} - X_{t_{i-1,n}}$, and denote our data by $\mathcal{X}_n$.
- Goal: estimation of the diffusion coefficient s_0^2 using a non-parametric Bayesian approach.
- Under our sampling scheme, consistent estimation of the drift b_0 is impossible. In many contexts, knowledge of the drift coefficient is even of no interest.
- We intentionally *misspecify* the model and act as if the drift were equal to **zero**. Ultimate mathematical justification: Girsanov's theorem (comes later). Misspecification is for facilitation of inference from the computational point of view: we avoid the intractable (true) likelihood.

Goal and first ideas

- Assume a discrete-time sample $\{X_{t_{i,n}}\}$ is available, where $t_{i,n} = i/n$, $i = 1, \ldots, n$. We also define $Y_{i,n} = X_{t_{i,n}} - X_{t_{i-1,n}}$, and denote our data by $\mathcal{X}_n$.
- Goal: estimation of the diffusion coefficient s_0^2 using a non-parametric Bayesian approach.
- Under our sampling scheme, consistent estimation of the drift b_0 is impossible. In many contexts, knowledge of the drift coefficient is even of no interest.
- We intentionally *misspecify* the model and act as if the drift were equal to **zero**. Ultimate mathematical justification: Girsanov's theorem (comes later). Misspecification is for facilitation of inference from the computational point of view: we avoid the intractable (true) likelihood.

Goal and first ideas

- Assume a discrete-time sample $\{X_{t_{i,n}}\}$ is available, where $t_{i,n} = i/n$, $i = 1, \ldots, n$. We also define $Y_{i,n} = X_{t_{i,n}} - X_{t_{i-1,n}}$, and denote our data by $\mathcal{X}_n$.
- Goal: estimation of the diffusion coefficient s_0^2 using a non-parametric Bayesian approach.
- Under our sampling scheme, consistent estimation of the drift b_0 is impossible. In many contexts, knowledge of the drift coefficient is even of no interest.
- We intentionally *misspecify* the model and act as if the drift were equal to **zero**. Ultimate mathematical justification: Girsanov's theorem (comes later). Misspecification is for facilitation of inference from the computational point of view: we avoid the intractable (true) likelihood.

Gaussian pseudo-likelihood

- With drift set to zero, the Gaussian (pseudo)-likelihood is given by

$$L_n(s) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi \int_{t_{i-1,n}}^{t_{i,n}} s^2(u)\,\mathrm{d}u}} \psi \left(\frac{Y_{i,n}}{\sqrt{\int_{t_{i-1,n}}^{t_{i,n}} s^2(u)\,\mathrm{d}u}} \right) \right\},$$

  where $\psi(u) = \exp(-u^2/2)$.
- As the likelihood depends on s through integrals $\int_{t_{i-1,n}}^{t_{i,n}} s^2(u)\,\mathrm{d}u$, it is computationally convenient to use a prior with realisations that are constant on intervals $[t_{i-1}, t_i)$.

Gaussian pseudo-likelihood

- With drift set to zero, the Gaussian (pseudo)-likelihood is given by

$$L_n(s) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi \int_{t_{i-1,n}}^{t_{i,n}} s^2(u)\, \mathrm{d}u}} \psi \left(\frac{Y_{i,n}}{\sqrt{\int_{t_{i-1,n}}^{t_{i,n}} s^2(u)\, \mathrm{d}u}} \right) \right\},$$

  where $\psi(u) = \exp(-u^2/2)$.
- As the likelihood depends on s through integrals $\int_{t_{i-1,n}}^{t_{i,n}} s^2(u)\, \mathrm{d}u$, it is computationally convenient to use a prior with realisations that are constant on intervals $[t_{i-1}, t_i)$.

Bins

- Let $m < n$. Then $n = mN + r$ with $0 \leq r < m$, and in fact $N = \lfloor \frac{n}{m} \rfloor$. Define bins

$$B_k = [t_{m(k-1),n}, t_{mk,n}), \quad k = 0, \ldots, N-1,$$
$$B_N = [t_{m(N-1),n}, 1].$$

- Consider piecewise constant functions

$$s = \sum_{k=1}^{N} \xi_k 1_{B_k}, \quad s^2 = \sum_{k=1}^{N} \xi_k^2 1_{B_k} = \sum_{k=1}^{N} \theta_k 1_{B_k}.$$

Bins

- Let $m < n$. Then $n = mN + r$ with $0 \leq r < m$, and in fact $N = \lfloor \frac{n}{m} \rfloor$. Define bins

$$B_k = [t_{m(k-1),n}, t_{mk,n}), \quad k = 0, \ldots, N-1,$$
$$B_N = [t_{m(N-1),n}, 1].$$

- Consider piecewise constant functions

$$s = \sum_{k=1}^{N} \xi_k 1_{B_k}, \quad s^2 = \sum_{k=1}^{N} \xi_k^2 1_{B_k} = \sum_{k=1}^{N} \theta_k 1_{B_k}.$$

Outline

Inverse gamma prior and posterior

Lemma 1

Assume $\theta_1, \ldots, \theta_N$ *are independent with the inverse gamma* $\mathrm{IG}(\alpha, \beta)$ *distribution. Then* $\theta_1, \ldots, \theta_N$ *are a posteriori independent and, for* $k = 1, \ldots, N-1$,

$$\theta_k \mid \mathcal{X}_n \sim \mathrm{IG}\left(\alpha + m/2, \beta + nZ_k/2\right).$$

with $Z_k = \sum_{i=(k-1)m+1}^{km} Y_{i,n}^2$, *whereas*

$$\theta_N \mid \mathcal{X}_n \sim \mathrm{IG}\left(\alpha + (m+r)/2, \beta + nZ_N/2\right),$$

with $Z_N = \sum_{i=(N-1)m+1}^{n} Y_{i,n}^2$.

Posterior mean

- The posterior mean of s^2 can be obtained from the posterior means of θ_k's. For instance, for $k < N$ the posterior mean of θ_k is equal to

$$\mathbb{E}_{\Pi_n}(\theta_k \mid \mathcal{X}_n) = \frac{\beta + nZ_k/2}{\alpha + m/2 - 1}. \tag{1}$$

- Conceptually the posterior mean of s^2 in this context is quite similar to a regressogram. Regressogram is a nonparametric regression function estimator similar to a histogram, in that it is also based on averaging over bins.

Posterior mean

- The posterior mean of s^2 can be obtained from the posterior means of θ_k's. For instance, for $k < N$ the posterior mean of θ_k is equal to

$$\mathbb{E}_{\Pi_n}(\theta_k \mid \mathcal{X}_n) = \frac{\beta + nZ_k/2}{\alpha + m/2 - 1}. \tag{1}$$

- Conceptually the posterior mean of s^2 in this context is quite similar to a regressogram. Regressogram is a nonparametric regression function estimator similar to a histogram, in that it is also based on averaging over bins.

Assumptions

Assumption 1

(a) For some $K > 0$ it holds that

$$|b_0(t,x)|^2 \leq K(1+|x|^2), \quad \forall x \in \mathbb{R},$$
$$|b_0(t,x) - b_0(t,y)| \leq K|x-y|, \quad \forall t \in [0,1], \quad \forall x, y \in \mathbb{R};$$

(b) the dispersion coefficient s_0 is Hölder continuous on $[0,1]$ with Hölder constant L and Hölder exponent $\lambda \in (0,1]$, $|s_0(t) - s_0(s)| \leq L|t-s|^\lambda$ for all $t, s \in [0,1]$, and is bounded away from zero and (by continuity also from infinity).

Assumptions

Assumption 1

(a) For some $K > 0$ it holds that

$$|b_0(t,x)|^2 \leq K(1+|x|^2), \quad \forall x \in \mathbb{R},$$
$$|b_0(t,x) - b_0(t,y)| \leq K|x-y|, \quad \forall t \in [0,1], \quad \forall x,y \in \mathbb{R};$$

(b) the dispersion coefficient s_0 is Hölder continuous on $[0,1]$ with Hölder constant L and Hölder exponent $\lambda \in (0,1]$, $|s_0(t) - s_0(s)| \leq L|t-s|^\lambda$ for all $t,s \in [0,1]$, and is bounded away from zero and (by continuity also from infinity).

L_2 asymptotics

Theorem 1

Assume s_0 is Hölder smooth of order λ, and let $N \asymp n^{1/(2\lambda+1)}$. Then the posterior for s^2 contracts around the truth s_0^2 at rate $\varepsilon_n \asymp n^{-\lambda/(2\lambda+1)}$: for any sequence h_n tending to infinity as $n \to \infty$, we have

$$\mathbb{E}\left[\Pi_n(\|s^2 - s_0^2\|_2 \geq h_n \varepsilon_n \mid \mathcal{X}_n)\right] \to 0$$

as $n \to \infty$.

L_∞ asymptotics

Theorem 2

Assume $N_n \asymp n^{1/(2\lambda+1)}$. If we let $\tilde{\varepsilon}_n \asymp n^{-\lambda/(2\lambda+2)}$, then for any sequence h_n tending to infinity (as $n \to \infty$) we have

$$\mathbb{E}\left[\Pi_n\Big(\sup_{x\in[0,1]} |s^2(x) - s_0^2(x)| \geq h_n\tilde{\varepsilon}_n \mid \mathcal{X}_n\Big)\right] \to 0$$

as $n \to \infty$.

Setup

- We present results for three dispersion coefficients s_1, s_2, s_3, up to a vertical shift given by

$$
\begin{aligned}
s_1(t) &= 3/2 + \sin(2(4t-2)) + 2\exp(-16(4t-2)^2),\\
s_2(t) &= 1/2 + 0.3\exp(-4(4t-1)^2) + 0.7\exp(-16(4t-3)^2),\\
s_3(t) &= W(\omega_0).
\end{aligned}
$$

- s_1, s_2 are standard test cases in nonparametric regression. They are infinitely smooth, but shape (curvature) is as important a feature as smoothness in nonparametric practice.
- s_3, being a realisation of a Brownian path, is an example of a very rough function (with Hölder index essentially a 1/2).

Setup

- We present results for three dispersion coefficients s_1, s_2, s_3, up to a vertical shift given by

$$\begin{aligned} s_1(t) &= 3/2 + \sin(2(4t-2)) + 2\exp(-16(4t-2)^2), \\ s_2(t) &= 1/2 + 0.3\exp(-4(4t-1)^2) + 0.7\exp(-16(4t-3)^2), \\ s_3(t) &= W(\omega_0). \end{aligned}$$

- s_1, s_2 are standard test cases in nonparametric regression. They are infinitely smooth, but shape (curvature) is as important a feature as smoothness in nonparametric practice.
- s_3, being a realisation of a Brownian path, is an example of a very rough function (with Hölder index essentially a 1/2).

Setup

- We present results for three dispersion coefficients s_1, s_2, s_3, up to a vertical shift given by

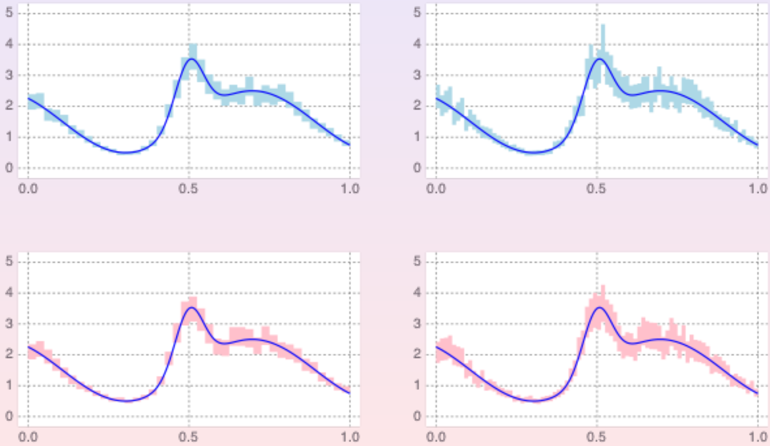
$$
\begin{aligned}
s_1(t) &= 3/2 + \sin(2(4t-2)) + 2\exp(-16(4t-2)^2),\\
s_2(t) &= 1/2 + 0.3\exp(-4(4t-1)^2) + 0.7\exp(-16(4t-3)^2),\\
s_3(t) &= W(\omega_0).
\end{aligned}
$$

- s_1, s_2 are standard test cases in nonparametric regression. They are infinitely smooth, but shape (curvature) is as important a feature as smoothness in nonparametric practice.
- s_3, being a realisation of a Brownian path, is an example of a very rough function (with Hölder index essentially a $1/2$).

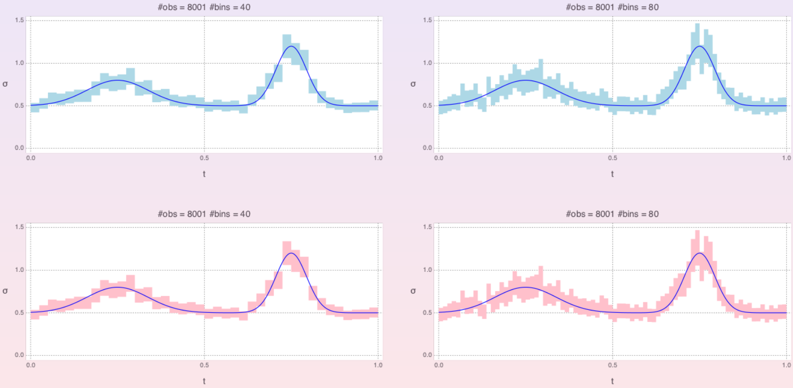
Further specifics

- Drift in the examples is either $b_0 \equiv 0$ or $b_1 \equiv 50$ or $b_1(x) = -10x + 20$ (top pictures vs. bottom pictures).
- Prior on θ_k's is non-informative independent $\mathrm{IG}(0.1, 0.1)$.
- The true curves in the pictures are given by solid lines, bands correspond to marginal 98% symmetric credible sets.

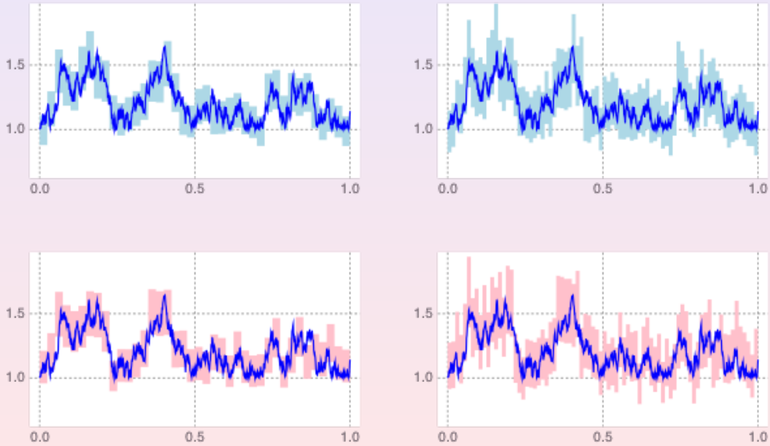
Example 1



Example 2


#obs = 8001 #bins = 40
#obs = 8001 #bins = 80
#obs = 8001 #bins = 40
#obs = 8001 #bins = 80
σ
t
0.0
0.5
1.0
1.5

Example 3



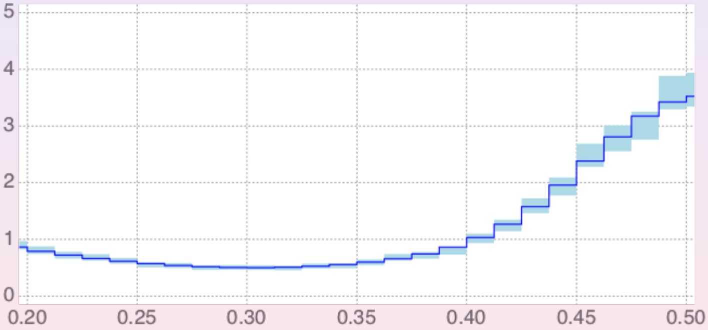
Some observations

- A strong nonzero drift hardly affects the obtained credible bands, and credible bands successfully recover the overall shape of the functions s_i.
- The functions s_i do not always pass through all the credible intervals, which is not surprising given the fact that these intervals are marginal.

Some observations

- A strong nonzero drift hardly affects the obtained credible bands, and credible bands successfully recover the overall shape of the functions s_i.
- The functions s_i do not always pass through all the credible intervals, which is not surprising given the fact that these intervals are marginal.

Zoomed-in picture of Example 1 (no drift)



Disjoint credible sets when the curvature is strong ($t = .45, .48$).

Number of bins

- Our theorems only specify that the optimal number of bins N_n is proportional to n^α, where the exponent α depends on the smoothness λ of a function to be estimated. This does not give a directly applicable recipe on how to choose the proportionality constant.
- In practice we recommend to use our theoretical results as guidance and to try out several choices of the number of bins. This is not unlike the scale-space smoothing approach in the frequentist literature, or bandwidth selection in kernel estimation.

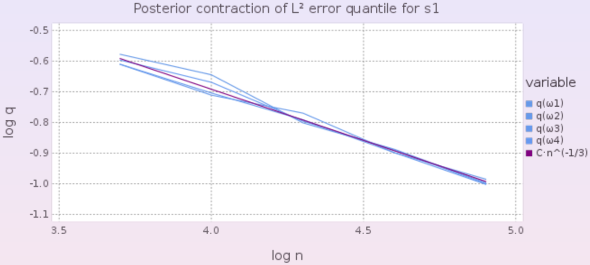
Number of bins

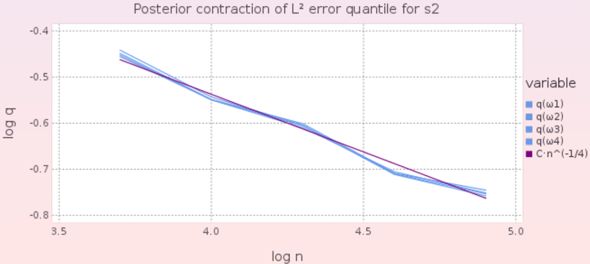
- Our theorems only specify that the optimal number of bins N_n is proportional to n^α, where the exponent α depends on the smoothness λ of a function to be estimated. This does not give a directly applicable recipe on how to choose the proportionality constant.
- In practice we recommend to use our theoretical results as guidance and to try out several choices of the number of bins. This is not unlike the scale-space smoothing approach in the frequentist literature, or bandwidth selection in kernel estimation.

Empirical contraction rates

- Of particular interest is the empirical size of the L_2- and L_∞-balls containing most of the posterior mass.
- To assess this, we approximate the distribution of the L_2- or L_∞-distance between posterior samples and the truth by sampling from the posterior. We do this for four different realisations of the model denoted by $X(\omega_1), \ldots, X(\omega_4)$.

L_2-rates

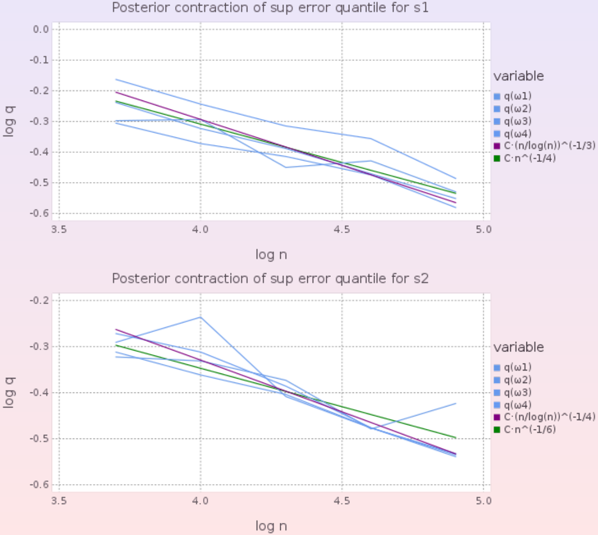

Posterior contraction of L² error quantile for s1
-0.5
-0.6
-0.7
-0.8
-0.9
-1.0
-1.1
log q
3.5
4.0
4.5
5.0
log n
variable
q(ω1)
q(ω2)
q(ω3)
q(ω4)
C·n^(-1/3)


Posterior contraction of L² error quantile for s2
-0.4
-0.5
-0.6
-0.7
-0.8
log q
3.5
4.0
4.5
5.0
log n
variable
q(ω1)
q(ω2)
q(ω3)
q(ω4)
C·n^(-1/4)

L_2-rate: discussion

- The empirical findings for function s_1 agree very well with the exponent $\frac{1}{3} = \frac{\lambda}{2\lambda+1}$ for $\lambda = 1$ obtained.
- The function s_3 is λ-Hölder smooth for any $\lambda < \frac{1}{2}$. The empirically determined exponent is in excellent agreement with the exponent $\frac{1}{4} = \frac{\lambda}{2\lambda+1}$ for $\lambda = \frac{1}{2}$.

L_∞-rates



L_∞-rate: discussion

- If the number of bins N_n is chosen in analogy to nonparametric kernel regression as $\log(N_n) = \log 5 + \frac{1}{2\lambda_i+1}\log(n/\log(n))$, the empirical findings suggest the rate $(n/\log n)^{\frac{1}{3}}$ or similar for s_1, and the rate $(n/\log n)^{\frac{1}{4}}$ or similar for s_2.
- The theoretical L_∞ posterior contraction rate we derived is possibly suboptimal. But empirical results are less conclusive than in the case of the L_2-norm.

Description

- The daily exchange rates (noon buying rates in New York City for cable transfers payable in foreign currencies) JPY/USD and USD/GBP from January 1, 1999, to March 20, 2010 are available as data sets DEXJPUS and DEXUSUK from the Federal Reserve.
- These exchange rate time series were considered in Hamrick et al. (2011). Discrete-time observations obtained from the SDE

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad X_0 = x, \quad t \in [0, T], \tag{2}$$

with space-dependent dispersion coefficient σ. Hamrick et al. proposed a maximum penalised quasi-likelihood method to estimate nonparametrically the diffusion coefficient σ^2.

Description

- The daily exchange rates (noon buying rates in New York City for cable transfers payable in foreign currencies) JPY/USD and USD/GBP from January 1, 1999, to March 20, 2010 are available as data sets DEXJPUS and DEXUSUK from the Federal Reserve.
- These exchange rate time series were considered in Hamrick et al. (2011). Discrete-time observations obtained from the SDE

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad X_0 = x, \quad t \in [0, T], \tag{2}$$

  with space-dependent dispersion coefficient σ. Hamrick et al. proposed a maximum penalised quasi-likelihood method to estimate nonparametrically the diffusion coefficient σ^2.

DEXJPUS

Note the higher volatility after 2008.

DEXUSUK

Note the higher volatility after 2008.

Non-stationarity

- Plots of both series appear to indicate that the data are nonstationary, and this is confirmed also by the outcomes of the augmented Dickey-Fuller test we performed using `urca` package in **R**. The test constitutes a standard unit root test in time series analysis.
- As stationarity is no prerequisite for application of our nonparametric Bayesian method, we did not pursue this question any further.

Comparison

- We used the estimates $\hat{\sigma}$ given in Hamrick et al. (2011) to calculate the induced estimates $t \mapsto \hat{s}(t) = \hat{\sigma}(X_t)$ of the historical volatility.
- We employed a non-informative $\mathrm{IG}(0.001, 0.001)$ prior for coefficients θ_k's.
- Our estimation results show that the volatility was high in final years of the decade 2000–2010, coinciding with the sub-prime mortgage crisis and the following recession. The model from Hamrick et al. (2011) does *not* appear to capture this fact.
- It is reassuring to see that our method recovers this relevant event from the data.
  Accordingly, we believe that including time-dependence of the volatility into the model is more appropriate in this example.

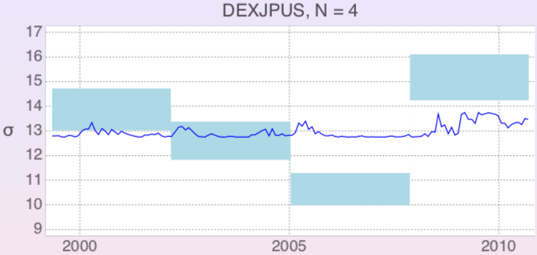
Comparison

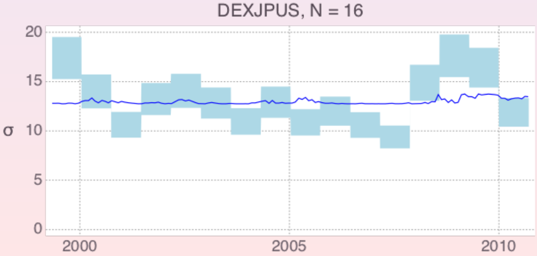
- We used the estimates $\hat{\sigma}$ given in Hamrick et al. (2011) to calculate the induced estimates $t \mapsto \hat{s}(t) = \hat{\sigma}(X_t)$ of the historical volatility.
- We employed a non-informative $\mathrm{IG}(0.001, 0.001)$ prior for coefficients θ_k's.
- Our estimation results show that the volatility was high in final years of the decade 2000–2010, coinciding with the sub-prime mortgage crisis and the following recession. The model from Hamrick et al. (2011) does *not* appear to capture this fact.
- It is reassuring to see that our method recovers this relevant event from the data.
  Accordingly, we believe that including time-dependence of the volatility into the model is more appropriate in this example.

Comparison

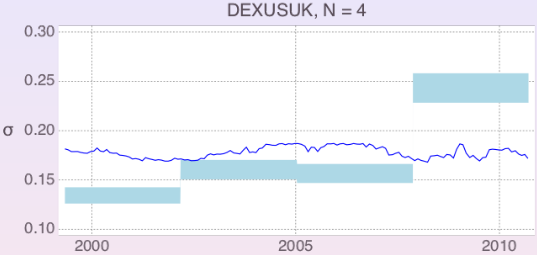
- We used the estimates $\hat{\sigma}$ given in Hamrick et al. (2011) to calculate the induced estimates $t \mapsto \hat{s}(t) = \hat{\sigma}(X_t)$ of the historical volatility.
- We employed a non-informative $\mathrm{IG}(0.001, 0.001)$ prior for coefficients θ_k's.
- Our estimation results show that the volatility was high in final years of the decade 2000–2010, coinciding with the sub-prime mortgage crisis and the following recession. The model from Hamrick et al. (2011) does *not* appear to capture this fact.
- It is reassuring to see that our method recovers this relevant event from the data.
  Accordingly, we believe that including time-dependence of the volatility into the model is more appropriate in this example.

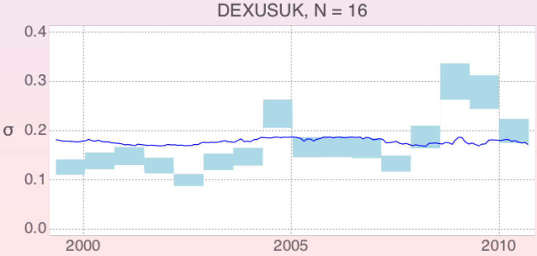
DEXJPUS


DEXJPUS, N = 4
17
16
15
14
13
12
11
10
9
σ
2000
2005
2010


DEXJPUS, N = 16
20
15
10
5
0
σ
2000
2005
2010

DEXUSUK


DEXUSUK, N = 4
σ
0.30
0.25
0.20
0.15
0.10
2000
2005
2010


DEXUSUK, N = 16
σ
0.4
0.3
0.2
0.1
0.0
2000
2005
2010

Outline

General priors on bins

The prior Π_n is defined as the law of a random function $s = \sum_{k=1}^{N_n} \xi_k \mathbf{1}_{B_k}$, where the random variables $\kappa \leq \xi_k \leq \mathcal{K}$, $k = 1, \ldots, N_n$, are independent and identically distributed with a density that is bounded away from zero on the interval $[\kappa, \mathcal{K}]$.

Harmless misspecifation of the drift

Proposition 1

Let Assumption 1 hold and assume that for $\varepsilon_n \to 0$

$$\mathbb{E}_{0,s_0}[\Pi_n(||s_0 - s||_2 \geq \varepsilon_n|\mathcal{X}_n)] \to 0$$

as $n \to \infty$. Then also

$$\mathbb{E}_{b,s_0}[\Pi_n(||s_0 - s||_2 \geq \varepsilon_n|\mathcal{X}_n)] \to 0.$$

Here $\mathbb{E}_{b,s_0}$ denotes expectation under drift b and $\mathbb{E}_{0,s_0}$ denotes expectation under zero drift.

Main result

Theorem 3

Under the model assumptions and $0 < \kappa \leq s_0(t) \leq \mathcal{K} < \infty$ for all $t \in [0,1]$. For any sequence $m_n \asymp n^{1-\alpha}$ (bin sizes), equivalently $N_n \asymp n^{\alpha}$, with $\alpha = \frac{1}{1+2\lambda}$, there exists a constant $\widetilde{M} > 0$, such that for $\varepsilon_n = \widetilde{M} n^{-\beta} \log^{\gamma} n$ with $\beta = \lambda/(2\lambda + 2)$ and arbitrary positive γ,

$$\mathbb{E}_{b_0,s_0}[\Pi_n(||s_0 - s||_2 \geq \varepsilon_n | \mathcal{X}_n)] \to 0 \quad (3)$$

as $n \to \infty$.

Comments on the rate

The rate $\varepsilon_n \asymp n^{-\beta} \log^{\gamma} n$ is up to the logarithmic factor the same as in Theorem 2 for the $||\cdot||_\infty$-norm, but slightly worse than in Theorem 1 for the $||\cdot||_2$-norm.

Technical origin: a probably suboptimal bound for the $||\cdot||_\infty$-norm of sets in the proof of the next lemma. Notation:

$$\mathcal{W}_i = 1 - \frac{Y_{i,n}^2}{\int_{t_{i-1,n}}^{t_{i,n}} s_0^2(u)\,\mathrm{d}u}, \quad f_s(z) = \frac{\int_z^{z+1/n} [s_0^2(u) - s^2(u)]\,\mathrm{d}u}{\int_z^{z+1/n} s^2(u)\,\mathrm{d}u}.$$

A typical lemma

In the proof one needs various results relying on empirical process theory. Here is an important example.

Lemma 2

Let the conditions of Theorem 3 hold and assume $b_0 = 0$. Then

$$\mathbb{P}_{0,s_0}\left(\sup_{f_s\in\mathcal{F}_{s_0,\widetilde{\varepsilon}_n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathcal{W}_i f_s(\frac{i-1}{n})\right|\geq\delta_n\right)\lesssim\frac{1}{n^{\lambda}\widetilde{\varepsilon}_n^2},$$

where $\mathcal{F}_{s_0,\widetilde{\varepsilon}_n} = \{f_s : \|s - s_0\|_\infty < \widetilde{\varepsilon}_n\}$ and δ_n is an arbitrary sequence of positive numbers, such that $\delta_n \asymp \widetilde{\varepsilon}_n^2$.

A sharper result (possible?) would lead to a better contraction rate.

More on the proof

One works with

$$\Pi_n(U^c_{s_0,\varepsilon_n} \mid \mathcal{X}_n) = \frac{\int_{U^c_{s_0,\varepsilon_n}} R_n(s)\Pi(\mathrm{d}s)}{\int_{\mathcal{S}} R_n(s)\Pi_n(\mathrm{d}s)} \leq \frac{\int_{U^c_{s_0,\varepsilon_n}} R_n(s)\Pi(\mathrm{d}s)}{\int_{V_{s_0,\tilde{\varepsilon}_n}} R_n(s)\Pi_n(\mathrm{d}s)} = \frac{\mathcal{N}_n}{\mathcal{D}_n},$$

and aims at exponential bounds for numerator and denominator. One has

$$\mathcal{D}_n \geq \inf_{V_{s_0,\tilde{\varepsilon}_n}} R_n(s) \times \Pi_n(V_{s_0,\tilde{\varepsilon}_n}),$$

and for the analysis of the infimum, Lemma 2 is used.

Metric entropy considerations are used all over the place.

Outline

Summary and outlook

- We propose an easily implementable nonparametric Bayesian method for estimating deterministic diffusion coefficient.
- The method can be thought of as a Bayesian equivalent of a histogram or a regressogram, and is useful e.g. for exploratory data analysis.
- We support the method by theory and simulation examples, both reported only partially here, and applied the method on real data examples, obtaining interesting results.
- Future work: many extensions (of the histogram approach) possible.

Two events

- 12th Bachelier Colloquium on Mathematical Finance and Stochastic Calculus, Metabief, January 15–20, 2018
- 17th Winter School on Mathematical Finance, Lunteren, January 22–24, 2018

Two events

- 12th Bachelier Colloquium on Mathematical Finance and Stochastic Calculus, Metabief, January 15–20, 2018
- 17th Winter School on Mathematical Finance, Lunteren, January 22–24, 2018

Thank you!