Adaptive posterior contraction results for empirical Bayes methods for diffusions

Jan van Waaij, Humboldt University of Berlin

Slides are on https://github.com/Jan-van-Waaij/WorkshopParis2018

Subject of today

- Posterior contraction rates for diffusion processes.
- Adaptation to unknown smoothness of the function.
- Empirical Bayes.

Statistical inference for diffusions

- Diffusions on the line: real-valued strong Markov processes with continuous paths,
- Under weak conditions a diffusion is described via an SDE

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

- We assume $\sigma \equiv 1$,
- $\theta : \mathbb{R} \to \mathbb{R}$ is measurable, 1-periodic and $\int_0^1 \theta(x)^2 dx < \infty$.
- Observations $X^T = (X_t : t \in [0, T])$ of

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

- Goal: estimate θ.

Posterior contraction for the Gaussian process prior

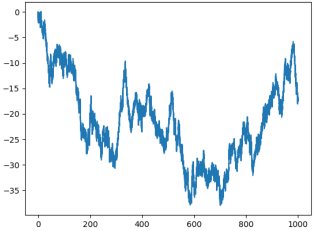
- Continuous stochastic process whose fdd's are multivariate Gaussian.
- Prior:
$$\theta = \sum_{k=1}^{\infty} k^{-\alpha-1/2} Z_k \phi_k.$$
- ☺ Posterior = Gaussian process + explicit formula
- True function $\theta_0 = \sum_{k=1}^{\infty} \theta_k \phi_k$ satisfies $\sum_{k=1}^{\infty} k^{2\beta}\theta_k^2 < \infty$.
- 😐 Minimax posterior convergence rate $T^{-\frac{\beta}{1+2\beta}}$ if and only if $\alpha = \beta$.
- ☹ Not adaptive!

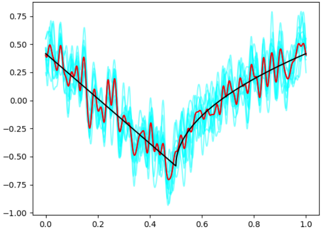
Example

- SDE $dX_t = \theta_0(X_t)dt + dW_t$,
- Prior:

$$\theta = \sum_{k=1}^{100} k^{-1} Z_k \phi_k$$

-





Solution I: Hierarichal Bayes

- Let the posterior choose the right smoothness.
- Hyperpriors on the hyperparameters.

$$1\sum_{k=1}^{\infty} k^{-\alpha-1/2} Z_k \phi_k$$

$E \sim \mathrm{Exp}(1)$		$J \sim$ geometric
$S = \frac{E^{1/2+\alpha}}{\sqrt{T}}$	$\pi(\alpha) \asymp e^{-T^{\frac{1}{1+2\alpha}}}$	$S^2 \sim$ inv. gamma
$\theta \mid S =$	$\theta \mid \alpha =$	$\theta \mid J, S =$
$S\sum_{k=1}^{\infty} k^{-\alpha-1/2} Z_k \phi_k$	$\sum_{k=1}^{\infty} k^{-\alpha-1/2} Z_k \phi_k$	$S\sum_{k=1}^{J} k^{-\alpha-1/2} Z_k \phi_k$
Adap. $\beta \leq \alpha + 1/2$	Adap. $\beta > 0$	Adap. $\beta > 0$.

Empirical Bayes

- $$s \sum_{k=1}^{\infty} k^{-\alpha-1/2} Z_k \phi_k$$
- Estimate the hyperparameters from the data.
- Use the plug-in posterior for the inference.
-
  - Empirical Bayes on s,
  - empirical Bayes on α,
  - empirical Bayes on (α, s).

Hierarchical vs. empirical Bayes

- Gaussian process prior is not adaptive.
- Hierarchical Bayes solution: equip hyperparameters with additional prior.
  - Prior is not Gaussian
  - Adaptivity + (near) optimal rates.
- Empirical Bayes solution: estimate hyperparameter from the data and use plug-in posterior for inference.
  - the (data-driven) prior still Gaussian.
  - But a lot unknown about the theoretical and computational performance...
  - The analysis is considerably harder.

Empirical Bayes

- Use prior Π_α
$$\theta = \sum_{k=1}^{\lfloor T \rfloor} k^{-1/2-\alpha} Z_k \phi_k,$$
- estimate α from the data,
- use
$$\Pi_{\hat{\alpha}} = \Pi_\alpha(\cdot \mid X^T)\Big|_{\alpha=\hat{\alpha}}$$
for the inference.

- Prior Π_α defined by $\sum_{k=1}^{\lfloor T \rfloor} k^{-\alpha-1/2} Z_k \phi_k$.
- "Prior behaves best when it puts a lot of prior mass around θ_0."
- That is when $\alpha = \beta + \mathcal{O}(1/\log T)$.
- Marginal maximum likelihood estimator (MMLE)

$$\hat{\alpha} = \mathrm{argmax}_{\alpha \in \Lambda} \int p_\theta(X^T) d\Pi_\alpha(\theta),$$

$$p_\theta(X^T) = \exp\left\{ \int_0^T \theta(X_t) dX_t - \frac{1}{2} \int_0^T \theta(X_t)^2 dt \right\}$$

- $$\Lambda = \left[\frac{1}{2} + \delta, \sqrt{\log T} \right]$$

  or

  $$\left\{ k/\log T : \frac{1}{2} + \delta \leq k/\log T \leq \sqrt{\log T} \right\}.$$

Theorem

When θ_0 is β-Sobolev smooth, $1/2 + \delta \leq \beta < \infty$, then for some $M > 0$,

$$\Pi_{\hat{\alpha}} \left(\theta : \|\theta - \theta_0\|_2 \leq M T^{-\frac{\beta}{1+2\beta}} \;\middle|\; X^T \right) \to 1$$

in $\mathbb{P}_{\theta_0}$-probability as $T \to \infty$.

Other options

1. Fix $\alpha > 1/2$ and priors
$$s \sum_{k=1}^{\infty} k^{-\alpha-1/2} Z_k \phi_k,$$
optimise over $s \in \Lambda = \left[T^{-\frac{1}{4+4\alpha}}, T^{\alpha}\right]$ (or discretised version).
2. Priors
$$s \sum_{k=1}^{\lfloor T \rfloor} k^{-\alpha-1/2} Z_k \phi_k,$$
$(\alpha, s) \in \left\{1/2 + \delta \leq \alpha \leq \sqrt{\log T}, T^{-\frac{1}{4+4\alpha}} \leq s \leq T^{\alpha}\right\}$ (or discretised version).

Outline of the proof

Show $\Pi_{\hat{\alpha}}\left(\|\theta-\theta_0\| > MT^{-\frac{\beta}{1+2\beta}} \mid X^T\right) \to 0$.

Ingredients of the proof:

1. Determine $\Lambda_0 \subseteq \Lambda$ where $\Pi_s(\cdot \mid X^T)$ enjoys good rates.
2. $\mathbb{P}_{\theta_0}(\hat{\alpha} \in \Lambda_0) \to 1$, as $T \to \infty$,
3. $$\mathbb{E}_{\theta_0}\left(\Pi_{\hat{\alpha}}\left(\|\theta-\theta_0\|_2 \geq MT^{-\frac{\beta}{1+2\beta}} \mid X^T\right)\right)$$
$$\leq \mathbb{E}_{\theta_0}\left(\sup_{\alpha\in\Lambda_0}\Pi_{\alpha}\left(\|\theta-\theta_0\|_2 \geq MT^{-\frac{\beta}{1+2\beta}} \mid X^T\right)\right)$$
$$+\,\mathbb{P}_{\theta_0}(\hat{\alpha} \notin \Lambda_0) \to 0.$$

Determining Λ_0

Let $K > 0$ be constant. There is a unique $\varepsilon_\alpha > 0$ so that

$$\Pi_\alpha(\|\theta - \theta_0\|_2 < K\varepsilon_\alpha) = e^{-T\varepsilon_\alpha^2}.$$

Let

$$\varepsilon_0 = \min_{\alpha \in \Lambda} \varepsilon_\alpha.$$

Let $L > 2$ be a constant and

$$\Lambda_0 = \{\alpha \in \Lambda : \varepsilon_\alpha \leq L\varepsilon_0\}.$$

Lemma

For $L > 1$ big enough, with $\mathbb{P}_{\theta_0}$-probability converging to one $\hat{\alpha} \in \Lambda_0$.

Step 1 Take $\bar{p} = p_\theta / p_{\theta_0}$ instead.

$$\begin{aligned} &\operatorname{argmax}_{\alpha \in \Lambda} \int p_\theta(X^T) d\Pi_\alpha(\theta) \\ =&\operatorname{argmax}_{\alpha \in \Lambda} \int \bar{p}_\theta(X^T) d\Pi_\alpha(\theta). \end{aligned}$$

Step 2 Let $\alpha_0 \in \Lambda, \varepsilon_{\alpha_0} \leq 2\varepsilon_0$. There are constants $0 < A < B$ so that with $\mathbb{P}_{\theta_0}$-probability converging to one,

$$\begin{aligned} &\int \bar{p}_\theta(X^T) d\Pi_{\alpha_0}(\theta) \geq e^{-AT\varepsilon_0^2} \\ > &e^{-BT\varepsilon_0^2} \geq \sup_{\alpha \in \Lambda \setminus \Lambda_0} \int \bar{p}_\theta(X^T) d\Pi_\alpha(\theta) \end{aligned}$$

Goal: show that

$$\mathbb{P}_{\theta_0}\left(\sup_{\alpha\in\Lambda\backslash\Lambda_0}\int \bar{p}_\theta(X^T)d\Pi_\alpha(\theta)\geq e^{-BT\varepsilon_0^2}\right)\to 0.$$

Choose $\alpha_1,\dots,\alpha_N\in\Lambda\backslash\Lambda_0$, so that

$$\Lambda\backslash\Lambda_0\subseteq=\bigcup_{k=1}^N I_k:=\bigcup_{k=1}^N\left[\alpha_k-T^{-2}/\log T,\alpha_k+T^{-2}/\log T\right].$$

$$\mathbb{P}_{\theta_0}\left(\sup_{\alpha\in\Lambda\backslash\Lambda_0}\int \bar{p}_\theta(X^T)d\Pi_\alpha(\theta)\geq e^{-BT\varepsilon_0^2}\right)$$
$$\leq T^3\max_{1\leq k\leq N}\mathbb{P}_{\theta_0}\left(\sup_{\alpha\in I_k}\int \bar{p}_\theta(X^T)d\Pi_\alpha(\theta)\geq e^{-BT\varepsilon_0^2}\right).$$

$$\text{Introduce } \Psi_\alpha(\theta) = \sum_{k=1}^{\lfloor T \rfloor} k^{\alpha_k - \alpha},$$

$$\text{when } \theta = \sum_{k=1}^{\lfloor T \rfloor} k^{-\alpha_k - 1/2} Z_k \phi_k,$$

$$\text{then } \Psi_\alpha(\theta) = \sum_{k=1}^{\lfloor T \rfloor} k^{-\alpha - 1/2} Z_k \phi_k,$$

So when $\theta \sim \Pi_{\alpha_k}$, then $\Psi_\alpha(\theta) \sim \Pi_\alpha$.

$$\sup_{\alpha \in I_k} \int \bar{p}_\theta(X^T) d\Pi_\alpha(\theta) = \sup_{\alpha \in I_k} \int \bar{p}_{\Psi_\alpha(\theta)}(X^T) d\Pi_{\alpha_k}(\theta)$$

$$\leq \int \sup_{\alpha \in I_k} \bar{p}_{\Psi_\alpha(\theta)}(X^T) d\Pi_{\alpha_k}(\theta)$$

We bound

$$
\begin{aligned}
&\mathbb{P}_{\theta_0}\left(\sup_{\alpha\in I_k}\int \bar{p}_\theta(X^T)d\Pi_\alpha(\theta)\geq e^{-BT\varepsilon_0^2}\right)\\
\leq&\mathbb{P}_{\theta_0}\left(\int \sup_{\alpha\in I_k}\bar{p}_{\Psi_\alpha(\theta)}(X^T)d\Pi_{\alpha_k}(\theta)\geq e^{-BT\varepsilon_0^2}\right)\\
=&\,\mathbb{E}_{\theta_0}\left[\mathbb{I}\left\{\int \sup_{\alpha\in I_k}\bar{p}_{\Psi_\alpha(\theta)}(X^T)d\Pi_{\alpha_k}(\theta)\geq e^{-BT\varepsilon_0^2}\right\}(\varphi+1-\varphi)\right]\\
\leq&\,\mathbb{E}_{\theta_0}\varphi+e^{BT\varepsilon_0^2}\sqrt{\int \mathbb{E}_\theta\sup_{\alpha\in I_k}\left(\bar{p}_{\Psi_\alpha(\theta)}(X^T)/\bar{p}_\theta(X^T)\right)^2}\\
&\times\sqrt{\int \mathbb{E}_\theta[1-\varphi]d\Pi_s(\theta)}.
\end{aligned}
$$

Thank you!