Posterior contraction in nonparametric inverse problems

Workshop on Inverse Problems, Heidelberg, 28–29 October 2016,

Bartek Knapik

Joint work with

Botond Szabó (Leiden/Budapest), Aad van der Vaart (Leiden), Harry van Zanten (Amsterdam, KdVI) and Jean-Bernard Salomond (Paris)

28 October 2016

Recovery of f from indirect observations

Observations

$$Y = Kf + noise,$$

where K is some known transformation of infinite-dimensional object of interest f.

Recovery of f from indirect observations

Observations

$$Y = Kf + noise,$$

where K is some known transformation of infinite-dimensional object of interest f.

Bayesian approach

We put a prior Π on f and we are interested in posterior contraction

$$\Pi(d(f, f_0) \gtrsim \varepsilon_n | Y)$$

Recovery of f from indirect observations

Observations

$$Y = Kf + noise,$$

where K is some known transformation of infinite-dimensional object of interest f.

Bayesian approach

We put a prior Π on f and we are interested in posterior contraction

$$\Pi(d(f, f_0) \gtrsim \varepsilon_n | Y)$$

There is a catch...

- "Classical" theory of posterior contraction does not give interesting results ("inverseness")

Frequentist vs. Bayesian

Frequentist

We need an estimation procedure, for instance, (penalized) maximum likelihood estimation, method of moments, etc.

Frequentist vs. Bayesian

Frequentist

We need an estimation procedure, for instance, (penalized) maximum likelihood estimation, method of moments, etc.

Bayesian

- Put a prior on f

Frequentist vs. Bayesian

Frequentist

We need an estimation procedure, for instance, (penalized) maximum likelihood estimation, method of moments, etc.

Bayesian

- Put a prior on f
- Data Y depends on f

Frequentist vs. Bayesian

Frequentist

We need an estimation procedure, for instance, (penalized) maximum likelihood estimation, method of moments, etc.

Bayesian

- Put a prior on f
- Data Y depends on f
- Bayes' theorem says how f depends on the data Y – the posterior distribution

Frequentist vs. Bayesian

Frequentist

We need an estimation procedure, for instance, (penalized) maximum likelihood estimation, method of moments, etc.

Bayesian

- Put a prior on f
- Data Y depends on f
- Bayes' theorem says how f depends on the data Y – the posterior distribution

Choosing a prior can be viewed as choosing or tuning the estimation procedure

Pragmatic Bayesian

How does the posterior behave from frequentist perspective?

Pragmatic Bayesian

How does the posterior behave from frequentist perspective?

Posterior contraction rate

The posterior $\Pi_n(\,\cdot\,|Y^{(n)})$ is said to contract at f_0 at rate $\varepsilon_n \downarrow 0$, if

$$\Pi_n(f : d(f, f_0) \geq M_n\varepsilon_n|Y^{(n)}) \to 0,$$

in $P_0^{(n)}$-probability, for every $M_n \to \infty$ as $n \to \infty$.

Pragmatic Bayesian

How does the posterior behave from frequentist perspective?

Posterior contraction rate

The posterior $\Pi_n(\,\cdot\,|Y^{(n)})$ is said to contract at f_0 at rate $\varepsilon_n \downarrow 0$, if

$$\Pi_n(f : d(f, f_0) \geq M_n \varepsilon_n | Y^{(n)}) \to 0,$$

in $P_0^{(n)}$-probability, for every $M_n \to \infty$ as $n \to \infty$.

There exist point estimators based on the posterior that also attain this rate (i.e., posterior mean or the centre of a ball capturing enough posterior mass).

General posterior contraction theorems

First: general conditions for consistency – Schwartz (1965)

General posterior contraction theorems

First: general conditions for consistency – Schwartz (1965)

Next: extension to rates – Ghoshal, Ghosh and van der Vaart (2000), Shen and Wasserman (2001)

General posterior contraction theorems

First: general conditions for consistency – Schwartz (1965)

Next: extension to rates – Ghoshal, Ghosh and van der Vaart (2000), Shen and Wasserman (2001)

The literature on posterior contraction is rich.
Usually closely related to the ideas in these seminal papers.

General posterior contraction theorems

First: general conditions for consistency – Schwartz (1965)

Next: extension to rates – Ghoshal, Ghosh and van der Vaart (2000), Shen and Wasserman (2001)

The literature on posterior contraction is rich.
Usually closely related to the ideas in these seminal papers.

General theory relies on specific, yet often natural, distances.

Wrong distance

Let $Y = (Y_1, Y_2, \dots)$

$$Y_i = \kappa_i f_i + \frac{1}{\sqrt{n}} Z_i, \qquad i = 1, 2, \dots.$$

Wrong distance

Let $Y = (Y_1, Y_2, \dots)$

$$Y_i = \kappa_i f_i + \frac{1}{\sqrt{n}} Z_i, \qquad i = 1, 2, \dots.$$

For any $f \in \ell_2$, and $\kappa \in \ell_2$ let $Kf = (\kappa_1 f_1, \kappa_2 f_2, \dots)$.

The natural distance between Kf_1 and Kf_2 (also from posterior contraction point of view)

$$\|Kf_1 - Kf_2\|^2 = \sum_{i=1}^{\infty} |\kappa_i f_{i,1} - \kappa_i f_{i,2}|^2$$

Wrong distance and posterior contraction

When $|\kappa_i| \downarrow 0$, even if $\|Kf - Kf_0\|$ is small, $\|f - f_0\|$ can be large.

Wrong distance and posterior contraction

When $|\kappa_i| \downarrow 0$, even if $\|Kf - Kf_0\|$ is small, $\|f - f_0\|$ can be large.

In general, there is no C such that

$$d(f, f_0) \leq Cd(Kf, Kf_0).$$

Therefore, we cannot say anything about

$$\Pi_n(d(f, f_0) \gtrsim \widetilde{\varepsilon}_n | Y^{(n)}) \to 0$$

based on

$$\Pi_n(d(Kf, Kf_0) \gtrsim \varepsilon_n | Y^{(n)}) \to 0.$$

Does Bayes deal with inverse problems?

Everything up until now does not show that Bayesian approach does not work.

Does Bayes deal with inverse problems?

Everything up until now does not show that Bayesian approach does not work.

We see that the general theoretical results are not applicable.

Does Bayes deal with inverse problems?

Everything up until now does not show that Bayesian approach does not work.

We see that the general theoretical results are not applicable.

Bayes should work and does work. Putting a prior is natural way of specifying a degree of regularization.

Does Bayes deal with inverse problems?

Everything up until now does not show that Bayesian approach does not work.

We see that the general theoretical results are not applicable.

Bayes should work and does work. Putting a prior is natural way of specifying a degree of regularization.

No general theory = direct evaluation of the posterior and its properties.

Sequence setting

We observe

$$Y_i = \kappa_i f_i + \frac{1}{\sqrt{n}} Z_i, \qquad i = 1, 2, \dots,$$

and put a product prior $\Pi = \bigotimes N(0, \lambda_i)$ on the sequence $f = (f_i)$.

We say that the true f_0 is Sobolev β-smooth if $\sum i^{2\beta} f_{0,i}^2 < \infty$.

Sequence setting

We observe

$$Y_i = \kappa_i f_i + \frac{1}{\sqrt{n}} Z_i, \qquad i = 1, 2, \dots,$$

and put a product prior $\Pi = \bigotimes N(0, \lambda_i)$ on the sequence $f = (f_i)$.

We say that the true f_0 is Sobolev β-smooth if $\sum i^{2\beta} f_{0,i}^2 < \infty$.

Posterior rate of contraction

- $\lambda_i = i^{-1-2\alpha}$, $\kappa_i \asymp i^{-p}$, $\alpha > 0$, $p \geq 0$: $n^{-\frac{\alpha \wedge \beta}{1+2\alpha+2p}}$
- $\lambda_i = i^{-1-2\alpha}$, $\kappa_i \asymp \exp(-\gamma i^p)$, $\alpha, \gamma > 0$, $p \geq 1$: $(\log n)^{-\frac{\alpha \wedge \beta}{p}}$
- $\lambda_i = \exp(-\alpha i^p)$, $\kappa_i \asymp \exp(-\gamma i^p)$, $\alpha, \gamma > 0$, $p \geq 1$: $(\log n)^{-\frac{\beta}{p}}$

K, van der Vaart, van Zanten (2011, 2013), Florens and Simoni (2012, 2013), Agapiou et al. (2013, 2014), Ray (2013)

Getting the right α: Adaptation in mildly ill-posed problem

Recall

$$\Pi_\alpha = \bigotimes_i N(0, i^{-1-2\alpha})$$

where now we consider α a hyperparameter.

Getting the right α: Adaptation in mildly ill-posed problem

Recall

$$\Pi_\alpha = \bigotimes_i N(0, i^{-1-2\alpha})$$

where now we consider α a hyperparameter.

In the setting described before

$$f|\alpha \sim \Pi_\alpha = \bigotimes_i N(0, i^{-1-2\alpha}) \qquad \text{and} \qquad Y|(f,\alpha) \sim \bigotimes_i N(\kappa_i f_{0,i}, n^{-1}),$$

hence

$$Y|\alpha \sim \bigotimes_i N(0, i^{-1-2\alpha}\kappa_i^{-2} + n^{-1}).$$

Empirical and hierarchical Bayes

Since the prior Π_α depends on the hyperparameter, so does the posterior $\Pi_\alpha(\,\cdot\,|Y)$.

Empirical and hierarchical Bayes

Since the prior Π_α depends on the hyperparameter, so does the posterior $\Pi_\alpha(\,\cdot\,|Y)$.

Empirical Bayes posterior

Replace α by the maximizer $\hat{\alpha}_n$ of the likelihood

$$\Pi_{\hat{\alpha}_n}(\,\cdot\,|Y) = \Pi_\alpha(\,\cdot\,|Y)\Big|_{\alpha=\hat{\alpha}_n}.$$

Empirical and hierarchical Bayes

Since the prior Π_α depends on the hyperparameter, so does the posterior $\Pi_\alpha(\,\cdot\,|Y)$.

Empirical Bayes posterior

Replace α by the maximizer $\hat{\alpha}_n$ of the likelihood

$$\Pi_{\hat{\alpha}_n}(\,\cdot\,|Y) = \Pi_\alpha(\,\cdot\,|Y)\Big|_{\alpha=\hat{\alpha}_n}.$$

Hierarchical Bayes

We use a prior $\lambda(\alpha)$ on α and then the full hierarchical prior is given by

$$\Pi = \int_0^\infty \lambda(\alpha)\Pi_\alpha \, d\alpha$$

Integration example

The Volterra operator

We observe the process $Y = (Y_t : t \in [0, 1])$

$$Y_t = \int_0^t \int_0^s f(u)\, du\, ds + \frac{1}{\sqrt{n}} W_t,$$

where $W = (W_t : t \in [0, 1])$ is a Brownian motion

Integration example

The Volterra operator

We observe the process $Y = (Y_t : t \in [0,1])$

$$Y_t = \int_0^t \int_0^s f(u)\, du\, ds + \frac{1}{\sqrt{n}} W_t,$$

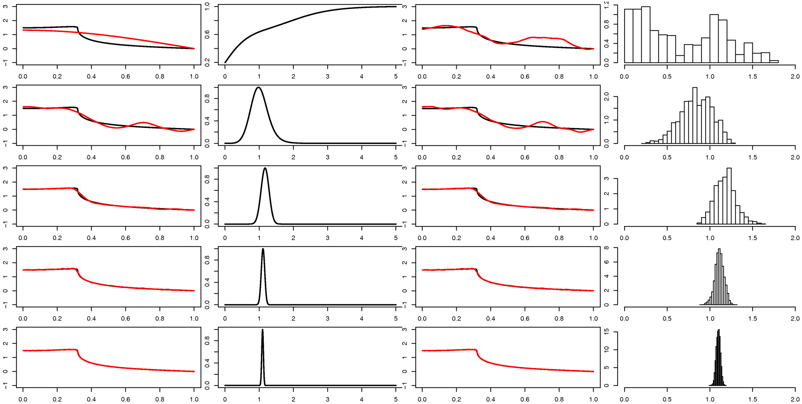
where $W = (W_t : t \in [0,1])$ is a Brownian motion

Spectral decomposition of the Volterra operator

The eigenvalues and eigenfunctions of $K^T K$ are

$$\kappa_i = \frac{1}{(i+1/2)\pi} \asymp i^{-1}, \quad e_i(x) = \sqrt{2}\cos\big((i+1/2)\pi x\big), \quad i = 0, 1, 2, \ldots.$$

Empirical and hierarchical Bayes



Black curve – true f_0, red curve – posterior mean

$n = 10^3, 10^5, 10^7, 10^9$ and 10^{11} (top to bottom)

Adaptive Bayes in mildly ill-posed problems

Let

$$\|f\|_\beta^2 = \sum_i i^{2\beta} f_i^2.$$

Theorem (K, Szabó, van der Vaart, van Zanten, 2012-2016)

For every $R > 0$ *and* $M_n \to \infty$

$$\sup_{\|f_0\|_\beta \le R} \mathrm{E}_0 \Pi_{\hat{\alpha}_n}\Big(\|f - f_0\| \ge M_n L_n\, n^{-\frac{\beta}{1+2\beta+2p}} \,|Y\Big) \to 0,$$

$$\sup_{\|f_0\|_\beta \le R} \mathrm{E}_0 \Pi\Big(\|f - f_0\| \ge M_n L_n\, n^{-\frac{\beta}{1+2\beta+2p}} \,|Y\Big) \to 0,$$

where $L_n = (\log n)^2 (\log\log n)^{1/2}$ *is a slowly varying term.*

Adaptive Bayes in mildly ill-posed problems

Let

$$\|f\|_\beta^2 = \sum_i i^{2\beta} f_i^2.$$

Theorem (K, Szabó, van der Vaart, van Zanten, 2012-2016)

For every $R > 0$ and $M_n \to \infty$

$$\sup_{\|f_0\|_\beta \le R} \mathrm{E}_0 \Pi_{\hat\alpha_n}\Big(\|f - f_0\| \geq M_n L_n\, n^{-\frac{\beta}{1+2\beta+2p}} \,|Y\Big) \to 0,$$

$$\sup_{\|f_0\|_\beta \le R} \mathrm{E}_0 \Pi\Big(\|f - f_0\| \geq M_n L_n\, n^{-\frac{\beta}{1+2\beta+2p}} \,|Y\Big) \to 0,$$

where $L_n = (\log n)^2 (\log\log n)^{1/2}$ is a slowly varying term.

Optimal!

Adaptive Bayes in mildly ill-posed problems

Let

$$\|f\|_\beta^2 = \sum_i i^{2\beta} f_i^2.$$

Theorem (K, Szabó, van der Vaart, van Zanten, 2012-2016)

For every $R > 0$ and $M_n \to \infty$

$$\sup_{\|f_0\|_\beta \le R} \mathrm{E}_0 \Pi_{\hat{\alpha}_n}\left(\|f - f_0\| \ge M_n L_n\, n^{-\frac{\beta}{1+2\beta+2p}} \,|Y\right) \to 0,$$

$$\sup_{\|f_0\|_\beta \le R} \mathrm{E}_0 \Pi\left(\|f - f_0\| \ge M_n L_n\, n^{-\frac{\beta}{1+2\beta+2p}} \,|Y\right) \to 0,$$

where $L_n = (\log n)^2 (\log\log n)^{1/2}$ is a slowly varying term.

Optimal!

up to a slowly varying term

Towards more general results

Let Kf denote the transformed parameter of interest f, and data by $Y \sim P_{Kf}$

Here $K : \mathscr{F} \ni f \mapsto Kf \in K\mathscr{F}$, d and d_K denote some metrics on $\mathscr{F}$ and $K\mathscr{F}$

Towards more general results

Let Kf denote the transformed parameter of interest f, and data by $Y \sim P_{Kf}$

Here $K : \mathscr{F} \ni f \mapsto Kf \in K\mathscr{F}$, d and d_K denote some metrics on $\mathscr{F}$ and $K\mathscr{F}$

Problems with general theory

- rates for $d_K\big(Kf_1, Kf_2\big)$ rather than $d\big(f_1, f_2\big)$
- when the problem is ill-posed, then $d_K\big(Kf_1, Kf_2\big)$ and $d\big(f_1, f_2\big)$ are not equivalent

$$\sup\Big\{d\big(f_1, f_2\big) \,:\, f_1, f_2 \in \mathscr{F} \;\text{ s.t. }\; d_K\big(Kf_1, Kf_2\big) \leq \delta\Big\} = \infty$$

Towards more general results

Simple idea

Consider sequence of sets S_n such that

$$\sup\Big\{d\big(f,f_0\big) : f \in S_n \ \text{ s.t. } \ d_K\big(Kf,Kf_0\big) \leq \delta\Big\} < \infty$$

Simple observation

Note that the prior in the previous examples did not depend on the inverse problem K, i.e., the sequence κ_i.

Standard choice of S_n

In the sequence setting S_n could be

$$S_n = \left\{ f \in \ell_2 : \sum_{i > k_n} f_i^2 \leq c\rho_n^2 \right\},$$

for given sequences of positive numbers $k_n \to \infty$ and $\rho_n \to 0$ and a constant $c > 0$.

Standard choice of S_n

In the sequence setting S_n could be

$$S_n = \Big\{ f \in \ell_2 : \sum_{i > k_n} f_i^2 \le c\rho_n^2 \Big\},$$

for given sequences of positive numbers $k_n \to \infty$ and $\rho_n \to 0$ and a constant $c > 0$.

If f_0 is such that for some sequence (s_i)

$$\sum_{i=1}^{\infty} f_{0,i}^2 s_i^{-2} < \infty,$$

then

$$\sup \Big\{ d(f, f_0) : f \in S_n \text{ s.t. } d_K(Kf, Kf_0) \le \delta \Big\} \lesssim \kappa_{k_n}^{-1}\delta + \rho_n + s_{k_n}.$$

Modulus of continuity and the contraction theorem

Let

$$\omega(S_n, f_0, d, d_K, \delta) = \sup\Big\{ d(f, f_0) : f \in S_n \;\text{ s.t. }\; d_K(Kf, Kf_0) \leq \delta \Big\}$$

Modulus of continuity and the contraction theorem

Let

$$\omega(S_n, f_0, d, d_K, \delta) = \sup\Big\{ d(f, f_0) : f \in S_n \ \text{ s.t. } \ d_K(Kf, Kf_0) \le \delta \Big\}$$

Theorem (K and Salomond, 201?)

Let $\varepsilon_n \to 0$ and let Π be a prior distribution for f such that

$$\mathrm{E}_0 \Pi(S_n^c \mid Y) \to 0$$

for some sequence of sets (S_n), $S_n \subset \mathscr{F}$, and

$$\mathrm{E}_0 \Pi\Big(d\big(Kf, Kf_0\big) \ge M_n \varepsilon_n \,\big|\, Y\Big) \to 0,$$

for any $M_n \to \infty$. Then

$$\mathrm{E}_0 \Pi\Big(d\big(f, f_0\big) \ge \omega(S_n, f_0, d, d_K, M_n \varepsilon_n) \,\big|\, Y\Big) \to 0.$$

Remaining posterior mass condition

Remaining posterior mass condition: $\mathrm{E}_0 \Pi\big(S_n^c \mid Y\big) \to 0$

Lemma (Lemma 1 in Ghosal and van der Vaart, 2007)

Let $\varepsilon_n \to 0$ and let (S_n) be a sequence of sets $S_n \subset \mathcal{F}$. If Π is the prior distribution on f satisfying

$$\frac{\Pi(S_n^c)}{\Pi(B_n(Kf_0, \varepsilon_n))} \lesssim \exp(-2n\varepsilon_n^2),$$

then

$$\mathrm{E}_0 \Pi\big(S_n^c \mid Y\big) \to 0.$$

Remaining posterior mass condition

Remaining posterior mass condition: $\mathrm{E}_0\Pi\big(S_n^c \mid Y\big) \to 0$

Lemma (Lemma 1 in Ghosal and van der Vaart, 2007)

Let $\varepsilon_n \to 0$ and let (S_n) be a sequence of sets $S_n \subset \mathcal{F}$. If Π is the prior distribution on f satisfying

$$\frac{\Pi(S_n^c)}{\Pi(B_n(Kf_0, \varepsilon_n))} \lesssim \exp(-2n\varepsilon_n^2),$$

then

$$\mathrm{E}_0\Pi\big(S_n^c \mid Y\big) \to 0.$$

Note that in the denominator we have the usual prior mass condition for the posterior contraction in the direct problem of recovering Kf.

Example 0: mildly ill-posed sequence setting

If f is Sobolev β-smooth, then Kf is Sobolev $(\beta + p)$-smooth.

The prior Π_α induces the prior $\bigotimes N(0, i^{-1-2\alpha-2p})$ on Kf.

Already Zhao (2000), and Belitser and Ghosal (2003) showed that in this case the posterior contracts at rate

$$n^{-\frac{(\alpha\wedge\beta)+p}{1+2\alpha+2p}}.$$

Example 0: mildly ill-posed sequence setting

Known:

$$\mathrm{E}_0\Pi\Big(\|Kf - Kf_0\| \geq M_n n^{-\frac{(\alpha\wedge\beta)+p}{1+2\alpha+2p}} \mid Y\Big) \to 0.$$

By taking $S_n = \Big\{f \in \ell_2 : \sum_{i>k_n} f_i^2 \leq c\rho_n^2\Big\}$ with

$$k_n = n^{\frac{1}{1+2\alpha+2p}}, \quad \rho_n = n^{-\frac{(\alpha\wedge\beta)}{1+2\alpha+2p}},$$

we get

$$\begin{aligned}\omega(S_n, f_0, \|\cdot\|, \|\cdot\|, M_n\varepsilon_n) &\lesssim \kappa_{k_n}^{-1}\delta + \rho_n + s_{k_n} \\ &\lesssim M_n n^{\frac{p}{1+2\alpha+2p}} \cdot n^{-\frac{(\alpha\wedge\beta)+p}{1+2\alpha+2p}} + n^{-\frac{(\alpha\wedge\beta)}{1+2\alpha+2p}} + n^{-\frac{\beta}{1+2\alpha+2p}} \lesssim M_n n^{-\frac{(\alpha\wedge\beta)}{1+2\alpha+2p}}.\end{aligned}$$

Example 1: Numerical differentiation

For all $f \in L_1([0, 1])$ define the operator K

$$Kf(x) = \int_0^x f(t)\,dt, \qquad \text{for } x \in [0, 1]$$

Note that the operator K is no longer defined on the Hilbert space. (No SVD)

Example 1: Numerical differentiation

For all $f \in L_1([0,1])$ define the operator K

$$Kf(x) = \int_0^x f(t)\,dt, \qquad \text{for } x \in [0,1]$$

Note that the operator K is no longer defined on the Hilbert space. (No SVD)

Consider the inverse regression problem

$$Y_i = Kf(x_i) + \sigma\epsilon_i, \qquad \epsilon_i \overset{iid}{\sim} N(0,1), \qquad i = 1,\ldots,n,$$

where the design points $x_1, x_2, \ldots, x_n \in [0,1]$ are fixed.

Example 1: Numerical differentiation

Prior

We put a prior on f such that the induced prior on Kf is well known, and we use B-spline basis $(B_{j,q})$ of order q associated with J equally spaced knots.

Example 1: Numerical differentiation

Prior

We put a prior on f such that the induced prior on Kf is well known, and we use B-spline basis $(B_{j,q})$ of order q associated with J equally spaced knots.

We consider the following hierarchical prior on f

$$\Pi = \begin{cases} J \sim \Pi_J \\ a_1, \dots a_J \overset{iid}{\sim} \Pi_{a,J} \\ f(x) = J \sum_{j=1}^{J-1} (a_{j+1} - a_j) B_{j,q-1}(x). \end{cases}$$

Example 1: Numerical differentiation

Prior

We put a prior on f such that the induced prior on Kf is well known, and we use B-spline basis $(B_{j,q})$ of order q associated with J equally spaced knots.

We consider the following hierarchical prior on f

$$\Pi = \begin{cases} J \sim \Pi_J \\ a_1, \dots a_J \stackrel{iid}{\sim} \Pi_{a,J} \\ f(x) = J \sum_{j=1}^{J-1} (a_{j+1} - a_j) B_{j,q-1}(x). \end{cases}$$

We then have

$$Kf(x) = \sum_{j=1}^{J} a_j B_{j,q}(x), \qquad \text{for } x \in [0,1],$$

and this prior is well suited for Hölder smooth functions.

Example 1: Numerical differentiation

Consider matrix Σ_n^q

$$(\Sigma_n^q)_{i,j} = \frac{1}{n} \sum_{l=1}^{n} B_{i,q}(x_l) B_{j,q}(x_l), \qquad i,j = 1, \dots, J.$$

Example 1: Numerical differentiation

Consider matrix Σ_n^q

$$(\Sigma_n^q)_{i,j} = \frac{1}{n}\sum_{l=1}^{n} B_{i,q}(x_l)B_{j,q}(x_l), \qquad i,j = 1,\ldots,J.$$

We require the following conditions on the design $x_1, x_2, \ldots, x_n$

- for all $v_1 \in \mathbb{R}^J$
$$J^{-1}\|v_1\|_J^2 \asymp v_1^T \Sigma_n^q v_1$$
- for all $v_2 \in \mathbb{R}^{J-1}$
$$(J-1)^{-1}\|v_2\|_{J-1}^2 \asymp v_2^T \Sigma_n^{(q-1)} v_2.$$

Example 1: Numerical differentiation

Consider matrix Σ_n^q

$$(\Sigma_n^q)_{i,j} = \frac{1}{n}\sum_{l=1}^{n} B_{i,q}(x_l)B_{j,q}(x_l), \qquad i,j = 1,\dots,J.$$

We require the following conditions on the design $x_1, x_2, \dots, x_n$

- for all $v_1 \in \mathbb{R}^J$

$$J^{-1}\|v_1\|_J^2 \asymp v_1^T \Sigma_n^q v_1$$

- for all $v_2 \in \mathbb{R}^{J-1}$

$$(J-1)^{-1}\|v_2\|_{J-1}^2 \asymp v_2^T \Sigma_n^{(q-1)} v_2.$$

These conditions are natural and satisfied by many designs, including the uniform design.

Example 1: Numerical differentiation

Theorem (K and Salomond, 201?)

Suppose the true f_0 is β-Hölder smooth, for $\beta \leq q-1$. Suppose that Π_J is such that for some constants $c_d, c_u > 0$ and $t \geq 0$,

$$\exp(-c_d j(\log j)^t) \leq \Pi_J(j \leq J \leq 2j), \qquad \Pi_J(J > j) \lesssim \exp(-c_u j(\log j)^t).$$

Suppose that for all $a_0 \in \mathbb{R}^J$, $\|a_0\|_\infty \leq H$, and some constant c

$$\Pi_{a,J}(\|a - a_0\|_J \leq \epsilon) \geq \exp(-cJ\log(1/\epsilon))$$

Then for some $C > 0$ and some $r > 0$

$$\mathrm{E}_0\Pi\Big(\|f - f_0\|_n \geq Cn^{-\frac{\beta}{3+2\beta}}(\log n)^{3r} \;\Big|\; Y\Big) \to 0.$$

Adaptive result, prior does not depend on β

Example 2: Deconvolution

For all $f \in L_2(\mathbb{R})$ define the operator K

$$Kf(x) = \int_{\mathbb{R}} f(t)\lambda(x-t)\,dt, \qquad \text{for } x \in \mathbb{R}$$

with the kernel $\lambda \in L_2(\mathbb{R})$ symmetric around 0.

Example 2: Deconvolution

For all $f \in L_2(\mathbb{R})$ define the operator K

$$Kf(x) = \int_{\mathbb{R}} f(t)\lambda(x-t)\,dt, \qquad \text{for } x \in \mathbb{R}$$

with the kernel $\lambda \in L_2(\mathbb{R})$ symmetric around 0.

Consider the inverse regression problem

$$Y_i = Kf(x_i) + \sigma\epsilon_i, \qquad \epsilon_i \overset{iid}{\sim} N(0,1), \qquad i = 1, \dots, n,$$

where the design points $x_1, x_2, \dots, x_n \in \mathbb{R}$ are fixed and satisfy certain conditions.

Example 2: Deconvolution

Prior

We consider the following hierarchical prior on f

$$\Pi = \begin{cases} J \sim \Pi_J \\ v \sim \Pi_v \\ w_1, \ldots, w_J | J \sim \bigotimes_{j=1}^{J} N(0,1) \\ f(x) = \sum_{j=1}^{J} \frac{w_j}{\sqrt{2\pi v^2}} \exp\Big(-\frac{(x-z_j)^2}{2v^2} \Big) \end{cases}$$

with certain (tail) conditions on Π_j and Π_v, and the nodes $z_1, \ldots, z_J$

Example 2: Deconvolution

Theorem (K and Salomond, 201?)

Suppose the true f_0 is β-Sobolev smooth, for $\beta \in \mathbb{N}$, supported on $[0, 1]$. Suppose that the Fourier transform of the kernel λ satisfies

$$|\hat{\lambda}(t)| \asymp |t|^{-p}, \qquad \text{for } p \in \mathbb{N}.$$

Then for some $C > 0$ and some $r > 0$

$$\mathrm{E}_0 \Pi\Big(\|f - f_0\|_n \geq C n^{-\frac{\beta}{1+2\beta+2p}} (\log n)^r \;\Big|\; Y\Big) \to 0.$$

Adaptive result, prior does not depend on β

Example 3: extremely ill-posed and spectral cut-off

Recovery of the initial condition for the heat equation

The Dirichlet problem for the heat equation

$$\frac{\partial}{\partial t}u(x,t) = \frac{\partial^2}{\partial x^2}u(x,t), \quad u(x,0) = f(x), \quad u(0,t) = u(1,t) = 0,$$

where u is defined on $[0,1] \times [0,T]$ and the function $f \in L_2[0,1]$ satisfies $f(0) = f(1) = 0$

Example 3: extremely ill-posed and spectral cut-off

Recovery of the initial condition for the heat equation

The Dirichlet problem for the heat equation

$$\frac{\partial}{\partial t}u(x,t) = \frac{\partial^2}{\partial x^2}u(x,t), \quad u(x,0) = f(x), \quad u(0,t) = u(1,t) = 0,$$

where u is defined on $[0,1] \times [0,T]$ and the function $f \in L_2[0,1]$ satisfies $f(0) = f(1) = 0$

The solution is given by

$$u(x,T) = \sqrt{2}\sum_i f_i e^{-i^2\pi^2 T}\sin(i\pi x) = \sum_i \kappa_i f_i e_i(x),$$

We observe a noisy version $u(\cdot\,,T)$, want to recover f

Example 3: extremely ill-posed and spectral cut-off

Consider the extremely ill-posed problem

$$Y_i = \kappa_i f_i + \frac{1}{\sqrt{n}} Z_i,$$

where $\kappa_i \asymp \exp(-\gamma i^p)$ for some $\gamma > 0$ and $p \geq 1$.

Example 3: extremely ill-posed and spectral cut-off

Consider the extremely ill-posed problem

$$Y_i = \kappa_i f_i + \frac{1}{\sqrt{n}} Z_i,$$

where $\kappa_i \asymp \exp(-\gamma i^p)$ for some $\gamma > 0$ and $p \geq 1$.

We take $\alpha \geq 0$ and $\xi > -2\gamma$, and consider the product prior

$$\Pi = \bigotimes_{i=1}^{k_n} N\big(0, i^{-\alpha} \exp(-\xi i^p)\big), \qquad k_n = \left(\frac{\log n}{\xi + 2\gamma} + O(\log\log n)\right)^{1/p}.$$

Example 3: extremely ill-posed and spectral cut-off

Consider the extremely ill-posed problem

$$Y_i = \kappa_i f_i + \frac{1}{\sqrt{n}} Z_i,$$

where $\kappa_i \asymp \exp(-\gamma i^p)$ for some $\gamma > 0$ and $p \geq 1$.

We take $\alpha \geq 0$ and $\xi > -2\gamma$, and consider the product prior

$$\Pi = \bigotimes_{i=1}^{k_n} N\big(0, i^{-\alpha} \exp(-\xi i^p)\big), \qquad k_n = \Big(\frac{\log n}{\xi + 2\gamma} + O(\log\log n)\Big)^{1/p}.$$

$$k_n \text{ solves } 1 = n i^{-\alpha} \exp(-(\xi + 2\gamma) i^p)$$

Example 3: extremely ill-posed and spectral cut-off

Theorem (K and Salomond, 201?)

Suppose the true f_0 is β-smooth. Then for every $M_n \to \infty$

$$E_0 \Pi\big(\|f - f_0\| \geq M_n (\log n)^{-\beta/p} \mid Y\big) \to 0.$$

Example 3: extremely ill-posed and spectral cut-off

Theorem (K and Salomond, 201?)

Suppose the true f_0 is β-smooth. Then for every $M_n \to \infty$

$$\mathrm{E}_0\Pi\big(\|f - f_0\| \geq M_n(\log n)^{-\beta/p} \mid Y\big) \to 0.$$

This prior recovers Kf_0 at a possibly suboptimal rate.

Example 3: extremely ill-posed and spectral cut-off

Theorem (K and Salomond, 201?)

Suppose the true f_0 is β-smooth. Then for every $M_n \to \infty$

$$\mathrm{E}_0 \Pi\big(\|f - f_0\| \geq M_n (\log n)^{-\beta/p} \mid Y\big) \to 0.$$

This prior recovers $K f_0$ at a possibly suboptimal rate.

$K f_0$ is analytic and should be recovered at the rate $n^{-1/2}(\log n)^{1/(2p)}$, but the prior considered here leads to the following suboptimal rate

$$(\log n)^{-\frac{\beta}{p} + \frac{\gamma\alpha}{p(\xi+2\gamma)}} n^{-\frac{\gamma}{\xi+2\gamma}},$$

if $\xi > 0$.

Conclusions

Thank you for your attention!