Simultaneous confidence bands with the volume-of-tube formula and spline estimators

Tatyana Krivobokova

joint work with
Thomas Kneib and Gerda Claeskens

University of Vienna

06 June 2011

Outline

1. Introduction
2. Volume-of-tube formula
3. Spline estimators
4. Confidence bands with spline estimators
5. Bayesian confidence bands
6. Extensions and conclusion

Confidence bands

Assuming

$$y_i = f(x_i) + \epsilon_i,\ i = 1, \dots, n$$

for $f \in \mathcal{F}$, $x_i \in \mathcal{X} \subset \mathbb{R}$ we aim to find a random interval $\mathcal{I}(x, \alpha)$ that depends on (y_i, x_i) only and for the significance level α

$$\inf_{f \in \mathcal{F}} P_f \Big\{ f(x) \in \mathcal{I}(x, \alpha), \forall x \in \mathcal{X} \Big\} = 1 - \alpha,$$

$\mathcal{X} \subset \mathbb{R}$ throughout the talk is a finite interval

Confidence bands

Typically $\mathcal{I}(x, \alpha)$ is based on some (linear) estimator $\widehat{f}$ of f

and for $x \in \mathcal{X}$ has the form

$$\left\{ \widehat{f}(x) - z_{n,\alpha}\sqrt{\operatorname{var} \widehat{f}(x)},\ \widehat{f}(x) + z_{n,\alpha}\sqrt{\operatorname{var} \widehat{f}(x)} \right\}$$

with $z_{n,\alpha}$ satisfying

$$\inf_{f \in \mathcal{F}} P_f \left\{ \frac{|\widehat{f}(x) - f(x)|}{\sqrt{\operatorname{var} \widehat{f}(x)}} > z_{n,\alpha},\ \forall x \in \mathcal{X} \right\} = \alpha$$

Suprema of Gaussian processes

If f could be estimated unbiasedly with a linear estimator $\widehat{f}$, that is

$$f \in \mathcal{F} = \left\{ f : \ f(x) = \mathrm{E}\, \widehat{f}(x), \forall x \right\}$$

then

$$\inf_{f \in \mathcal{F}} P_f \left\{ \frac{|\widehat{f}(x) - f(x)|}{\sqrt{\mathrm{var}\, \widehat{f}(x)}} > z_{n,\alpha}, \forall x \in \mathcal{X} \right\} = P \left\{ \sup_{x \in \mathcal{X}} \frac{|\widehat{f}(x) - \mathrm{E}\, \widehat{f}(x)|}{\sqrt{\mathrm{var}\, \widehat{f}(x)}} > z_{n,\alpha} \right\}$$

which can be approximated by

$$P \left\{ \sup_{x \in \mathcal{X}} |Z(x)| > z_{n,\alpha} \right\}$$

for a differentiable zero mean unit variance Gaussian process $Z(x)$

Suprema of Gaussian processes

Under certain regularity conditions on $\mathrm{cov}\{Z(x)\}$, it is known (see e.g. Hall, 1991, PTRF)

$$\liminf_{n\to\infty} \log(n) \inf_{A,B} \sup_{z} \left| P\left\{ \sup_{x\in\mathcal{X}} Z(x)/B - A < z \right\} - \exp(-e^{-z}) \right| > 0$$

Independent of A, B the convergence can not be faster than $\log(n)^{-1}$

However, bootstrap approximations can achieve a faster rate: for kernel estimators with bandwidth h it is $(nh)^{-1/2}(\log n)^2$

If f is estimated with a bias, then

$$\frac{|\widehat{f}(x) - \mathrm{E}\,\widehat{f}(x) + \mathrm{E}\,\widehat{f}(x) - f(x)|}{\sqrt{\mathrm{var}\,\widehat{f}(x)}} \approx \left| Z(x) + \frac{\mathrm{E}\,\widehat{f}(x) - f(x)}{\sqrt{\mathrm{var}\,\widehat{f}(x)}} \right|$$

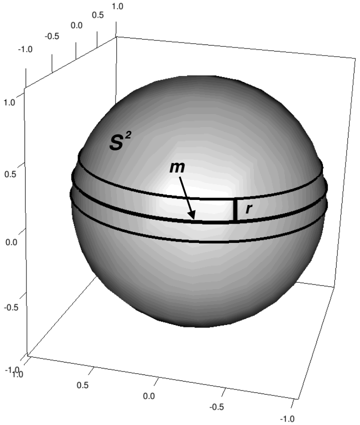
which depends on f

Moreover, any nonparametric estimator $\widehat{f}(x) = \widehat{f}(x, \lambda)$,
with λ estimated with e.g. GCV, introducing extra variability

Smoothing parameter choice is crucial for bootstrap approximations

Volume-of-tube formula

$$V(T_r) = V(S^{n-1})P(U \in T_r)$$



T_r tube of a radius r around m

$m : \mathcal{X} \to S^{n-1}$ regular curve

S^{n-1} unit sphere in $\mathbb{R}^n$

U uniformly distributed over S^{n-1}

$V(S^{n-1}) = 2\,\pi^{n/2}/\Gamma(n/2)$

Volume-of-tube formula

For $U = (U_1, \ldots, U_n)^t$ and $m = \{m_1(x), \ldots, m_n(x)\}^t$, $x \in \mathcal{X}$

$$\begin{aligned} V(T_r)/V(S^{n-1}) &= P(U \in T_r) = P\left(\inf_{x\in\mathcal{X}} \|U - m(x)\|^2 < r^2\right) \\ &= P\left(2\{1 - \sup_{x\in\mathcal{X}} U^t m(x)\} < r^2\right) \\ &= P\left(\sup_{x\in\mathcal{X}} U^t m(x) > 1 - r^2/2\right) \end{aligned}$$

Supremum of a Gaussian process

Since any random variable uniformly distributed over S^{n-1}

$$U = \frac{\epsilon}{\|\epsilon\|},\ \epsilon = (\epsilon_1, \ldots, \epsilon_n)^t \sim \mathcal{N}(0_n, I_n)$$

$$\begin{aligned} P\left\{\sup_{x\in\mathcal{X}} \frac{\epsilon^t}{\|\epsilon\|} m(x) > 1 - \frac{r^2}{2}\right\} &= P\left\{\sup_{x\in\mathcal{X}} \epsilon^t m(x) > \|\epsilon\| \left(1 - \frac{r^2}{2}\right)\right\} \\ &= P\left\{\sup_{x\in\mathcal{X}} Z(x) > \|\epsilon\| \left(1 - \frac{r^2}{2}\right)\right\} \end{aligned}$$

for a zero mean unit variance Gaussian processes $Z(x)$

with $\text{cov}\{Z(x_1), Z(x_2)\} = m(x_1)^t m(x_2)$

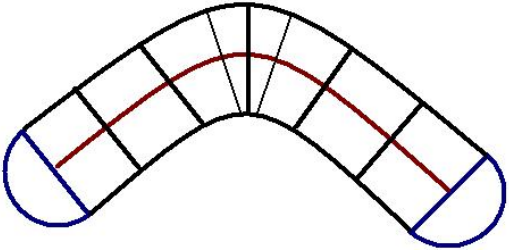
Thus, it holds exactly

$$
\begin{aligned}
P\left\{\sup_{x\in\mathcal{X}} Z(x) > z\right\} &= \int_z^\infty P\left\{\sup_{x\in\mathcal{X}} U^t m(x) > z/\xi\right\} g(\xi,n)d\xi \\
&= \int_z^\infty V(T_{r^*})/V(S^{n-1})g(\xi,n)d\xi
\end{aligned}
$$

for $r^* = \sqrt{2(1-z/\xi)}$ and

$g(\xi, n)$ as the density of a χ_n distributed random variable

Weyl's (1939) formula



$$V(T_r) = \frac{\pi^{\frac{n-2}{2}}}{\Gamma(\frac{n}{2})}\kappa_0 \left(r^2 - \frac{r^4}{4}\right)^{\frac{n-1}{2}} + \frac{2\pi^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})}\int_{1-\frac{r^2}{2}}^{1} (1-x^2)^{\frac{n-3}{2}} dx$$

with $\kappa_0 = \int_{\mathcal{X}} \|m'(x)\| dx$ as the length of a non-closed curve m

The formula is conservative if T_r has self-overlap

Tail approximation

Let $Z(x)$ be a zero mean unit variance Gaussian process with $\text{cov}\{Z(x_1), Z(x_2)\} = m(x_1)^t m(x_2)$, $m \in C^1$

Then for a non-closed curve m and $z \to \infty$

$$P\left\{\sup_{x\in\mathcal{X}} |Z(x)| > z\right\} = \frac{\kappa_0}{\pi} \exp\left(-\frac{z^2}{2}\right) + 2\{1 - \Phi(z)\} + o\left(e^{-z^2/2}\right)$$

with $\kappa_0 = |m| = \int_{\mathcal{X}} \|m^{'}(x)\| dx$ as the length of m and $\Phi(\cdot)$ as the c.d.f. of the standard normal distribution

Rigorous proofs and cases $\mathcal{X} \subset \mathbb{R}^d$, $d > 1$ are given in Sun (1993, AoS) and Sun & Loader (1994, AoS)

Some references

In the regression context the simultaneous confidence bands based on the volume-of-tube formula have been constructed using

- local polynomials (Sun & Loader, 1994 AoS)
- regression (least squares) splines (Zhou, Shen, Wolfe, 1998, AoS)

Both approaches ignored

- data-driven smoothing parameter
- bias

resulting in the coverage about $5 - 10\%$ less than the nominal

From the data pairs (y_i, x_i) that follow

$$y_i = f(x_i) + \varepsilon_i, \; x_i \in \mathcal{X}, \; \varepsilon_i \sim \mathcal{N}(0, \sigma^2), \; i = 1, \ldots, n$$

we estimate $f \in W_2^q(\mathcal{X})$ with penalized splines, that is solving

$$\min_{s \in \mathcal{S}(p,k)} \left[\frac{1}{n} \sum_{i=1}^{n} \{y_i - s(x_i)\}^2 + \lambda \int_{\mathcal{X}} \{s^{(q)}(x)\}^2 dx \right]$$

with $\mathcal{S}(p, k)$ as the spline space of degree p based on k knots

For $p = 2q - 1, \; k = n$ the solution is the smoothing spline estimator

Spline estimators

Denoting with $N(x) = \{N_1(x), \ldots, N_{k+p+1}(x)\}$ some basis in $\mathcal{S}(p,k)$, so that $s(x) = N(x)\beta \in \mathcal{S}(p,k)$, the spline estimator of f

$$\widehat{f}(x) = N(x)\widehat{\beta} = N(x)(N^tN + \lambda n D)^{-1}N^tY =: \ell(x,\lambda)^tY$$

with $D = \int_{\mathcal{X}} N^{(q)}(x)N^{(q)}(x)^t dx$ and $Y = (y_1, \ldots, y_n)^t$

For a fixed $\lambda > 0$ spline estimator $\widehat{f}(x)$ is a linear estimator of f

In practice λ is estimated by GCV or AIC from the data

Some asymptotics

Reference: Claeskens, K., Opsomer (2009, Biometrika)

Low rank spline estimators ($k \ll n$) have two parameters: k and λ

Under standard assumptions (regularity of data and knots,
$k = o(n)$, $\lambda \to 0$ with $\lambda n \to \infty$) λ is identifiable if $\lambda^{1/(2q)} k \to \infty$

$k = c(q, f, \sigma) n^{\nu/(2q+1)}$ for $\nu > 1$ and $c(q, f, \sigma) > 0$ ensures that

- λ is identifiable
- bias due to approximation of f by $s \in \mathcal{S}(p, k)$ is negligible decreasing with $k^{-2q\nu}$

Some asymptotics

For k fixed and satisfying $k = c(q, f, \sigma)n^{\nu/(2q+1)}$ the rate for

$$\lambda_0 = \arg\min_{\lambda>0} \mathrm{E}_f \|\widehat{f}(\lambda) - f\|_{n,2}^2$$

with $\|f\|_{n,2}^2 = n^{-1}\sum_{i=1}^n f(x_i)^2$, depends on certain conditions on f

For $f \in \mathcal{A}_r^q = \left\{ f \in W_2^q, n^{-1}\sum_{i=1}^k \tilde{f}_i^2 n^r i^{2qr} < a_r < \infty \right\}$, $r \in [1, 2]$

$$\lambda_0 = O\left(n^{-\frac{2q}{2qr+1}}\right) \quad \text{and} \quad \mathrm{E}_f \|\widehat{f}(\lambda_0) - f\|_{n,2}^2 = O\left(n^{-\frac{2qr}{2qr+1}}\right)$$

with $\tilde{f}$ as generalized Fourier coefficients of f

In particular, if $f \in W_2^{2q}$ and natural or periodic boundary conditions hold then $f \in \mathcal{A}_2^q$ and

$$\lambda_0 = \left[n\, 4q \|f^{(2q)}\|_2^2 \, \tilde{c}(q,\sigma)\{1 + o(1)\} \right]^{-\frac{2q}{4q+1}}$$

Summarizing

- once large enough, k does not matter
- spline estimators "adapt" to unknown (boundary) properties of f
- larger λ_0 (smaller bias) for "smoother" f

Confidence bands with spline estimators

As for any linear smoother for a spline estimator $\widehat{f}(x) = \ell(x,\lambda)^t Y$

$$
\begin{aligned}
\frac{|\widehat{f}(x) - f(x)|}{\sqrt{\mathrm{var}\widehat{f}(x)}} &= \frac{|\ell(x,\lambda)^t Y - \ell(x,\lambda)^t f + \ell(x,\lambda)^t f - f(x)|}{\sigma\|\ell(x,\lambda)\|} \\
&= \left|\frac{\varepsilon^t \ell(x,\lambda)}{\sigma\|\ell(x,\lambda)\|} + \frac{\ell(x,\lambda)^t f - f(x)}{\sigma\|\ell(x,\lambda)\|}\right| \\
&= |Z(x,\lambda) + \delta(x,\lambda,f)|
\end{aligned}
$$

for $\varepsilon = \{y_1 - f(x_1), \ldots, y_n - f(x_n)\}^t$ and standardized bias $\delta(x,\lambda,f)$

If λ were known and $\delta(x, \lambda, f) = 0$ then

$$\begin{aligned}
\alpha &= P\left\{\sup_{x \in \mathcal{X}} \frac{|\varepsilon^t \ell(x, \lambda)|}{\sigma \|\ell(x, \lambda)\|} > z\right\} = P\left\{\sup_{x \in \mathcal{X}} \frac{|\widehat{f}(x) - f(x)|}{\sigma \|\ell(x, \lambda)\|} > z\right\} \\
&= \frac{\kappa_0}{\pi} \exp\left(-\frac{z^2}{2}\right) + 2\{1 - \Phi(z)\} + o\left(e^{-z^2/2}\right)
\end{aligned}$$

with κ_0 as the length of $\ell(x, \lambda)/\|\ell(x, \lambda)\|$, resulting in

$$\left\{\widehat{f}(x, \lambda) - z\sigma\|\ell(x, \lambda)\|,\ \widehat{f}(x, \lambda) + z\sigma\|\ell(x, \lambda)\|\right\}$$

Confidence band width

The width of the confidence band is determined by

$z\sigma\|\ell(x,\lambda)\|$ with $z = \sqrt{\log(\kappa_0^2\{1 + O(1)\})}$

It is straightforward to show that for $c_1(p), c_2(p) > 0$

$\kappa_0 = c_1(p)k$ and $\|\ell(x,\lambda)\| \leq c_2(p)n^{-1}\lambda^{-1/(2q)}$

So that the width of the band is

$$O_p\left(n^{-1}\lambda^{-1/(2q)}\sqrt{\log k^2}\right)$$

Confidence band width

If the band is constructed with λ_0, then its width varies between

$$O_p\left(n^{-q/(2q+1)}\sqrt{\log k^2}\right) \text{ and } O_p\left(n^{-2q/(4q+1)}\sqrt{\log k^2}\right)$$

being narrower for "smoother" functions

A narrower band can also be obtained by taking a smaller k, which satisfies $k = c(q, f, \sigma)n^{\nu/(2q+1)}$, $\nu > 1$, $c(q, f, \sigma) > 0$

Variability due to $\widehat{\lambda}$

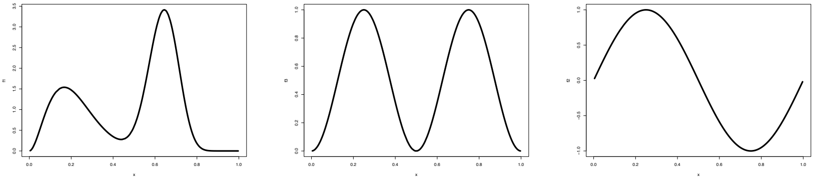
Under mild regularity assumptions

$$\frac{\ell(x,\widehat{\lambda})^t\varepsilon}{\sigma\|\ell(x,\widehat{\lambda})\|} = \frac{\ell(x,\lambda)^t\varepsilon}{\sigma\|\ell(x,\lambda)\|} + O_p\left(\lambda^{\frac{1}{4q}}\right)$$

with a smaller error term for smaller q and smaller λ

Simulation setting

The volume-of-tube formula is applied ignoring the bias and the variability due to estimation of λ with GCV to the three functions



Spline estimators are based on B-splines basis with $p = 3$, $q = 2$

Residual variance is taken $\sigma = 0.3$,
similar results were obtained for $\sigma = 0.1$ and $\sigma = 0.6$

For the nominal coverage of 0.95 using 1000 Monte Carlo samples

	$n = 50$		$n = 250$		$n = 500$	
	$k = 15$	40	$k = 40$	100	$k = 40$	200
f_1	0.91	0.86	0.88	0.90	0.90	0.89
	(0.91)	(0.91)	(0.44)	(0.45)	(0.33)	(0.33)
f_2	0.85	0.86	0.73	0.81	0.88	0.88
	(0.70)	(0.71)	(0.32)	(0.33)	(0.25)	(0.25)
f_3	0.93	0.91	0.88	0.92	0.92	0.92
	(0.64)	(0.69)	(0.29)	(0.29)	(0.20)	(0.20)

- undercoverage independent of n
- k has little influence

Centering around E $\widehat{f}$

	$n=50$		$n=250$		$n=500$	
	$k=15$	40	$k=40$	100	$k=40$	200
f_1	0.91	0.86	0.88	0.90	0.90	0.89
E $\widehat{f_1}$	0.96	0.97	0.96	0.94	0.95	0.94
f_2	0.85	0.86	0.73	0.81	0.88	0.88
E $\widehat{f_2}$	0.95	0.95	0.96	0.94	0.95	0.94
f_3	0.93	0.91	0.88	0.92	0.92	0.92
E $\widehat{f_3}$	0.94	0.94	0.95	0.95	0.93	0.94

Bayesian model for smoothing

Representing $s(x) = N(x)\beta = X(x)\alpha + B(x)b$, where $X(x) = \sum_{i=0}^{q-1} \alpha_i x^i$, $\dim(b) = k + p + 1 - q =: \tilde{k}$ and assuming

$$y_i = X(x_i)\alpha + B(x_i)b + \varepsilon_i,\ b \sim \mathcal{N}(0_k, \sigma_b^2 D_{\tilde{k}}^{-1}),\ \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

leads to a standard linear mixed model (or empirical Bayes model)

Putting priors on α, σ_b^2 and σ^2 leads to a full Bayesian model for s

Bayesian estimator

The best linear predictor (BLUP) of f is known to have the form

$$\widetilde{f}(x) = N(x)\widetilde{\beta} = N(x)(N^t N + \sigma^2/\sigma_b^2 D)^{-1} N^t Y = \ell(x, \lambda_b)^t Y$$

with $\lambda_b = \sigma^2/(n\sigma_b^2)$ estimated either

from the likelihood (empirical Bayes) with a mixed model software or using full Bayesian techniques based on MCMC

With respect to the marginal distribution of Y

$$Z_b(x) = \frac{\widetilde{f}(x) - f(x)}{\sqrt{\operatorname{var}\left\{\widetilde{f}(x) - f(x)\right\}}} \approx \frac{N(x)(\widetilde{\beta} - \beta)}{\sqrt{\operatorname{var}\left\{N(x)(\widetilde{\beta} - \beta)\right\}}} \sim \mathcal{N}(0,1)$$

Denoting $\ell_b(x, \lambda_b) = (N^t N + \sigma^2/\sigma_b^2 D)^{-1/2} N(x)^t$ and $\varepsilon_b = (N^t N + \sigma^2/\sigma_b^2 D)^{1/2}(\tilde{\beta} - \beta) \sim \mathcal{N}(0, \sigma^2 I_{\tilde{k}})$

$$Z_b(x) = \frac{\ell_b(x, \lambda_b)^t \varepsilon_b}{\sigma \|\ell_b(x, \lambda_b)\|} \sim \mathcal{N}(0,1)$$

For a known λ_b it holds

$$\begin{aligned}\alpha &= P\left\{\sup_{x\in\mathcal{X}} |Z_b(x)| > z_b\right\} \\ &= P\left\{\sup_{x\in\mathcal{X}} \frac{|\widetilde{f}(x,\lambda_b) - f(x)|}{\sigma\|\ell_b(x,\lambda_b)\|} > z_b\right\} \\ &= \frac{\kappa_{b,0}}{\pi}\exp\left(-\frac{z_b^2}{2}\right) + 2\{1-\Phi(z_b)\} + o\left(e^{-z_b^2/2}\right)\end{aligned}$$

$\kappa_{b,0}$ is the length of $\ell_b(x,\lambda_b)/\|\ell_b(x,\lambda_b)\|$ (dimension $k+p+1$)

Bayesian confidence bands

The Bayesian confidence band for $x \in \mathcal{X}$

$$\left\{ \widetilde{f}(x, \lambda_b) - z_b \sigma \|\ell_b(x, \lambda_b)\|, \ \widetilde{f}(x, \lambda_b) + z_b \sigma \|\ell_b(x, \lambda_b)\| \right\}$$

Note $\|\ell_b(x, \lambda_b)\|^2 = N(x)(N^t N + \lambda_b D)^{-1} N(x)^t$ while

$\|\ell(x, \lambda)\|^2 = N(x)(N^t N + \lambda D)^{-1} N^t N (N^t N + \lambda D)^{-1} N(x)^t$

How much the variability due to estimation of λ_b matters?

Simulation results

	$n = 50$		$n = 250$		$n = 500$	
	$k = 15$	40	$k = 40$	100	$k = 40$	200
f_1 EB	0.96	0.97	0.98	0.99	0.99	0.99
	(0.97)	(1.04)	(0.56)	(0.58)	(0.43)	(0.44)
f_1 FB	0.95	0.96	0.98	0.99	0.99	1.00
	(0.96)	(1.01)	(0.56)	(0.57)	(0.43)	(0.44)
f_2 EB	0.97	0.97	0.98	0.99	1.00	1.00
	(0.75)	(0.78)	(0.42)	(0.42)	(0.32)	(0.32)
f_2 FB	0.96	0.97	0.99	0.99	0.99	0.99
	(0.76)	(0.78)	(0.42)	(0.43)	(0.32)	(0.33)
f_3 EB	0.97	0.98	0.99	1.00	0.99	0.99
	(0.70)	(0.71)	(0.37)	(0.37)	(0.28)	(0.28)
f_3 FB	0.97	0.98	1.00	1.00	0.99	0.99
	(0.71)	(0.72)	(0.38)	(0.38)	(0.29)	(0.29)

Simulation results

The confidence bands were obtained

- from the MCMC samples in the full Bayesian model
- from the volume-of-tube formula in the empirical Bayesian model

The results suggest that

- the volume-of-tube formula is the attractive alternative to MCMC
- both bands are conservative for $f \in W_2^q$ independent of n and k

Bayesian bands are based on the posterior probability that a particular realisation of a given stochastic process is in the band, given the data

$$N\beta | y_1, \ldots y_n \sim \mathcal{N}\left(N\tilde{\beta}, \sigma^2 N(N^t N + \sigma^2/\sigma_b^2 D)^{-1} N^t\right)$$

It is known that the sample paths of Bayesian smoothing splines are not in W_2^q with probability 1 (e.g. Wahba, 1978, JRSSB)

Hence, the Bayesian bands consist a.s. of functions outside W_2^q

Bayesian nonparametric regression

Zhao (2000, AoS) considered a Gaussian white noise mode, assuming the regression function to lay in W^q and found that

“ there is no independent normal prior ... with support on W^q such that the corresponding Bayes estimator attains the optimal rate ...“

A certain mixture of Gaussians is proved to have these properties, but it depends crucially on the unknown q

Define

$$\lambda_b^0 = \arg\min_{\lambda_b > 0} \mathrm{E}_f \left\{ -l_p(\lambda_b; y_1, \ldots, y_n) \right\}$$

with $l_p(\lambda_b; y)$ as the profiled log-likelihood in the Bayesian model

Then for $f \in \mathcal{A}_r^q$, $r \in [1, 2]$

$$\lambda_b^0 = \left[n\, \tilde{c}(q, \sigma) \| f^{(q)} \|_2^2 \{1 + o(1)\} \right]^{-\frac{2q}{2q+1}}$$

Thus, for $f \in \mathcal{A}_r^q$, $r \in [1, 2]$

$$\frac{\lambda_0}{\lambda_b^0} = O\left\{ n^{\frac{4q^2(r-1)}{(2qr+1)(2q+1)}} \right\}$$

Smoothing parameter estimators

Let $\widehat{\lambda}_b$ and $\widehat{\lambda}_0$ be the estimators of λ_b^0 and λ_0

Many simulation studies (e.g. Kohn, Ansley, Tharm, 1991, JASA) show that in finite samples $\widehat{\lambda}_b$ is competitive with $\widehat{\lambda}_0$

This can be attributed to the huge variance of $\widehat{\lambda}_0$ (K., 2011)

$$\frac{\mathrm{var}(\widehat{\lambda}_0)}{\mathrm{var}(\widehat{\lambda}_b^0)} \approx 11q \left(\frac{\lambda_0}{\lambda_b^0}\right)^{2+1/(2q)}$$

Thus, using $\widehat{\lambda}_b^0$ for confidence bands may be preferable

The Bayesian bands are based on the marginal distribution of the data

$$\begin{aligned}
\alpha &= P\left\{\sup_{x\in\mathcal{X}} \frac{|\ell(x,\lambda_b)^t y - f(x)|}{\sigma\|\ell_b(x,\lambda_b)\|} > z_b\right\} \\
&= P\left\{\sup_{x\in\mathcal{X}} \frac{|\ell(x,\lambda_b)^t y - f(x)|}{\sigma\|\ell(x,\lambda_b)\|}\frac{\|\ell(x,\lambda_b)\|}{\|\ell_b(x,\lambda_b)\|} > z_b\right\}
\end{aligned}$$

If the bands are build based on the conditional distribution then

$$\alpha = P_f\left\{\sup_{x\in\mathcal{X}} \frac{|\ell(x,\lambda_b)^t y - f(x)|}{\sigma\|\ell(x,\lambda_b)\|} > z_b^*\right\}$$

If one would still use a $z_b \leq z_b^*$ as a critical value in the band

$$\left\{\widehat{f}(x, \lambda_b) - z_b \sigma \|\ell(x, \lambda_b)\|,\ \widehat{f}(x, \lambda_b) + z_b \sigma \|\ell(x, \lambda_b)\|\right\}$$

then it would be too narrow for a stochastic f

If $f \in W_2^q$ then λ_b^0 undersmooths $\widehat{f}(\lambda_b^0)$ for about the "right" amount

$$\frac{|\ell(x, \lambda_b^0)^t y - f(x)|}{\sigma \|\ell(x, \lambda_b^0)\|} = \frac{|\ell(x, \lambda_0)^t y - f(x)|}{\sigma \|\ell(x, \lambda_0)\|} \frac{\|\ell(x, \lambda_0)\|}{\|\ell_b(x, \lambda_0)\|} \{1 + o_p(1)\}$$

	$n = 50$		$n = 250$		$n = 500$	
	$k = 15$	40	$k = 40$	100	$k = 40$	200
f_1	0.92 (0.90)	0.94 (0.93)	0.96 (0.50)	0.95 (0.50)	0.96 (0.38)	0.96 (0.39)
f_2	0.93 (0.68)	0.93 (0.69)	0.96 (0.37)	0.94 (0.37)	0.96 (0.28)	0.95 (0.28)
f_3	0.95 (0.63)	0.95 (0.64)	0.96 (0.33)	0.97 (0.33)	0.97 (0.25)	0.97 (0.25)

Very similar results were obtained for samples sizes n up to 5000, different signal-to-noise ratios and different functions

In case of heteroscedastic data, the residual variance can be modelled as a smooth function with the (empirical) Bayesian splines

$$
\begin{aligned}
y_i &= X(x_i)\alpha + B(x_i)b + \varepsilon_i, \; b \sim \mathcal{N}\left(0_k, \sigma_b^2 D_{\tilde{k}}^{-1}\right) \\
\varepsilon_i &\sim \mathcal{N}\left(0, \exp\left\{X_\varepsilon(x_i)\gamma + B_\varepsilon(x_i)d\right\}\right), \; d \sim \mathcal{N}\left(0, \sigma_\varepsilon^2 D_\varepsilon^{-1}\right)
\end{aligned}
$$

Similarly σ_b^2 can be estimated as a smooth function, leading to a spatially adaptive smoothing parameter

All model parameters are estimated from the corresponding likelihood and the volume-of-tube formula can be applied as usual

Simultaneous confidence bands can also be used to test

$$\mathrm{H}_0: \ f(x) = X(x)\alpha \quad \text{vs} \quad \mathrm{H}_1: \ f(x) = X(x)\alpha + B(x)b, \ \ \forall x \in \mathcal{X}$$

for a $(q-1)$-degree polynomial $X(x)$

The goodness-of-fit test is performed by building a confidence band around $B(x)b$ and checking if it uniformly encloses the zero line

Simulations showed that this test is at least as good as, or even outperforms, the likelihood-ratio test

All bands and tests are implemented in the R-package `AdaptFitOS`

Conclusion

Confidence bands based on the volume-of-tube formula

- are very simple and fast to obtain in practice
- relay on $\varepsilon \sim \mathcal{N}(0_n, \sigma^2 I_n)$
- can be used in the Bayesian framework

Combination with (empirical) Bayesian spline estimators allows

- to use a more stable smoothing parameter estimator
- for a simple (but slightly conservative) bias correction
- for simple extensions to more complicated models