BNP NETWORK

NICOSIA

april 2022

Lecture 2

Bayesian Uncertainty Quantification

a review

Aad van der Vaart

TU Delft

April 2022

Semiparametric BvM

Castillo

Castillo & Rousseau

Franssen & Nguyen & vdV

Ray & vdV

Nickl et al.

Semiparametric BvM

$\theta \sim \Pi$ on $\mathbb{R}^1$, $\eta \sim \Pi_\eta$ (nuisance parameter, general)

$X_1 \cdots X_n \mid \theta, \eta \overset{iid}{\sim} P_{\theta,\eta}$

Semiparametric BvM

$\theta \sim \Pi$ on $\mathbb{R}^1$, $\eta \sim \Pi_\eta$ (nuisance parameter, general)

$X_1 \cdots X_n \mid \theta, \eta \overset{iid}{\sim} P_{\theta,\eta}$

$\dot{\ell}_{\theta,\eta} = \frac{\partial}{\partial \theta} \log p_{\theta\eta}$ ordinary score

$g = \frac{\partial}{\partial t}\Big|_{t=0} \log p_{\theta,\eta_t}$ $\quad {}_\eta\dot{\mathcal{P}}_{\theta,\eta} = \{\text{all scores for } \eta\}$

for curve $t \mapsto \eta_t$ with $\eta_0 = \eta$

$\Pi_{\theta\eta} \dot{\ell}_{\theta\eta} = \underset{g \in \overline{\text{lin}}\, {}_\eta\dot{\mathcal{P}}_{\theta\eta}}{\operatorname{argmin}} \; P_{\theta\eta} (\dot{\ell}_{\theta\eta} - g)^2$ projection onto $\overline{\text{lin}}\, {}_\eta\dot{\mathcal{P}}_{\theta\eta}$

$\tilde{\ell}_{\theta,\eta} := \dot{\ell}_{\theta,\eta} - \Pi_{\theta\eta} \dot{\ell}_{\theta\eta}$ $\quad \tilde{i}_{\theta\eta} = \mathrm{var}_{\theta,\eta} (\tilde{\ell}_{\theta,\eta}(X))$

efficient score $\quad$ efficient information

Semiparametric BvM

$\theta \sim \Pi$ on $\mathbb{R}^1$, $\eta \sim \Pi_\eta$ (nuisance parameter, general)

$X_1 \cdots X_n \mid \theta, \eta \overset{iid}{\sim} P_{\theta,\eta}$

$\dot{\ell}_{\theta,\eta} = \frac{\partial}{\partial\theta} \log p_{\theta\eta}$ ordinary score

$g = \frac{\partial}{\partial t}\Big|_{t=0} \log p_{\theta,\eta_t}$ $\qquad {}_\eta\dot{\mathcal{P}}_{\theta,\eta} = \{$all scores for $\eta\}$

for curve $t \mapsto \eta_t$ with $\eta_0 = \eta$

$\Pi_{\theta\eta}\dot{\ell}_{\theta\eta} = \underset{g \in \overline{\text{lin}}\, {}_\eta\dot{\mathcal{P}}_{\theta\eta}}{\operatorname{argmin}} P_{\theta\eta}(\dot{\ell}_{\theta\eta} - g)^2$ projection onto $\overline{\text{lin}}\, {}_\eta\dot{\mathcal{P}}_{\theta\eta}$

$\tilde{\ell}_{\theta,\eta} := \dot{\ell}_{\theta,\eta} - \Pi_{\theta\eta}\dot{\ell}_{\theta\eta}$ efficient score $\qquad \tilde{i}_{\theta\eta} = \operatorname{var}_{\theta,\eta}(\tilde{\ell}_{\theta,\eta}(X))$ efficient information

DESIRED THM For $\hat{\theta}_n$ with $\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{i}_{\theta_0\eta_0}^{-1} \tilde{\ell}_{\theta_0\eta_0}(X_i) + o_{P_{\theta_0\eta_0}}(1)$

$$\sqrt{n}(\theta - \hat{\theta}_n) \mid X_1 \cdots X_n \rightsquigarrow N(0, \tilde{i}_{\theta_0\eta_0}^{-1}) \text{ in } P^n_{\theta_0\eta_0}\text{-probability.}$$

Gives correct credible intervals, as before.

Bernstein-von Mises - TV

(Castillo, 2008)

THM If $\exists\ \eta \mapsto \tilde{\eta}_n(\theta,\eta)$ with

- $$\sum_{i=1}^{n} \log \frac{p_{\theta,\eta}}{p_{\theta_0,\tilde{\eta}_n(\theta,\eta)}}(X_i) = \sqrt{n}(\theta-\theta_0)\mathbb{G}_n \tilde{\ell}_{\theta_0,\eta_0} - \frac{1}{2} n \tilde{I}_{\theta_0,\eta_0}(\theta-\theta_0)^2 + R_n(\theta,\eta), \text{ for}$$

$$\sup_{\substack{\theta\in\Theta_n \\ \eta\in\mathcal{H}_n}} \frac{R_n(\theta,\eta)}{1+n|\theta-\theta_0|^2} \to 0$$

likelihood expansion

- $$\sup_{\substack{\theta\in\Theta_n \\ \eta\in\mathcal{H}_n}} \frac{\log \frac{d\Pi_\eta \circ \tilde{\eta}_n(\theta,\eta)^{-1}}{d\Pi_\eta \circ \tilde{\eta}_n(\theta_0,\eta)^{-1}}}{1+n(\theta-\theta_0)^2} \to 0$$

change of measure

- $\Pi_n(\theta\in\Theta_n, \eta\in\mathcal{H}_n \mid X_1\cdots X_n) \to 1$

  $\Pi_n(\eta \in \tilde{\eta}_n(\theta,\mathcal{H}_n) \mid X_1 - X_n, \theta=\theta_0) \to 1$

localisation

then desired theorem holds in TV.

$\tilde{\eta}_n$ least favourable, expansion in both θ and η, typically need rate on η.

If estimating θ does not become more difficult by not knowing η,
then likelihood expansion will only be in θ and change of measure condition disappears.

→ BvM valid under weak conditions, prior on η unimportant.

Example: estimate point of symmetry θ in symmetric density $p_{\theta,\eta}(x) = \eta(|x-\theta|)$.

Otherwise the prior on η may cause a bias in the posterior for θ
if not chosen with care (oversmoothing)

EXAMPLE – COX MODEL

(Castillo, 2008)

$\log \lambda \sim$ Riemann-Liouville (α) ($\alpha - \frac{1}{2}$ times integrated BM)

$\theta \perp\!\!\!\perp \lambda$

$T \mid Z \sim$ hazard $\lambda(t) e^{\theta Z}$

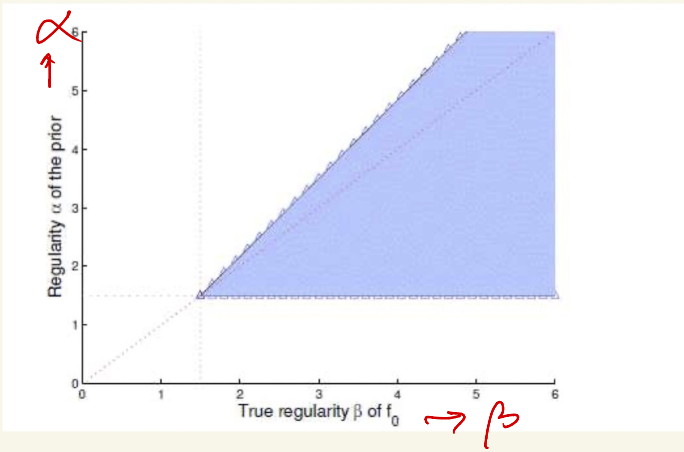
$T \perp\!\!\!\perp C \mid Z$

$X = (T \wedge C, Z, \Delta = 1_{T \le C})$

least favourable

THM If $\log \lambda_0 \in C^{\beta}[0,\tau]$, $\beta > 3/2$, $\gamma_0 \in C^{2\beta/3}[0,\tau]$

then BvM holds for $\alpha \in \left(3/2, \frac{4\beta}{3} - \frac{1}{2}\right)$



Bernstein-von Mises - weak

(Castillo & Rousseau, 2015)

THM if $\exists\ t \mapsto \eta_t(\theta,\eta)$ with $\forall t$

$$\sum_{i=1}^n \log \frac{p_{\theta+n^{-1/2}t,\ \eta_{n^{-1/2}t}(\theta,\eta)}}{p_{\theta,\eta}}(X_i) = \frac{t}{\sqrt{n}} \sum_{i=1}^n \tilde{l}_{\theta_0\eta_0}(X_i) - t\left(\tilde{i}_{\theta_0\eta_0} + o_p(1)\right)\sqrt{n}(\theta-\theta_0) - \frac{1}{2}t^2 \tilde{i}_{\theta_0\eta_0} + o_p(1)$$

likelihood expansion

$$\frac{\int_{\Theta_n}\int_{\mathcal{H}_n} \prod_{i=1}^n p_{\theta+n^{-1/2}t,\ \eta_{n^{-1/2}t}(\theta,\eta)}(X_i)\, d\Pi(\theta)\, d\Pi_\eta(\eta)}{\int_{\Theta_n}\int_{\mathcal{H}_n} \prod_{i=1}^n p_{\theta,\eta}(X_i)\, d\Pi(\theta)\, d\Pi_\eta(\eta)} \xrightarrow{P^n_{\theta_0\eta_0}} 1,$$

change of measure

$\Pi(\theta\in\Theta_n, \eta\in\mathcal{H}_n \mid X_1\cdots X_n) \to 1$

then $\sqrt{n}(\theta-\hat{\theta}_n) \mid X_1\cdots X_n \rightsquigarrow N(0, \tilde{i}^{-1}_{\theta_0\eta_0})$ in $P^n_{\theta_0\eta_0}$-probability

centering as before

Conditions might be true if

- $\frac{\partial}{\partial t}\Big|_{t=0} \log p_{\theta+t,\ \eta_t(\theta,\eta)} = \tilde{l}_{\theta,\eta}$
- $\sup_{\eta\in\mathcal{H}_n} \sqrt{n}\, P_{\theta_0\eta_0} \tilde{l}_{\theta_0,\eta} \xrightarrow{P_{\theta_0\eta_0}} 0$ ← need rate on η (?)

EXAMPLES - MIXTURES

(Franssen & Ngagen & van der Vaart, 2022)

$\eta \sim DP(\alpha)$

$$p_{\theta,\eta}(x,y) = \int z e^{-zx}\, z\theta e^{-z\theta y}\, d\eta(z)$$

$$p_{\theta,\eta}(x,y) = \int \varphi(x-z)\, \varphi(y-\theta_1-\theta_2 z)\, d\eta(z)$$

- likelihood expansion using empirical process theory
- change of measure valid because $DP(\alpha)$ gives few distinct values z.
- no rate on η needed by convex structure of model

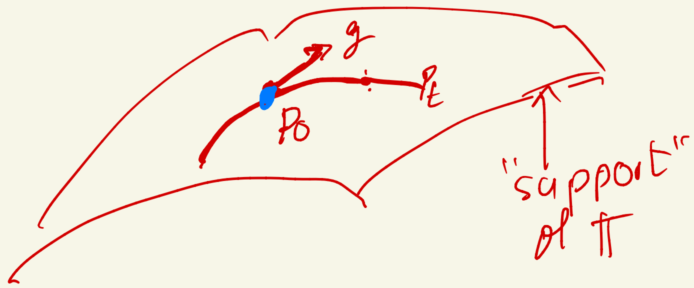
Semiparametric BvM – Functionals

$p \sim \Pi$

$X_1 \cdots X_n \mid p \overset{iid}{\sim} \Pi$

Estimate $\chi(p) \in \mathbb{R}$

Semiparametric BvM – Functionals



$p \sim \Pi$

$X_1 \cdots X_n \mid p \overset{iid}{\sim} \Pi$

Estimate $\chi(p) \in \mathbb{R}$

$$\dot{\mathcal{P}}_{p_0} = \left\{ g = \frac{\partial}{\partial t}\Big|_{t=0} \log p_t \ , \ \text{some submodel } t \mapsto p_t \right\}$$

$\tilde{\chi}_{p_0}$ such that $\frac{d}{dt}\Big|_{t=0} \chi(p_t) = \int g \, \tilde{\chi}_{p_0} \, dP_0$

efficient influence function

Semiparametric BvM – Functionals

$p \sim \Pi$

$X_1 \cdots X_n \mid p \overset{iid}{\sim} \Pi$

Estimate $\chi(p) \in \mathbb{R}$

$$\dot{\mathcal{P}}_{p_0} = \left\{ g = \frac{\partial}{\partial t|_{t=0}} \log p_t \text{ , some submodel } t \mapsto p_t \right\}$$

$\tilde{\chi}_{p_0}$ such that $\frac{d}{dt|_{t=0}} \chi(p_t) = \int g \, \tilde{\chi}_{p_0} \, dP_0$ and $\tilde{\chi}_{p_0} \in \dot{\mathcal{P}}_{p_0}$

efficient influence function

DESIRED THM For $\hat{\chi}_n$ with $\sqrt{n}\left(\hat{\chi}_n - \chi(p_0)\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\chi}_{p_0}(X_i) + o_{P_0}(1)$,

$$\sqrt{n}\left(\chi(p) - \hat{\chi}_n\right) \mid X_1 \cdots X_n \rightsquigarrow N\left(0, P_0 \tilde{\chi}_{p_0}^2\right), \text{ in } P_0^n\text{-probability}$$

(efficient information)$^{-1}$

Valid under similar conditions as before.

Missing Data / Causality

$Y \perp\!\!\!\perp A \mid Z$, $A \in \{0,1\}$. Observe Y iff $A=1$. Estimate EY.

Missing Data / Causality

$Y \perp\!\!\!\perp A \mid Z$, $A \in \{0,1\}$. Observe Y iff $A=1$. Estimate EY.

$$p_{f,a,b}(Z, A, YA) = f(Z)\, a(Z)^{A} (1-a(Z))^{1-A}\, b(Z)^{YA} (1-b(Z))^{(1-Y)A}$$

$a(z)$: $p(A=1 \mid Z=z)$

$b(z)$: $p(Y=1 \mid Z=z)$

EY: $\chi(f,a,b) = \int b(z) f(z)\, dz$

Missing Data / Causality

$Y \perp\!\!\!\perp A \mid Z$, $A \in \{0,1\}$. Observe Y iff $A=1$. Estimate EY.

$$p_{f,a,b}(Z, A, YA) = f(Z)\, a(Z)^A (1-a(Z))^{1-A}\; b(Z)^{YA} (1-b(Z))^{(1-Y)A}$$

a: $P(A=1 \mid Z=z)$ $\qquad$ b: $P(Y=1 \mid Z=z)$

$$\chi(f,a,b) = \int b(z) f(z)\, dz$$

$t \mapsto p_{f_t, a_t, b_t}$

submodels

$$f_t = (1+tg)f$$
$$a_t = \Psi(\Psi^{-1}(a) + t\alpha)$$
$$b_t = \Psi(\Psi^{-1}(b) + t\beta)$$

Ψ: logistic function

Missing Data / Causality

$Y \perp\!\!\!\perp A \mid Z$, $A \in \{0,1\}$. Observe Y iff $A=1$. Estimate EY.

$$p_{f,a,b}(Z, A, YA) = f(Z)\, a(Z)^{A} (1-a(Z))^{1-A}\, b(Z)^{YA} (1-b(Z))^{(1-Y)A}$$

$a(Z)$: $p(A=1 \mid Z=z)$ $\quad$ $b(Z)$: $p(Y=1 \mid Z=z)$

$$\chi(f,a,b) = \int b(z) f(z)\, dz$$

$t \mapsto p_{f_t, a_t, b_t}$

submodels

$$f_t = (1 + t\gamma) f$$
$$a_t = \Psi(\Psi^{-1}(a) + t\alpha)$$
$$b_t = \Psi(\Psi^{-1}(b) + t\beta)$$

Ψ: logistic function

$$\frac{d}{dt}\Big|_{t=0} \log p_{f_t, a_t, b_t}(Z, A, AY) = \gamma(Z) + (A - a(Z))\alpha(Z) + A(Y - b(Z))\beta(Z)$$

$$\frac{d}{dt}\Big|_{t=0} \chi(f_t, a_t, b_t) = \int b(1-b)\beta\, dF + \int b\gamma\, dF.$$

Missing Data / Causality

$Y \perp A \mid Z$, $A \in \{0,1\}$. Observe Y iff $A=1$. Estimate EY.

$$p_{f,a,b}(Z, A, YA) = f(Z)\, a(Z)^A (1-a(Z))^{1-A}\, b(Z)^{YA} (1-b(Z))^{(1-Y)A}$$

$a(Z)$: $P(A=1 \mid Z=z)$ $\qquad$ $b(Z)$: $P(Y=1 \mid Z=z)$

$\chi(f,a,b) = \int b(z) f(z)\, dz$

$t \mapsto p_{f_t, a_t, b_t}$

submodels

$$f_t = (1+t\gamma) f$$
$$a_t = \Psi(\Psi^{-1}(a) + t\alpha)$$
$$b_t = \Psi(\Psi^{-1}(b) + t\beta)$$

Ψ: logistic function

$$\frac{d}{dt}\Big|_{t=0} \log p_{f_t a_t, b_t}(Z, A, AY) = \gamma(Z) + (A - a(Z))\alpha(Z) + A(Y - b(Z))\beta(Z)$$

$$\frac{d}{dt}\Big|_{t=0} \chi(f_t, a_t, b_t) = \int b(1-b)\beta\, dF + \int b\gamma\, dF.$$

LEM $\quad \hat{\chi}_{f,a,b}(Z, A, AY) = \frac{A}{a(Z)}(Y - b(Z)) + b(Z) - EY.$

Missing Data / Causality

$Y \perp\!\!\!\perp A \mid Z$, $A \in \{0,1\}$. Observe Y iff $A=1$. Estimate EY.

$$p_{f,a,b}(Z,A,YA) = f(Z)\, a(Z)^{A}(1-a(Z))^{1-A}\, b(Z)^{YA}(1-b(Z))^{(1-Y)A}$$

$a(Z)$: $P(A=1 \mid Z=z)$ $b(Z)$: $P(Y=1 \mid Z=z)$

$$\chi(f,a,b) = \int b(z) f(z)\, dz$$

$t \mapsto p_{f_t,a_t,b_t}$ submodels

$$f_t = (1+t\gamma) f$$
$$a_t = \Psi(\Psi^{-1}(a)+t\alpha)$$
$$b_t = \Psi(\Psi^{-1}(b)+t\beta)$$

Ψ: logistic function

$$\frac{d}{dt}\Big|_{t=0} \log p_{f_t, a_t, b_t}(Z,A,AY) = \underbrace{\gamma(Z) + (A-a(Z))\alpha(Z) + A(Y-b(Z))\beta(Z)}_{\text{score}(\gamma,\alpha,\beta)}$$

$$\frac{d}{dt}\Big|_{t=0} \chi(f_t,a_t,b_t) = \int b(1-b)\beta\, dF + \int b\gamma\, dF.$$

LEM $\hat{\chi}_{f,a,b}(Z,A,AY) = \frac{A}{a(Z)}(Y-b(Z)) + b(Z) - EY.$

proof Verify that $E_{f,a,b}\left[\hat{\chi}_{f,a,b}(Z,A,AY)\,\text{score}(\gamma,\alpha,\beta)\right] = \int b(1-b)\beta\, dF + \int b\gamma\, dF.$ ∎

Missing Data / Causality

$Y \perp A \mid Z$, $A \in \{0,1\}$. Observe Y iff $A=1$. Estimate EY.

$$p_{f,a,b}(Z, A, YA) = f(Z)\, a(Z)^{A} (1-a(Z))^{1-A}\, b(Z)^{YA} (1-b(Z))^{(1-Y)A}$$

$a(Z)$: $P(A=1 \mid Z=z)$ $\qquad$ $b(Z)$: $P(Y=1 \mid Z=z)$

$\chi(f,a,b) = \int b(z) f(z)\, dz$

$t \mapsto p_{f_t, a_t, b_t}$ $\qquad$ $f_t = (1+t\gamma) f$

submodels $\qquad$ $a_t = \Psi(\Psi^{-1}(a) + t\alpha)$

$b_t = \Psi(\Psi^{-1}(b) + t\beta)$

Ψ: logistic function

$$\frac{d}{dt}\Big|_{t=0} \log p_{f_t, a_t, b_t}(Z, A, AY) = \underbrace{\gamma(Z) + (A - a(Z))\alpha(Z) + A(Y - b(Z))\beta(Z)}_{\text{score}(\gamma,\alpha,\beta)}$$

$$\frac{d}{dt}\Big|_{t=0} \chi(f_t, a_t, b_t) = \int b(1-b)\beta\, dF + \int b\gamma\, dF.$$

LEM $\quad \hat{\chi}_{f,a,b}(Z, A, AY) = \frac{A}{a(Z)}(Y - b(Z)) + b(Z) - EY.$

score(γ,α,β) too
$$\begin{cases} \gamma = b - \int b\, dF \\ \alpha = 0 \\ \beta = \frac{1}{a} \end{cases}$$

<u>proof</u> Verify that $E_{f,a,b}\left[\hat{\chi}_{f,a,b}(Z, A, AY)\, \text{score}(\gamma,\alpha,\beta)\right] = \int b(1-b)\beta\, dF + \int b\gamma\, dF.$ $\square$

Missing Data/Causality

(Ray & vdV, 2019).

$Y \perp A \mid Z$, $A \in \{0,1\}$, Observe Y iff $A=1$, Estimate EY

$F \sim DP(\alpha) \perp b \sim \Psi(GP) \perp a$

$Z_1 \cdots Z_n \mid F \overset{iid}{\sim} F$

$A_1 \cdots A_n \mid Z_1 \cdots Z_n \overset{ind}{\sim} \text{logisticreg}(a(Z_i))$

$Y_1 \cdots Y_n \mid Z_1 \cdots Z_n \overset{ind}{\sim} \text{logisticreg}(b(Z_i))$

drops out of posterior

THM BvM holds if

$$\sup_{b \in B_n} |\mathbb{G}_n(b - b_0)| \xrightarrow{P} 0$$

$b \in B_n \leftarrow$ set of posterior mass $\to 1$.

likelihood expansion.

$$\frac{\int_{B_n} \prod_{i=1}^n p_{a_0, b - \frac{t}{\sqrt{n}} \frac{1}{a_0}}(X_i)\, d\Pi(b)}{\int_{B_n} \prod_{i=1}^n p_{a_0, b}(X_i)\, d\Pi(b)} \xrightarrow{P} 1$$

change of measure.

- Need b smoother than necessary
- Better prior $b = \Psi\left(GP + \frac{\lambda_n}{\hat{a}_n}\right)$, $\lambda_n \sim N(0, \sigma_n^2)$, $\sigma_n \gg \varepsilon_n^b$
- Smooth prior on F gives even stronger conditions on b.

PDEs - Data Assimilation

(Stuart, Agapiou, Nickl, Wang, ...)

Nonparametric Bayes is attractive for uncertainty quantification in inverse problems.

We observe the solution u_f of a PDE plus noise, which depends on an unknown parameter f.

We obtain the posterior of f.

Example heat equation
$$\begin{cases} \frac{\partial}{\partial t} u_f(x,t) = \frac{\partial^2}{\partial x^2} u_f(x,t) \\ u_f(\cdot,0) = f \end{cases}$$

Observe $u_f(\cdot,1)$ + noise

Example Volterra equation
$$\begin{cases} u_f' = f \\ u_f(0) = 0 \end{cases} \rightarrow u_f(x) = \int_0^x f(t)\,dt$$

Schrödinger Equation

$$\begin{cases} \frac{1}{2}\Delta u_f = u_f f & \text{on } \mathcal{O} \subset \mathbb{R}^d \\ u_f = g & \text{on } \partial\mathcal{O} \end{cases}$$

open

g given

$$y = u_f + \frac{1}{\sqrt{n}}\dot{W}$$

white noise

Schrödinger Equation

$$\begin{cases} \frac{1}{2}\Delta u_f = u_f f & \text{on } \mathcal{O} \subset \mathbb{R}^d \\ u_f = g & \text{on } \partial\mathcal{O} \end{cases}$$

open

g given

$$Y = u_f + \frac{1}{\sqrt{n}}\dot{W}$$ white noise $\longrightarrow$ $Y_i \mid f, X_i \sim N(u_f(X_i), 1)$, $i = 1, \cdots, n$

$$p_f(y \mid x) \propto \exp\left(-\frac{1}{2}(y - u_f(x))^2\right)$$

Schrödinger Equation

$$\begin{cases} \frac{1}{2}\Delta u_f = u_f f & \text{on } \mathcal{O} \subset \mathbb{R}^d \\ u_f = g & \text{on } \partial\mathcal{O} \end{cases}$$

open

g given

$$y = u_f + \frac{1}{\sqrt{n}} \dot{W}$$ white noise $\longrightarrow$ $y_i \mid f, X_i \sim N(u_f(X_i), 1), \quad i = 1, \cdots, n$

$$p_f(y|x) \propto \exp\left(-\frac{1}{2}(y - u_f(x))^2\right)$$

paths $\quad f_t = f + t\varphi$

scores $\quad \frac{\partial}{\partial t}\Big|_{t=0} \log p_{f_t}(y) = (y - u_f(x))\, \underbrace{\dot{u}_f \varphi(x)}_{= \frac{d}{dt}\big|_{t=0} u_{f_t}(x)}$

Schrödinger Equation

$$\begin{cases} \frac{1}{2}\Delta u_f = u_f f & \text{on } \mathcal{O} \subset \mathbb{R}^d \\ u_f = g & \text{on } \partial\mathcal{O} \end{cases}$$

open

g given

$$y = u_f + \frac{1}{\sqrt{n}} \dot{W}$$ white noise $\longrightarrow$ $$y_i \mid f, X_i \sim N(u_f(X_i), 1), \quad i = 1, \dots, n$$

$$p_f(y|x) \propto \exp\left(-\frac{1}{2}(y - u_f(x))^2\right)$$

paths $\quad f_t = f + t\varphi$

scores $\quad \frac{\partial}{\partial t}\Big|_{t=0} \log p_{f_t}(y) = (y - u_f(x)) \underbrace{\dot{u}_f \varphi(x)}_{= \frac{d}{dt}\big|_{t=0} u_{f_t}(x)}$

$$\frac{1}{2}\Delta u_{f_t} = u_{f_t} f_t \longrightarrow \frac{1}{2}\Delta \frac{d}{dt}\Big|_{t=0} u_{f_t} = \frac{d}{dt}\Big|_{t=0} u_{f_t} f + u_f \varphi$$

$\longrightarrow$ $v = \dot{u}_f \varphi$ satisfies $$\begin{cases} \frac{1}{2}\Delta v = v f + u_f \varphi & \text{on } \mathcal{O} \\ v = 0 & \text{on } \partial\mathcal{O}. \end{cases}$$

Schrödinger Equation

$$\begin{cases} \frac{1}{2}\Delta u_f = u_f f & \text{on } \mathcal{O} \subset \mathbb{R}^d \\ u_f = g & \text{on } \partial\mathcal{O} \end{cases}$$

open

g given

$$y = u_f + \frac{1}{\sqrt{n}} \dot{W} \quad \text{white noise} \quad \longrightarrow \quad y_i \mid f, X_i \sim N(u_f(X_i), 1), \; i = 1, \dots, n$$

$$p_f(y|x) \propto \exp\left(-\frac{1}{2}(y - u_f(x))^2\right)$$

paths $\quad f_t = f + t\varphi$

scores $\quad \frac{\partial}{\partial t}\Big|_{t=0} \log p_{f_t}(y) = (y - u_f(x)) \underbrace{\dot{u}_f \varphi(x)}_{= \frac{d}{dt}\big|_{t=0} u_{f_t}(x)}$

$$\frac{1}{2}\Delta u_{f_t} = u_{f_t} f_t \quad \longrightarrow \quad \frac{1}{2}\Delta \frac{d}{dt}\Big|_{t=0} u_{f_t} = \frac{d}{dt}\Big|_{t=0} u_{f_t} f + u_f \varphi$$

$$\longrightarrow \quad v = \dot{u}_f \varphi \text{ satisfies } \begin{cases} \frac{1}{2}\Delta v = v f + u_f \varphi & \text{on } \mathcal{O} \\ v = 0 & \text{on } \partial\mathcal{O}. \end{cases}$$

Schrödinger Equation

$$\begin{cases} \frac{1}{2}\Delta u_f = u_f f & \text{on } \mathcal{O} \subset \mathbb{R}^d \\ u_f = g & \text{on } \partial\mathcal{O} \end{cases}$$

open

g given

$$y = u_f + \frac{1}{\sqrt{n}} \dot{W} \quad \text{white noise} \longrightarrow \quad y_i \mid f, X_i \sim N(u_f(X_i), 1), \; i = 1, \dots, n$$

$$p_f(y|x) \propto \exp\left(-\tfrac{1}{2}(y - u_f(x))^2\right)$$

paths $\quad f_t = f + t\varphi$

scores $\quad \frac{\partial}{\partial t}\Big|_{t=0} \log p_{f_t}(y) = \overbrace{(y - u_f(x)) \underbrace{\dot{u}_f \varphi(x)}_{= \frac{d}{dt}|_{t=0} u_{f_t}(x)}}^{\text{score}(\varphi)}$

Influence function for $\chi(f) = \int \psi(x) f(x)\,dx$:

some fixed ψ

$$\tilde{\chi}_f(y, X) = (y - u_f(X))\, \dot{u}_f \varphi_0(X)$$

must solve

$$E_f\, \tilde{\chi}_f(y, X)\, [\text{Score}(\varphi)] = \int \psi(x)\varphi(x)\,dx, \; \forall \varphi$$

Schrödinger Equation

$$\begin{cases} \frac{1}{2}\Delta u_f = u_f f & \text{on } \mathcal{O} \subset \mathbb{R}^d \\ u_f = g & \text{on } \partial\mathcal{O} \end{cases}$$

open

g given

$$y = u_f + \frac{1}{\sqrt{n}}\dot{W} \quad \text{white noise} \longrightarrow \quad y_i \mid f, X_i \sim N(u_f(X_i), 1), \; i = 1, \cdots, n$$

$$p_f(y|x) \propto \exp\left(-\tfrac{1}{2}(y - u_f(x))^2\right)$$

paths $\quad f_t = f + t\varphi$

scores $\quad \frac{\partial}{\partial t}\Big|_{t=0} \log p_{f_t}(y) = (y - u_f(x)) \underbrace{\dot{u}_f \varphi(x)}_{= \frac{d}{dt}\big|_{t=0} u_{f_t}(x)}$

Influence function for $\chi(f) = \int \psi(x) \cancel{f(x)} dx$:

$$\tilde{\chi}_f(y, X) = (y - u_f(X))\, \dot{u}_f \varphi_0(X)$$

must solve

$$E_f\, \tilde{\chi}_f(y, X)\,[\text{Score}(\varphi)] = \int \psi(x)\varphi(x)\,dx, \quad \forall \varphi$$

i.e. $\quad \langle \dot{u}_f \varphi_0, \dot{u}_f \varphi \rangle = \langle \psi, \varphi \rangle$

$$\varphi_0 = (\dot{u}_f^T \dot{u}_f)^{-1} \psi$$

$$\tilde{\chi}_f(y, X) = (y - u_f(X))\, \dot{u}_f\, (\dot{u}_f^T \dot{u}_f)^{+} \psi$$

Schrödinger Equation

(Nickl 2018)

$$\begin{cases} \frac{1}{2}\Delta u_f = u_f f & \text{on } \mathcal{O} \subset \mathbb{R}^d \\ u_f = g & \text{on } \partial\mathcal{O} \end{cases}$$

open

$$\log f \sim \sum_{j=1}^{J_n} \sum_{\ell} f_{j,\ell} \Psi_{j\ell} \qquad f_{j,k} \sim \text{Uniform}\left[-2^{-j(\alpha+d/2)}, 2^{-j(\alpha+d/2)}\right]$$

wavelet

Estimate $\chi_\psi(f) = \int \psi(x) f(x)\, dx$, given $\psi \in C_c^q(\mathcal{O})$, $q > 2 + \frac{3d}{2}$

THM if $f_0 \in C^\beta(\mathcal{O})$ for $\beta > (2+\frac{d}{2}) \vee d$ and $2^{J_n} \sim n^{\frac{1}{2\beta+4+d}}$

then BvM holds for $\chi_\psi(f)$.

THM Infinite-dimensional BvM holds for $(\chi_\psi(f) : \|\psi\|_{C_c^q(\mathcal{O})} \leq 1)$.

$\chi_\psi(f)$, $\psi \in C_c^q(\mathcal{O})$ identifies f.

Nonparametric Smoothing

Bayesian Nonparametric Smoothing

f a function

$f \sim \Pi$

$X \mid f \sim P_f$

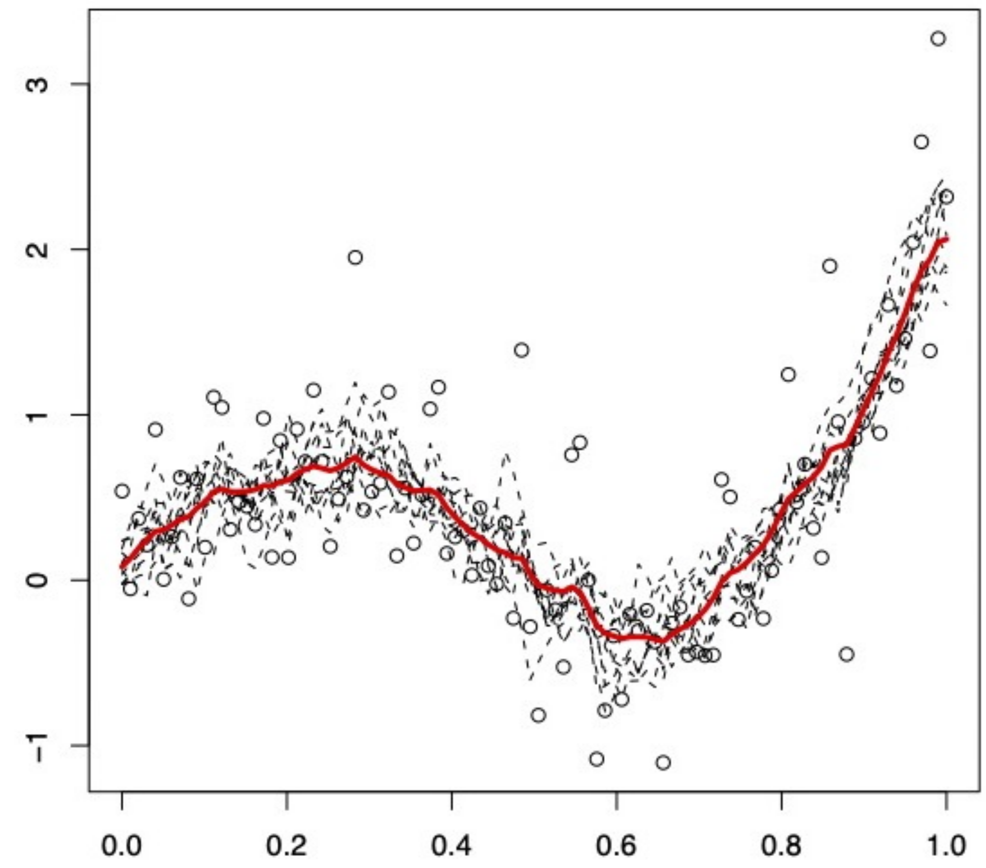
$\Pi(f \in \cdot \mid X)$

posterior mean/mode/median recovers true function
spread in $\Pi(f \in \cdot \mid X)$ should show confidence in this recovery

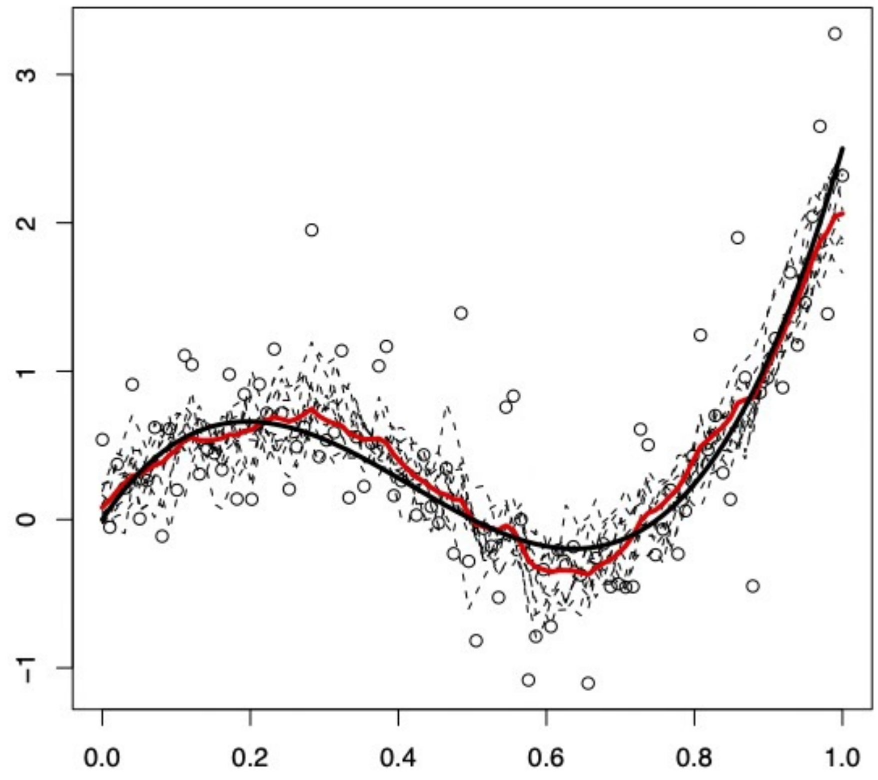
"Smoothing"

$\rightarrow$ bias-variance trade-off (prior "regularises")

$\rightarrow$ no $\sqrt{n}$-rate



3
2
1
0
−1
0.0
0.2
0.4
0.6
0.8
1.0



3
2
1
0
−1
0.0
0.2
0.4
0.6
0.8
1.0

History

Wahba, 1975

J. R. Statist. Soc. B (1983),
45, *No.* 1, *pp.* 133–150

Bayesian "Confidence Intervals" for the Cross-validated Smoothing Spline

By GRACE WAHBA

University of Wisconsin, USA

[Received August 1981. Revised August 1982]

SUMMARY

We consider the model $Y(t_i) = g(t_i) + \epsilon_i$, $i = 1, 2, \ldots, n$, where $g(t)$, $t \in [0, 1]$ is a smooth function and the $\{\epsilon_i\}$ are independent $N(0, \sigma^2)$ errors with σ^2 unknown. The cross-validated smoothing spline can be used to estimate g non-parametrically from observations on $Y(t_i)$, $i = 1, 2, \ldots, n$, and the purpose of this paper is to study confidence intervals for this estimate. Properties of smoothing splines as Bayes estimates are used to derive confidence intervals based on the posterior covariance function of the estimate. A small Monte Carlo study with the cubic smoothing spline is carried out to suggest by example to what extent the resulting 95 per cent confidence intervals can be expected to cover about 95 per cent of the true (but in practice unknown) values of $g(t_i)$, $i = 1, 2, \ldots, n$. The method was also applied to one example of a two-dimensional thin plate smoothing spline. An asymptotic theoretical argument is presented to explain why the method can be expected to work on fixed smooth functions (like those tried), which are "smoother" than the sample functions from the prior distributions on which the confidence interval theory is based.

Keywords: SPLINE SMOOTHING; CROSS-VALIDATION; CONFIDENCE INTERVALS

1. INTRODUCTION

Consider the model

$$Y(t_i) = g(t_i) + \epsilon_i, \quad i = 1, 2, \ldots, n, \quad t_i \in [0, 1], \tag{1.1}$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_n)' \sim N(0, \sigma^2 I_{n \times n})$, σ^2 is unknown and $g(\cdot)$ is a fixed but unknown function with $m - 1$ continuous derivatives and $\int_0^1 (g^{(m)}(t))^2 dt < \infty$. The smoothing spline estimate of g given $Y(t_i) = y_i$, $i = 1, 2, \ldots, n$, which we will call $g_{n,\lambda}$, is the minimizer of

$$n^{-1} \sum_{i=1}^{n} (g(t_i) - y_i)^2 + \lambda \int_0^1 (g^{(m)}(t))^2 dt$$

Works great!

Cox, 1993

The Annals of Statistics
1993, Vol. 21, No. 2, 903–923

AN ANALYSIS OF BAYESIAN INFERENCE FOR NONPARAMETRIC REGRESSION[1]

BY DENNIS D. COX

Rice University

The observation model $y_i = \beta(i/n) + \varepsilon_i$, $1 \le i \le n$, is considered, where the ε's are i.i.d. with mean zero and variance σ^2 and β is an unknown smooth function. A Gaussian prior distribution is specified by assuming β is the solution of a high order stochastic differential equation. The estimation error $\delta = \beta - \hat{\beta}$ is analyzed, where $\hat{\beta}$ is the posterior expectation of β. Asymptotic posterior and sampling distributional approximations are given for $\|\delta\|^2$ when $\|\cdot\|$ is one of a family of norms natural to the problem. It is shown that the frequentist coverage probability of a variety of $(1 - \alpha)$ posterior probability regions tends to be larger than $1 - \alpha$, but will be infinitely often less than any $\varepsilon > 0$ as $n \to \infty$ with prior probability 1. A related continuous time signal estimation problem is also studied.

1. Introduction. In this article we consider Bayesian inference for a class of nonparametric regression models. Suppose we observe

$$Y_{ni} = \beta(t_{ni}) + \varepsilon_i, \qquad 1 \le i \le n, \tag{1.1}$$

where $t_{ni} = i/n$, $\beta\colon [0, 1] \to \mathbb{R}$ is an unknown smooth function, and $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. random errors with mean 0 and known variance $\sigma^2 < \infty$. The ε_i are modeled as $N(0, \sigma^2)$. A Gaussian prior for β will now be specified. Let $m \ge 2$ and for some constants $a_0, \ldots, a_m$ with $a_m \neq 0$ let

$$L = \sum_{i=0}^{m} a_i D^i$$

Fails miserably!

(penalised least squares with splines
= posterior mode of IBM prior)