Hölder-continuous versions of posterior distributions in Bayesian Statistical Inference

EMANUELE DOLERA
Università di Modena e Reggio Emilia

Modena, June 9^{th}, 2015

(work in progress with E. MAININI)

INTRODUCTION

Basic elements of Bayesian Statistical Inference

$(\mathbb{X}, \mathrm{d}_{\mathbb{X}})$	sample space (Polish)
$(\Theta, \mathrm{d}_{\Theta})$	parameter space (Polish)
$[\Theta]$	set of all p.m.'s on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$ endowed with the topology of weak convergence of p.m.'s (Polish)
$\{m(\cdot, \theta)\}_{\theta \in \Theta}$	statistical model, i.e. a probability kernel from $(\Theta, \mathscr{B}(\Theta))$ into $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$
π	prior p.m. on $(\Theta, \mathscr{B}(\Theta))$

Bayesian paradigm and posterior distribution

There are $(\Omega, \mathscr{A}, \mathrm{P})$ and two random elements

$$\tilde{x} : \Omega \to \mathbb{X} \qquad \text{sample (observable)}$$

$$\tilde{\theta} : \Omega \to \Theta \qquad \text{random parameter}$$

such that, for all $A \in \mathscr{B}(\mathbb{X})$ and $B \in \mathscr{B}(\Theta)$,

$$\mathrm{P}[\tilde{x} \in A, \tilde{\theta} \in B] = \int_B m(A, \theta)\pi(\mathrm{d}\theta) \ . \tag{1}$$

Bayesian paradigm and posterior distribution

There are $(\Omega, \mathscr{A}, \mathrm{P})$ and two random elements

$$\tilde{x} : \Omega \to \mathbb{X} \qquad \text{sample (observable)}$$

$$\tilde{\theta} : \Omega \to \Theta \qquad \text{random parameter}$$

such that, for all $A \in \mathscr{B}(\mathbb{X})$ and $B \in \mathscr{B}(\Theta)$,

$$\mathrm{P}[\tilde{x} \in A, \tilde{\theta} \in B] = \int_B m(A, \theta)\pi(\mathrm{d}\theta) \ . \tag{1}$$

The posterior distribution $\pi(\cdot, \tilde{x})$ stands for the r.c.d. $\mathrm{P}[\tilde{\theta} \in \cdot \mid \tilde{x}]$, where $\pi(\cdot, x)$ is a probability kernel from $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$ into $(\Theta, \mathscr{B}(\Theta))$ such that, for all $A \in \mathscr{B}(\mathbb{X})$ and $B \in \mathscr{B}(\Theta)$,

$$\mathrm{P}[\tilde{x} \in A, \tilde{\theta} \in B] = \int_A \pi(B, x)\mathrm{P} \circ \tilde{x}^{-1}(\mathrm{d}x) \ . \tag{2}$$

Some common question in Bayesian statistics (I)

Question 1: Since the main tool of Bayesian statistics, i.e. the *Bayes theorem*, is not always applicable (e.g., it does not work on non-dominated models), **how to evaluate (or, at least, approximate) the posterior distribution?**

Some common question in Bayesian statistics (I)

Question 1: Since the main tool of Bayesian statistics, i.e. the *Bayes theorem*, is not always applicable (e.g., it does not work on non-dominated models), **how to evaluate (or, at least, approximate) the posterior distribution?**

See, e.g., Kolmogorov (1933), Renyi (1955), Dubins (1975), Pfanzagl (1979).

Some common question in Bayesian statistics (I)

Question 1: Since the main tool of Bayesian statistics, i.e. the *Bayes theorem*, is not always applicable (e.g., it does not work on non-dominated models), **how to evaluate (or, at least, approximate) the posterior distribution?**

See, e.g., Kolmogorov (1933), Renyi (1955), Dubins (1975), Pfanzagl (1979).

For example, given $x_0 \in \mathrm{supp}(\mathrm{P} \circ \tilde{x}^{-1})$, one wonders whether

$$\frac{\mathrm{P}[\tilde{x} \in U_\delta(x_0), \tilde{\theta} \in \cdot]}{\mathrm{P}[\tilde{x} \in U_\delta(x_0)]}$$

is a good approximation (for small δ) to $\pi(\cdot, x_0)$ or not, maybe uniformly in x_0.

Some common question in Bayesian statistics (II)

Question 2: **How to formalize the well-known concept of continuous dependence from the data for a posterior distribution?**

Some common question in Bayesian statistics (II)

Question 2: **How to formalize the well-known concept of continuous dependence from the data for a posterior distribution?**

See, e.g., Zabell (1979), Tjur (1982, 1984), Ionescu-Tulcea (1984).

Some common question in Bayesian statistics (II)

Question 2: **How to formalize the well-known concept of continuous dependence from the data for a posterior distribution?**

See, e.g., Zabell (1979), Tjur (1982, 1984), Ionescu-Tulcea (1984).

This concept would establish a point of contact with the theory of *continuous dependence on initial data*, well-known in the field of ODE's and PDE's.

Recent consideration of this question in the realm of Inverse Problems, tackled from a Bayesian point of view. See Cotter, Dashti, Robinson and Stuart (2010), Vollmer (2013).

Some common question in Bayesian statistics (III)

Question 3: **How to handle the presence of an error term $\tilde{\varepsilon}$, independent of $(\tilde{\theta}, \tilde{x})$, w.r.t. the original model? More formally, are**

$$\mathsf{E}\left[\left|\mathsf{P}[\tilde{\theta} \in B \mid \tilde{x}] - \mathsf{P}[\tilde{\theta} \in B \mid \tilde{x} + \tilde{\varepsilon}]\right|\right]$$

or

$$\mathsf{E}\left[\mathrm{d}_{[\Theta]}(\mathsf{P}[\tilde{\theta} \in \cdot \mid \tilde{x}], \mathsf{P}[\tilde{\theta} \in \cdot \mid \tilde{x} + \tilde{\varepsilon}])\right]$$

"small" if $\tilde{\varepsilon}$ is "small" in some sense?

Some common question in Bayesian statistics (III)

Question 3: **How to handle the presence of an error term $\tilde{\varepsilon}$, independent of $(\tilde{\theta}, \tilde{x})$, w.r.t. the original model? More formally, are**

$$\mathsf{E}\left[\left|\mathsf{P}[\tilde{\theta} \in B \mid \tilde{x}] - \mathsf{P}[\tilde{\theta} \in B \mid \tilde{x} + \tilde{\varepsilon}]\right|\right]$$

or

$$\mathsf{E}\left[\mathrm{d}_{[\Theta]}(\mathsf{P}[\tilde{\theta} \in \cdot \mid \tilde{x}], \mathsf{P}[\tilde{\theta} \in \cdot \mid \tilde{x} + \tilde{\varepsilon}])\right]$$

"small" if $\tilde{\varepsilon}$ is "small" in some sense?

See, e.g., Lindley and Smith (1972) for the linear model.

Some common question in Bayesian statistics (IV)

In a compact space $\mathbb{X}$, consider a sequence of partitions $\Pi_m := \{A_{m,1}, \dots, A_{m,k_m+1}\}$, for $m \in \mathbb{N}$, so that Π_{m+1} is a refinement of Π_m and $\mathscr{B}(\mathbb{X}) = \sigma\left(\bigcup_{m\geq 1} \Pi_m\right)$. Put $\varepsilon_m := \max_{1\leq i\leq k_m+1} \mathrm{diam}(A_{m,i})$ and assume that $\varepsilon_m \downarrow 0$.

Some common question in Bayesian statistics (IV)

In a compact space $\mathbb{X}$, consider a sequence of partitions $\Pi_m := \{A_{m,1}, \ldots, A_{m,k_m+1}\}$, for $m \in \mathbb{N}$, so that Π_{m+1} is a refinement of Π_m and $\mathscr{B}(\mathbb{X}) = \sigma\left(\bigcup_{m\geq 1} \Pi_m\right)$. Put $\varepsilon_m := \max_{1\leq i\leq k_m+1} \mathrm{diam}(A_{m,i})$ and assume that $\varepsilon_m \downarrow 0$.

For any $\mathfrak{p} \in \mathbb{M}$ (= set of all p.m.'s on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$), define $\mathfrak{p}_{\varepsilon_m}(\cdot) := \sum_{i=1}^{k_m+1} \mathfrak{p}(A_{m,i})\delta_{a_{m,i}}$ with $a_{m,i} \in A_{m,i}$. Then, given $\tilde{\xi}^{(n)} := (\tilde{\xi}_1, \ldots, \tilde{\xi}_n)$ exchangeable r.v.'s, define $\tilde{\xi}_{j,m} := a_{m,i}$ if $\tilde{\xi}_j \in A_{m,i}$ for $j, m \in \mathbb{N}$ and $i = 1, \ldots, k_m + 1$, and put $\xi_m^{(n)} := (\xi_{1,m}, \ldots, \xi_{n,m})$.

Some common question in Bayesian statistics (IV)

In a compact space $\mathbb{X}$, consider a sequence of partitions $\Pi_m := \{A_{m,1}, \dots, A_{m,k_m+1}\}$, for $m \in \mathbb{N}$, so that Π_{m+1} is a refinement of Π_m and $\mathscr{B}(\mathbb{X}) = \sigma\left(\bigcup_{m\geq 1} \Pi_m\right)$. Put $\varepsilon_m := \max_{1\leq i\leq k_m+1} \mathrm{diam}(A_{m,i})$ and assume that $\varepsilon_m \downarrow 0$.

For any $\mathfrak{p} \in \mathbb{M}$ (= set of all p.m.'s on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$), define $\mathfrak{p}_{\varepsilon_m}(\cdot) := \sum_{i=1}^{k_m+1} \mathfrak{p}(A_{m,i})\delta_{a_{m,i}}$ with $a_{m,i} \in A_{m,i}$. Then, given $\tilde{\xi}^{(n)} := (\tilde{\xi}_1, \dots, \tilde{\xi}_n)$ exchangeable r.v.'s, define $\tilde{\xi}_{j,m} := a_{m,i}$ if $\tilde{\xi}_j \in A_{m,i}$ for $j, m \in \mathbb{N}$ and $i = 1, \dots, k_m + 1$, and put $\xi_m^{(n)} := (\xi_{1,m}, \dots, \xi_{n,m})$.

<u>Question 4</u>: **How close is** $\mathrm{P}[\tilde{\mathfrak{p}}_{\varepsilon_m} \in \cdot \mid \xi_m^{(n)}]$ **to the original posterior?**

Some common question in Bayesian statistics (IV)

In a compact space $\mathbb{X}$, consider a sequence of partitions $\Pi_m := \{A_{m,1}, \ldots, A_{m,k_m+1}\}$, for $m \in \mathbb{N}$, so that Π_{m+1} is a refinement of Π_m and $\mathscr{B}(\mathbb{X}) = \sigma\left(\bigcup_{m\geq 1} \Pi_m\right)$. Put $\varepsilon_m := \max_{1\leq i\leq k_m+1} \operatorname{diam}(A_{m,i})$ and assume that $\varepsilon_m \downarrow 0$.

For any $\mathfrak{p} \in \mathbb{M}$ (= set of all p.m.'s on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$), define $\mathfrak{p}_{\varepsilon_m}(\cdot) := \sum_{i=1}^{k_m+1} \mathfrak{p}(A_{m,i})\delta_{a_{m,i}}$ with $a_{m,i} \in A_{m,i}$. Then, given $\tilde{\xi}^{(n)} := (\tilde{\xi}_1, \ldots, \tilde{\xi}_n)$ exchangeable r.v.'s, define $\tilde{\xi}_{j,m} := a_{m,i}$ if $\tilde{\xi}_j \in A_{m,i}$ for $j, m \in \mathbb{N}$ and $i = 1, \ldots, k_m + 1$, and put $\xi_m^{(n)} := (\xi_{1,m}, \ldots, \xi_{n,m})$.

Question 4: **How close is** $\mathrm{P}[\tilde{\mathfrak{p}}_{\varepsilon_m} \in \cdot \mid \xi_m^{(n)}]$ **to the original posterior?**

See, e.g., Regazzini and Sazonov (2000), and Arjas (1996).

Some common question in Bayesian statistics (V)

Let $\{\tilde{\xi}_n\}_{n\geq 1}$ on $(\Omega, \mathscr{A}, \mathrm{P})$ be a sequence of *exchangeable* r.v.'s, and put $\pi_n(\cdot, (\tilde{\xi}_1, \dots, \tilde{\xi}_n))$ the posterior distribution based on a sample of n observations.

Some common question in Bayesian statistics (V)

Let $\{\tilde{\xi}_n\}_{n\geq 1}$ on $(\Omega, \mathscr{A}, \mathrm{P})$ be a sequence of *exchangeable* r.v.'s, and put $\pi_n(\cdot, (\tilde{\xi}_1, \dots, \tilde{\xi}_n))$ the posterior distribution based on a sample of n observations.

Question 5 : given another sequence $\{\tilde{\eta}_n\}_{n\geq 1}$ on $(\Omega, \mathscr{A}, \mathrm{P})$, i.i.d. with common distribution $m(\cdot, \theta_0)$, **is it true that** $\pi_n(\cdot, (\tilde{\eta}_1, \dots, \tilde{\eta}_n)) \Rightarrow \delta_{\theta_0}$, P-**a.s., as** n **goes to infinity? (Bayesian consistency)**

Some common question in Bayesian statistics (V)

Let $\{\tilde{\xi}_n\}_{n\geq 1}$ on $(\Omega, \mathscr{A}, \mathrm{P})$ be a sequence of *exchangeable* r.v.'s, and put $\pi_n(\cdot, (\tilde{\xi}_1, \ldots, \tilde{\xi}_n))$ the posterior distribution based on a sample of n observations.

Question 5 : given another sequence $\{\tilde{\eta}_n\}_{n\geq 1}$ on $(\Omega, \mathscr{A}, \mathrm{P})$, i.i.d. with common distribution $m(\cdot, \theta_0)$, **is it true that** $\pi_n(\cdot, (\tilde{\eta}_1, \ldots, \tilde{\eta}_n)) \Rightarrow \delta_{\theta_0}$, P-**a.s., as** n **goes to infinity? (Bayesian consistency)**

See, e.g., Diaconis and Freedman (1986), Barron, Schervish and Wasserman (1999), Ghosal, Ghosh and van der Vaart (2000).

Our main problem

The previous problems can be traced back to a common "purely mathematical" question:
Find sufficient conditions on $\{m(\cdot,\theta)\}_{\theta\in\Theta}$ and π for the existence of a **specific version** $\pi^*(\cdot,x)$ of the posterior for which

$$\boxed{\mathrm{d}_{[\Theta]}(\pi^*(\cdot,x_1),\pi^*(\cdot,x_2)) \leq C_\alpha \mathrm{d}_{\mathbb{X}}(x_1,x_2)^\alpha \qquad (3)}$$

holds for all $x_1,x_2\in\mathbb{X}$, with suitable $\alpha\in(0,1]$ and $C_\alpha\geq 0$.

Our main problem

The previous problems can be traced back to a common “purely mathematical” question:
Find sufficient conditions on $\{m(\cdot,\theta)\}_{\theta\in\Theta}$ and π for the existence of a **specific version** $\pi^*(\cdot,x)$ of the posterior for which

$$\boxed{\mathrm{d}_{[\Theta]}(\pi^*(\cdot,x_1),\pi^*(\cdot,x_2)) \leq C_\alpha \mathrm{d}_{\mathbb{X}}(x_1,x_2)^\alpha \qquad (3)}$$

holds for all $x_1,x_2\in\mathbb{X}$, with suitable $\alpha\in(0,1]$ and $C_\alpha\geq 0$.

Condition (3) amounts to proving that $\pi(\cdot,x)$ admits a Hölder-continuous versions, as a function from $(\mathbb{X},\mathrm{d}_{\mathbb{X}})$ into $([\Theta],\mathrm{d}_{[\Theta]})$.

Our main problem

The previous problems can be traced back to a common "purely mathematical" question:
Find sufficient conditions on $\{m(\cdot,\theta)\}_{\theta\in\Theta}$ and π for the existence of a **specific version** $\pi^*(\cdot,x)$ of the posterior for which

$$\boxed{\mathrm{d}_{[\Theta]}(\pi^*(\cdot,x_1),\pi^*(\cdot,x_2)) \leq C_\alpha \mathrm{d}_{\mathbb{X}}(x_1,x_2)^\alpha \qquad (3)}$$

holds for all $x_1, x_2 \in \mathbb{X}$, with suitable $\alpha \in (0,1]$ and $C_\alpha \geq 0$.

Condition (3) amounts to proving that $\pi(\cdot,x)$ admits a Hölder-continuous versions, as a function from $(\mathbb{X},\mathrm{d}_{\mathbb{X}})$ into $([\Theta],\mathrm{d}_{[\Theta]})$.

The choice of $\mathrm{d}_{[\Theta]}$ plays a crucial role in the problem.

Weak metrics

$$\mathrm{d}^{(P)}_{[\Theta]}(\mu_1,\mu_2) := \inf\{\varepsilon > 0 \mid \mu_1(B) \le \mu_2(B^{\varepsilon}) + \varepsilon,\ \forall\, B \in \mathscr{B}(\Theta)\}$$

$$\mathrm{d}^{(FM)}_{[\Theta]}(\mu_1,\mu_2) := \sup_{\substack{h:\Theta\to\mathbb{R} \\ \|h\|_{BL}\le 1}} \left| \int_{\Theta} h(\theta)\mu_1(\mathrm{d}\theta) - \int_{\Theta} h(\theta)\mu_2(\mathrm{d}\theta) \right|$$

$$\mathrm{d}^{(W,\beta)}_{[\Theta]}(\mu_1,\mu_2) := \inf_{\nu\in\mathcal{F}(\mu_1,\mu_2)} \left(\int_{\Theta^2} [\mathrm{d}_{\Theta}(\theta_1,\theta_2)]^{\beta}\, \nu(\mathrm{d}\theta_1 \mathrm{d}\theta_2) \right)^{1/\beta}$$

Weak metrics

$$\mathrm{d}^{(P)}_{[\Theta]}(\mu_1,\mu_2) := \inf\{\varepsilon > 0 \mid \mu_1(B) \leq \mu_2(B^\varepsilon) + \varepsilon,\ \forall\, B \in \mathscr{B}(\Theta)\}$$

$$\mathrm{d}^{(FM)}_{[\Theta]}(\mu_1,\mu_2) := \sup_{\substack{h:\Theta\to\mathbb{R} \\ \|h\|_{BL}\leq 1}} \left| \int_\Theta h(\theta)\mu_1(\mathrm{d}\theta) - \int_\Theta h(\theta)\mu_2(\mathrm{d}\theta) \right|$$

$$\mathrm{d}^{(W,\beta)}_{[\Theta]}(\mu_1,\mu_2) := \inf_{\nu\in\mathcal{F}(\mu_1,\mu_2)} \left(\int_{\Theta^2} [\mathrm{d}_\Theta(\theta_1,\theta_2)]^\beta\, \nu(\mathrm{d}\theta_1 \mathrm{d}\theta_2) \right)^{1/\beta}$$

Notation: $B^\varepsilon := \{\theta \in \Theta \mid \mathrm{d}_\Theta(\theta,\theta_0) < \varepsilon$ for some $\theta_0 \in B\}$; $\|h\|_{BL} := \|h\|_\infty + \|h\|_L$, $\|h\|_\infty := \sup_{\theta\in\Theta} |h(\theta)|$ and $\|h\|_L := \sup_{\theta_1\neq\theta_2}[|h(\theta_1) - h(\theta_2)|/\mathrm{d}_\Theta(\theta_1,\theta_2)]$; $\mathcal{F}(\mu_1,\mu_2)$ stands for the Fréchet class with fixed marginals μ_1 and μ_2.

Weak metrics

$$\mathrm{d}^{(P)}_{[\Theta]}(\mu_1,\mu_2) := \inf\{\varepsilon>0 \mid \mu_1(B) \leq \mu_2(B^\varepsilon)+\varepsilon,\ \forall\, B\in\mathscr{B}(\Theta)\}$$

$$\mathrm{d}^{(FM)}_{[\Theta]}(\mu_1,\mu_2) := \sup_{\substack{h:\Theta\to\mathbb{R}\\ \|h\|_{BL}\leq 1}} \left| \int_\Theta h(\theta)\mu_1(\mathrm{d}\theta) - \int_\Theta h(\theta)\mu_2(\mathrm{d}\theta)\right|$$

$$\mathrm{d}^{(W,\beta)}_{[\Theta]}(\mu_1,\mu_2) := \inf_{\nu\in\mathcal{F}(\mu_1,\mu_2)} \left(\int_{\Theta^2}[\mathrm{d}_\Theta(\theta_1,\theta_2)]^\beta\,\nu(\mathrm{d}\theta_1\mathrm{d}\theta_2)\right)^{1/\beta}$$

Notation: $B^\varepsilon := \{\theta\in\Theta \mid \mathrm{d}_\Theta(\theta,\theta_0)<\varepsilon \text{ for some } \theta_0\in B\}$; $\|h\|_{BL} := \|h\|_\infty + \|h\|_L$, $\|h\|_\infty := \sup_{\theta\in\Theta}|h(\theta)|$ and $\|h\|_L := \sup_{\theta_1\neq\theta_2}[|h(\theta_1)-h(\theta_2)|/\mathrm{d}_\Theta(\theta_1,\theta_2)]$; $\mathcal{F}(\mu_1,\mu_2)$ stands for the Fréchet class with fixed marginals μ_1 and μ_2.

Warning: $\mathrm{d}^{(W,\beta)}_{[\Theta]}$ is defined for $\beta\geq 1$, only when $\int_\Theta[\mathrm{d}_\Theta(\theta,\theta_0)]^\beta\mu(\mathrm{d}\theta)<+\infty$ for some $\theta_0\in\Theta$, for $i=1,2$.

Strong "metrics"

$$\mathrm{d}^{(TV)}_{[\Theta]}(\mu_1,\mu_2) := \sup_{B\in\Theta} |\mu_1(B)-\mu_2(B)| = \frac{1}{2}\int_\Theta |f_1(\theta)-f_2(\theta)|\lambda(\mathrm{d}\theta)$$

$$\mathrm{d}^{(H)}_{[\Theta]}(\mu_1,\mu_2) := \left(\int_\Theta [\sqrt{f_1(\theta)}-\sqrt{f_2(\theta)}]^2\lambda(\mathrm{d}\theta)\right)^{1/2}$$

$$\mathrm{d}^{(KL)}_{[\Theta]}(\mu_1,\mu_2) := \int_\Theta \log\left(\frac{f_1(\theta)}{f_2(\theta)}\right) f_1(\theta)\lambda(\mathrm{d}\theta)$$

provided that $\mu_i \ll \lambda$, with $f_i := \frac{\mathrm{d}\mu_i}{\mathrm{d}\lambda}$, $i = 1,2$.

Strong "metrics"

$$
\begin{aligned}
\mathrm{d}^{(TV)}_{[\Theta]}(\mu_1,\mu_2) &:= \sup_{B\in\Theta} |\mu_1(B)-\mu_2(B)| = \frac{1}{2}\int_\Theta |f_1(\theta)-f_2(\theta)|\lambda(\mathrm{d}\theta) \\
\mathrm{d}^{(H)}_{[\Theta]}(\mu_1,\mu_2) &:= \left(\int_\Theta [\sqrt{f_1(\theta)}-\sqrt{f_2(\theta)}]^2\lambda(\mathrm{d}\theta)\right)^{1/2} \\
\mathrm{d}^{(KL)}_{[\Theta]}(\mu_1,\mu_2) &:= \int_\Theta \log\left(\frac{f_1(\theta)}{f_2(\theta)}\right) f_1(\theta)\lambda(\mathrm{d}\theta)
\end{aligned}
$$

provided that $\mu_i \ll \lambda$, with $f_i := \frac{\mathrm{d}\mu_i}{\mathrm{d}\lambda}$, $i=1,2$.

Strong metrics are (in general) stronger than weak ones.

Two simple example with positive answer

A parametric example: $\mathbb{X} = \Theta = \mathbb{R}$; $m(\cdot, \theta) = \mathcal{N}[\theta, s^2](\cdot)$; $\pi(\cdot) = \mathcal{N}[\mu, \sigma^2](\cdot)$; $\pi(\cdot \mid x) = \mathcal{N}[\frac{\sigma^2 x + s^2 \mu}{\sigma^2 + s^2}, \frac{\sigma^2 s^2}{\sigma^2 + s^2}](\cdot)$. Then,

$$\mathrm{d}^{(TV)}_{[\Theta]}(\pi(\cdot \mid x_1), \pi(\cdot \mid x_2)) = C(\mu, \sigma^2, s^2)|x_1 - x_2| \ .$$

Two simple example with positive answer

A parametric example: $\mathbb{X} = \Theta = \mathbb{R}$; $m(\cdot, \theta) = \mathcal{N}[\theta, s^2](\cdot)$; $\pi(\cdot) = \mathcal{N}[\mu, \sigma^2](\cdot)$; $\pi(\cdot \mid x) = \mathcal{N}[\frac{\sigma^2 x + s^2 \mu}{\sigma^2 + s^2}, \frac{\sigma^2 s^2}{\sigma^2 + s^2}](\cdot)$. Then,

$$\mathrm{d}^{(TV)}_{[\Theta]}(\pi(\cdot \mid x_1), \pi(\cdot \mid x_2)) = C(\mu, \sigma^2, s^2)|x_1 - x_2| \ .$$

A nonparametric example: $\mathbb{X} = \mathbb{X}_1^n$ with $\mathbb{X}_1$ any metric space; $\Theta = \mathbb{M}$, the set of all p.m.'s on $(\mathbb{X}_1, \mathscr{B}(\mathbb{X}_1))$; $m(A_1 \times \cdots \times A_n, \theta) = \prod_{i=1}^n \mathfrak{p}(A_i)$; $\pi(\cdot) = \mathcal{D}[\alpha](\cdot)$; $\pi(\cdot \mid (x_1, \ldots, x_n)) = \mathcal{D}[\alpha + \sum_{i=1}^n \delta_{x_i}](\cdot)$. Then,

$$\mathrm{d}^{(W,\beta)}_{\mathbb{P}}(\pi(\cdot \mid \mathbf{x}), \pi(\cdot \mid \mathbf{y})) \le C \, \mathrm{d}^{(W,\beta)}_{\mathbb{M}} \left(\frac{1}{n} \sum_{i=1}^n \delta_{x_i}, \frac{1}{n} \sum_{i=1}^n \delta_{y_i} \right) .$$

An example of the complexity of the problem

A vast class of *nonparametric priors* can be characterized as *normalized random measure with independent increments*, i.e. $\tilde{\mathfrak{p}} = \tilde{\mu}/\tilde{\mu}(\mathbb{X})$, where $\tilde{\mu}$ is a *completely random measure* (c.r.m).

An example of the complexity of the problem

A vast class of *nonparametric priors* can be characterized as *normalized random measure with independent increments*, i.e. $\tilde{\mathfrak{p}} = \tilde{\mu}/\tilde{\mu}(\mathbb{X})$, where $\tilde{\mu}$ is a *completely random measure* (c.r.m).

A c.r.m. without fixed jumps is characterized by the Lévy-Khintchine representation

$$\mathsf{E}\left[e^{-\int_{\mathbb{X}} f(x)\tilde{\mu}(\mathrm{d}x)}\right] = \exp\left\{-\int_0^{+\infty}\int_{\mathbb{X}}[1-e^{-sf(x)}]\nu(\mathrm{d}s,\mathrm{d}x)\right\}$$

$\int_0^{+\infty}(1 \wedge s)\nu(\mathrm{d}s, A) < +\infty$. Typically, $\nu(\mathrm{d}s,\mathrm{d}x) = \rho_x(\mathrm{d}s)\alpha(\mathrm{d}x)$.

An example of the complexity of the problem

A vast class of *nonparametric priors* can be characterized as *normalized random measure with independent increments*, i.e. $\tilde{\mathfrak{p}} = \tilde{\mu}/\tilde{\mu}(\mathbb{X})$, where $\tilde{\mu}$ is a *completely random measure* (c.r.m).

A c.r.m. without fixed jumps is characterized by the Lévy-Khintchine representation

$$\mathsf{E}\left[e^{-\int_{\mathbb{X}} f(x)\tilde{\mu}(\mathrm{d}x)}\right] = \exp\left\{-\int_0^{+\infty}\int_{\mathbb{X}}[1-e^{-sf(x)}]\nu(\mathrm{d}s,\mathrm{d}x)\right\}$$

$\int_0^{+\infty}(1\wedge s)\nu(\mathrm{d}s, A) < +\infty$. Typically, $\nu(\mathrm{d}s,\mathrm{d}x) = \rho_x(\mathrm{d}s)\alpha(\mathrm{d}x)$.
The posterior is given by normalizing $\tilde{\mu} \mid (\tilde{\xi}_1,\dots,\tilde{\xi}_n), \tilde{U}_n =^{(d)} \tilde{\mu}_{\tilde{U}_n} + \sum_{i=1}^k \tilde{J}_i^{(\tilde{U}_n)}\delta_{\tilde{\xi}_i^*}$. See Regazzini, Lijoi and Prünster (2003), James, Lijoi and Prünster (2005, 2009).

Why a general theorem?

There are, in general, recurrent difficulties in checking the validity of (3):

a) apart from the Bayes theorem, there is no general formula yielding the posterior explicitly
b) the posterior is a non-linear function of model, prior and data
c) even when the posterior is known, the evaluation of the distance $\mathrm{d}_{[\Theta]}$ is often impracticable.

Why a general theorem?

There are, in general, recurrent difficulties in checking the validity of (3):

a) apart from the Bayes theorem, there is no general formula yielding the posterior explicitly
b) the posterior is a non-linear function of model, prior and data
c) even when the posterior is known, the evaluation of the distance $\mathrm{d}_{[\Theta]}$ is often impracticable.

Thus, it is not convenient to evaluate the posterior explicitly and verify the Hölder-continuity directly. It is preferable to find sufficient conditions on the prior or, better, to some finite-dimensional objects which characterize the prior.

THE DOMINATED CASE

Bayes theorem

Basic assumption of dominated model: $m(\cdot, \theta) \ll \lambda$ for all $\theta \in \Theta$, for a suitable σ-finite measure λ on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$.

Bayes theorem

Basic assumption of dominated model: $m(\cdot,\theta) \ll \lambda$ for all $\theta \in \Theta$, for a suitable σ-finite measure λ on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$.

Theorem (Bayes)

If $(\theta, x) \mapsto f(x \mid \theta) := \frac{m(\mathrm{d}x,\theta)}{\lambda(\mathrm{d}x)}$ *is* $\mathscr{B}(\Theta) \otimes \mathscr{B}(\mathbb{X})/\mathscr{B}(\mathbb{R})$*, then there exists a set* $A_0 \in \mathscr{B}(\mathbb{X})$*, with* $P \circ \tilde{x}^{-1}(A_0) = 1$*, such that, for all* $x \in A_0$*, one has* $\rho(x) := \int_\Theta f(x \mid \theta)\pi(\mathrm{d}\theta) > 0$ *and*

$$\pi(B, x) = \frac{\int_B f(x \mid \theta)\pi(\mathrm{d}\theta)}{\rho(x)} \qquad (\forall\, B \in \mathscr{B}(\Theta)) \,. \tag{4}$$

Bayes theorem

Basic assumption of dominated model: $m(\cdot,\theta) \ll \lambda$ for all $\theta \in \Theta$, for a suitable σ-finite measure λ on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$.

Theorem (Bayes)

If $(\theta, x) \mapsto f(x \mid \theta) := \frac{m(\mathrm{d}x,\theta)}{\lambda(\mathrm{d}x)}$ *is* $\mathscr{B}(\Theta) \otimes \mathscr{B}(\mathbb{X})/\mathscr{B}(\mathbb{R})$*, then there exists a set* $A_0 \in \mathscr{B}(\mathbb{X})$*, with* $P \circ \tilde{x}^{-1}(A_0) = 1$*, such that, for all* $x \in A_0$*, one has* $\rho(x) := \int_\Theta f(x \mid \theta)\pi(\mathrm{d}\theta) > 0$ *and*

$$\pi(B, x) = \frac{\int_B f(x \mid \theta)\pi(\mathrm{d}\theta)}{\rho(x)} \qquad (\forall\, B \in \mathscr{B}(\Theta)) \,. \tag{4}$$

If $\mathscr{B}(\mathbb{X})$ is countable generated (e.g., by the separability of $\mathbb{X}$), the above measurability of $(\theta, x) \mapsto f(x \mid \theta)$ follows.

An elementary case: Θ is a finite set

Theorem

If $\Theta = \{\theta_1, \dots, \theta_k\}$ and $\pi(\{\theta_i\}) > 0$ for all $i = 1, \dots, k$, then (3), with $\mathrm{d}_{[\Theta]} = \mathrm{d}_{[\Theta]}^{(TV)}$, is equivalent to requiring that $x \mapsto \frac{f(x \mid \theta_i)}{f(x \mid \theta_j)}$ are Hölder-continuous of exponent $\alpha \in (0, 1]$ for all $i, j = 1, \dots, k$.

An elementary case: Θ is a finite set

Theorem

If $\Theta = \{\theta_1, \dots, \theta_k\}$ *and* $\pi(\{\theta_i\}) > 0$ *for all* $i = 1, \dots, k$*, then (3), with* $\mathrm{d}_{[\Theta]} = \mathrm{d}_{[\Theta]}^{(TV)}$*, is equivalent to requiring that* $x \mapsto \frac{f(x \mid \theta_i)}{f(x \mid \theta_j)}$ *are Hölder-continuous of exponent* $\alpha \in (0, 1]$ *for all* $i, j = 1, \dots, k$*.*

A strong point is the absence of extra hypotheses on $(\mathbb{X}, \mathrm{d}_{\mathbb{X}})$.

An elementary case: Θ is a finite set

Theorem

If $\Theta = \{\theta_1, \dots, \theta_k\}$ and $\pi(\{\theta_i\}) > 0$ for all $i = 1, \dots, k$, then (3), with $\mathrm{d}_{[\Theta]} = \mathrm{d}_{[\Theta]}^{(TV)}$, is equivalent to requiring that $x \mapsto \frac{f(x \mid \theta_i)}{f(x \mid \theta_j)}$ are Hölder-continuous of exponent $\alpha \in (0, 1]$ for all $i, j = 1, \dots, k$.

- A strong point is the absence of extra hypotheses on $(\mathbb{X}, \mathrm{d}_{\mathbb{X}})$.
- Another merit of this simple setting is to highlight the role of the regularity of $f(x \mid \theta)$ w.r.t. the x-variable.

Functional analysis approach

Theorem

If $\mathbb{X} = \overline{\mathbb{U}}$, $\mathbb{U} \subseteq \mathbb{R}^d$ *is an open subset with Lipschitz boundary, and*

$$\int_\Theta \left\| \frac{f(x \mid \theta)}{\rho(x)} \right\|_{\mathrm{W}_x^{1,p}(\mathbb{U})} \pi(\mathrm{d}\theta) < +\infty \tag{5}$$

holds for some $p > d$*, then (3) is in force with* $\mathrm{d}_{[\Theta]} = \mathrm{d}_{[\Theta]}^{(TV)}$*.*

Functional analysis approach

Theorem

If $\mathbb{X} = \overline{\mathbb{U}}$, $\mathbb{U} \subseteq \mathbb{R}^d$ *is an open subset with Lipschitz boundary, and*

$$\int_{\Theta} \left\| \frac{f(x \mid \theta)}{\rho(x)} \right\|_{\mathrm{W}_x^{1,p}(\mathbb{U})} \pi(\mathrm{d}\theta) < +\infty \tag{5}$$

holds for some $p > d$*, then (3) is in force with* $\mathrm{d}_{[\Theta]} = \mathrm{d}_{[\Theta]}^{(TV)}$*.*

A strong point is the absence of extra hypotheses on $(\Theta, \mathrm{d}_\Theta)$.

Functional analysis approach

Theorem

If $\mathbb{X} = \overline{\mathbb{U}}$, $\mathbb{U} \subseteq \mathbb{R}^d$ is an open subset with Lipschitz boundary, and

$$\int_\Theta \left\| \frac{f(x \mid \theta)}{\rho(x)} \right\|_{\mathrm{W}_x^{1,p}(\mathbb{U})} \pi(\mathrm{d}\theta) < +\infty \tag{5}$$

holds for some $p > d$, then (3) is in force with $\mathrm{d}_{[\Theta]} = \mathrm{d}_{[\Theta]}^{(TV)}$.

- A strong point is the absence of extra hypotheses on $(\Theta, \mathrm{d}_\Theta)$.
- The proof exploits the fact that $x \mapsto h(\theta)\pi(\mathrm{d}\theta, x)$ is in $\mathrm{W}^{1,p}(\mathbb{U})$ uniformly w.r.t. $h \in \mathrm{L}^\infty(\Theta)$ with $\|h\|_\infty \leq 1$. The conclusion follows from the Morrey embedding theorem for $\mathrm{W}^{1,p}(\mathbb{U})$ with $p > d$.

Functional analysis approach

If the evaluation of the norm in (6) is difficult, we provide a simpler condition, at the expense of further assumptions on $\mathbb{X}$ and ρ.

Theorem

Suppose that $\mathbb{X} = \overline{\mathbb{U}}$, where $\mathbb{U} \subseteq \mathbb{R}^d$ is a bounded open subset with Lipschitz boundary. If $\rho \in \mathrm{W}^{1,p}(\mathbb{U})$, $\rho(x) \geq R > 0$ a.e. in $\mathbb{U}$ and

$$\int_{\Theta} \|f(x \mid \theta)\|_{\mathrm{W}_x^{1,p}(\mathbb{U})} \, \pi(\mathrm{d}\theta) < +\infty \tag{6}$$

hold for some $p > d$, then (6) follows with the same p.

Geometric approach: Amari-Rao calculus

This approach exploits the fact that a dominated statistical model with densities $\{g(\cdot \mid \tau)\}_{\tau \in \mathbb{T}}$ can be viewed as a Riemannian manifold with Riemannian metric given by the Fisher information matrix. Here, this theory can be applied to the (unusual!) model $\left\{ \frac{f(x \mid \theta)}{\rho(x)} \right\}_{x \in \mathbb{X}}$, with $\mathbb{X}$ some d-dimensional set, which is required to have *finite Fisher information*, i.e.

$$\begin{aligned} & I_{i,j}(x) \\ := \; & \int_\Theta \left[\frac{\partial}{\partial x_i} \log\left(\frac{f(x \mid \theta)}{\rho(x)} \right) \right] \cdot \left[\frac{\partial}{\partial x_j} \log\left(\frac{f(x \mid \theta)}{\rho(x)} \right) \right] \frac{f(x \mid \theta)}{\rho(x)} q(\mathrm{d}\theta) \\ < \; & +\infty \end{aligned}$$

for all $i, j = 1, \ldots, d$.

Geometric approach: Amari-Rao calculus

Another important element is the *Amari formula*, which links the geodesic distance induced by this Riemannian structure to the Hellinger distance: $\mathrm{d}^{(H)}_{[\Theta]}(f_1, f_2) = \varphi(\mathrm{d}^{(Geo)}_{[\Theta]}(f_1, f_2))$ for a suitable Lipschitz continuous φ.

Geometric approach: Amari-Rao calculus

Another important element is the *Amari formula*, which links the geodesic distance induced by this Riemannian structure to the Hellinger distance: $\mathrm{d}^{(H)}_{[\Theta]}(f_1, f_2) = \varphi(\mathrm{d}^{(Geo)}_{[\Theta]}(f_1, f_2))$ for a suitable Lipschitz continuous φ.

Theorem

If $\mathbb{X} \subseteq \mathbb{R}^d$ and $I_{i,j}(x)$ is uniformly bounded, then (3) is in force with $\mathrm{d}_{[\Theta]} = \mathrm{d}^{(H)}_{[\Theta]}$.

Geometric approach: Otto calculus

This approach is based on the Otto-Villani theory of the Wasserstein space $([\Theta]_2, \mathrm{d}_{[\Theta]}^{(W,2)})$, where

$$[\Theta]_2 := \left\{ \mu \in [\Theta] \mid \int_{\Theta} [\mathrm{d}_\Theta(\theta, \theta_0)]^2 \mu(\mathrm{d}\theta) < +\infty \right\} .$$

This space can be (formally) viewed as an infinite-dimensional Riemannian manifold, provided that Θ has itself a structure of smooth and complete Riemannian manifold.

Geometric approach: Otto calculus

This approach is based on the Otto-Villani theory of the Wasserstein space $([\Theta]_2, \mathrm{d}_{[\Theta]}^{(W,2)})$, where

$$[\Theta]_2 := \left\{ \mu \in [\Theta] \mid \int_\Theta [\mathrm{d}_\Theta(\theta, \theta_0)]^2 \mu(\mathrm{d}\theta) < +\infty \right\} .$$

This space can be (formally) viewed as an infinite-dimensional Riemannian manifold, provided that Θ has itself a structure of smooth and complete Riemannian manifold.

A key element of the theory is the Benamou-Brenier formula:

$$\mathrm{d}_{[\Theta]}^{(W,2)}(\mu^0, \mu^1) = \inf \left\{ \int_0^1 \|\mathbf{v}_t\|_{\mu_t} \mathrm{d}t \right\} \tag{7}$$

where the inf is taken on the solutions of $\frac{\partial \mu_t}{\partial t} + \nabla \bullet (\mathbf{v}_t \mu_t) = 0$ with $\mu^0 = \mu_0$ and $\mu^1 = \mu_1$.

Geometric approach: Otto calculus

In (7), $\|\cdot\|_{\mu_t}$ stands for the Riemannian L^2-norm on $\mathrm{T}_{\mu_t}([\Theta]_2)$ and the integral yields the length of $\{\mu_t\}_{t\in[0,1]}$. Hence, $\mathrm{d}_{[\Theta]}^{(W,2)}$ can be viewed as a geodesic distance.

Geometric approach: Otto calculus

In (7), $\|\cdot\|_{\mu_t}$ stands for the Riemannian L^2-norm on $\mathrm{T}_{\mu_t}([\Theta]_2)$ and the integral yields the length of $\{\mu_t\}_{t\in[0,1]}$. Hence, $\mathrm{d}_{[\Theta]}^{(W,2)}$ can be viewed as a geodesic distance.

Assume that $\mathbb{X}$ is a smooth, complete Riemannian manifold and let $\alpha : [0,1] \to \mathbb{X}$ be the geodesic connecting x_1 with x_2.

Geometric approach: Otto calculus

In (7), $\|\cdot\|_{\mu_t}$ stands for the Riemannian L^2-norm on $\mathrm{T}_{\mu_t}([\Theta]_2)$ and the integral yields the length of $\{\mu_t\}_{t\in[0,1]}$. Hence, $\mathrm{d}^{(W,2)}_{[\Theta]}$ can be viewed as a geodesic distance.

Assume that $\mathbb{X}$ is a smooth, complete Riemannian manifold and let $\alpha : [0,1] \to \mathbb{X}$ be the geodesic connecting x_1 with x_2.

If q coincides with vol and $\int_\Theta [\mathrm{d}_\Theta(\theta,\theta_0)]^2 f(x \mid \theta) q(\mathrm{d}\theta) < +\infty$ for all $x \in \mathbb{X}$, then $\gamma(t) := \frac{f(\alpha(t) \mid \theta)}{\rho(\alpha(t))} q(\mathrm{d}\theta)$ is a curve in $[\Theta]_2$ connecting $\pi(\cdot \mid x_1)$ with $\pi(\cdot \mid x_2)$. Therefore, from (7),

$$\mathrm{d}^{(W,2)}_{[\Theta]}(\pi(\cdot \mid x_1), \pi(\cdot \mid x_2)) \leq \text{length}[\gamma] \leq \int_0^1 \|\bar{\mathbf{v}}_t\|_{\gamma(t)} \mathrm{d}t$$

where $\bar{\mathbf{v}}_t = \nabla_\theta u(t,\theta)$ solves $\frac{\partial \gamma(t)}{\partial t} + \nabla \bullet (\gamma(t) \nabla_\theta u(t,\theta)) = 0$ for all $t \in [0,1]$, with $u(t,\theta) = 0$ on $\theta \in \partial\Theta$.

Geometric approach: Otto calculus

We get, in the end, the (singular) elliptic problem: find $u = u(t,\theta) \in \mathrm{H}^1_\theta(\Theta, \pi(\mathrm{d}\theta \mid x))$ such that

$$\begin{aligned} & \frac{\rho(\alpha(t))\alpha'(t)\nabla_x f(\alpha(t) \mid \theta) - f(\alpha(t) \mid \theta)\alpha'(t)\nabla_x \rho(\alpha(t))}{\rho(\alpha(t))} \\ = \quad & -\nabla_\theta \bullet (f(\alpha(t) \mid \theta)\nabla_\theta u(t,\theta)) \end{aligned} \tag{8}$$

in Θ, with $u(t,\theta) = 0$ on $\partial\Theta$.

Geometric approach: Otto calculus

We get, in the end, the (singular) elliptic problem: find $u = u(t,\theta) \in \mathrm{H}^1_\theta(\Theta, \pi(\mathrm{d}\theta \mid x))$ such that

$$\begin{aligned} & \frac{\rho(\alpha(t))\alpha'(t)\nabla_x f(\alpha(t) \mid \theta) - f(\alpha(t) \mid \theta)\alpha'(t)\nabla_x \rho(\alpha(t))}{\rho(\alpha(t))} \\ = \quad & -\nabla_\theta \bullet (f(\alpha(t) \mid \theta)\nabla_\theta u(t,\theta)) \end{aligned} \tag{8}$$

in Θ, with $u(t,\theta) = 0$ on $\partial\Theta$.

Theorem

Let Θ be a smooth, complete and compact Riemannian manifold with positive Ricci curvature, and let π coincide with the volume form. Suppose that $f(x \mid \theta) \geq f_0 > 0$ for all $x \in \mathbb{X}$ and $\theta in \mathbb{X}$, and that $f(x \mid \theta) \in \mathrm{H}^1_x(\mathbb{X})$. Then, (3) is in force with $\mathrm{d}_{[\Theta]} = \mathrm{d}^{(W,2)}_{[\Theta]}$ and with $\mathrm{d}_{\mathbb{X}}$ the geodesic distance on $\mathbb{X}$.

THE NONPARAMETRIC CASE

The nonparametric setting consists in choosing $\Theta = \mathbb{M}$, the space of all the p.m.'s on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$ and $m(A, \theta) = \mathfrak{p}(A)$, with the usual change of notation $\theta \leftrightarrow \mathfrak{p}$. The distance $\mathrm{d}_{\mathbb{M}}$ is chosen among the weak metrics.

The nonparametric setting consists in choosing $\Theta = \mathbb{M}$, the space of all the p.m.'s on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$ and $m(A, \theta) = \mathfrak{p}(A)$, with the usual change of notation $\theta \leftrightarrow \mathfrak{p}$. The distance $\mathrm{d}_{\mathbb{M}}$ is chosen among the weak metrics.

A central point in Bayesian Nonparametrics is the characterization of π through simpler objects, typically of finite-dimensional type. Here, we deem convenient to introduce the *laws of the observations* $\{\mu_n\}_{n \geq 1}$, referred to *an auxiliary exchangeable sequence* $\{\tilde{\xi}_n\}_{n \geq 1}$, to *define* π through

$$\mu_n(A_1 \times \cdots \times A_n) = \int_{\mathbb{M}} \left[\prod_{i=1}^{n} \mathfrak{p}(A_i)\right] \pi(\mathrm{d}\mathfrak{p}) \tag{9}$$

The nonparametric setting consists in choosing $\Theta = \mathbb{M}$, the space of all the p.m.'s on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$ and $m(A, \theta) = \mathfrak{p}(A)$, with the usual change of notation $\theta \leftrightarrow \mathfrak{p}$. The distance $\mathrm{d}_{\mathbb{M}}$ is chosen among the weak metrics.

A central point in Bayesian Nonparametrics is the characterization of π through simpler objects, typically of finite-dimensional type. Here, we deem convenient to introduce the *laws of the observations* $\{\mu_n\}_{n\geq 1}$, referred to *an auxiliary exchangeable sequence* $\{\tilde{\xi}_n\}_{n\geq 1}$, to *define* π through

$$\mu_n(A_1 \times \cdots \times A_n) = \int_{\mathbb{M}} \left[\prod_{i=1}^{n} \mathfrak{p}(A_i)\right] \pi(\mathrm{d}\mathfrak{p}) \qquad (9)$$

The general philosophy will be to find *sufficient conditions* on the sequence $\{\mu_n\}_{n\geq 1}$ to obtain (3).

Even if it may seem unusual to define π through the sequence $\{\mu_n\}_{n\geq 1}$, it is almost always simpler to obtain this sequence from the actual definition of π than to evaluate directly the posterior.

Even if it may seem unusual to define π through the sequence $\{\mu_n\}_{n\geq 1}$, it is almost always simpler to obtain this sequence from the actual definition of π than to evaluate directly the posterior.

Noteworthy examples of nonparametric (class of) priors are:

a) normalized random measures with independent increments
b) neutral-to-the-right
c) priors for cumulative hazards or hazard rates
d) exchangeable partition probability functions
e) Poisson-Kingman models and Gibbs-type priors
f) species sampling models
g) mixture models
h) Polya trees
i) hierarchical models.

A structure lemma

We state the following lemma in an abstract setting. Then, we shall use it with $\mathbb{X} \subseteq \mathbb{R}^d$ or $\mathbb{X}$ a d-dimensional Riemannian manifold with positive curvature, with reference measure $\lambda = \mathscr{L}^d$ or $\lambda = \text{vol}$, respectively.

A structure lemma

We state the following lemma in an abstract setting. Then, we shall use it with $\mathbb{X} \subseteq \mathbb{R}^d$ or $\mathbb{X}$ a d-dimensional Riemannian manifold with positive curvature, with reference measure $\lambda = \mathscr{L}^d$ or $\lambda = \mathrm{vol}$, respectively.

Lemma (Assumptions)

Consider the sequence $\{\mu_n\}_{n\geq 1}$ satisfying the symmetry and the compatibility assumptions. Suppose $\mathrm{d}\mu_1 = \rho_1 \mathrm{d}\lambda$ for some probability density function ρ_1 on $(\mathbb{X}, \mathscr{B}(\mathbb{X}), \lambda)$ and

$$\mu_2(A_1 \times A_2) = u \int_{A_1 \times A_2} \rho_2(x_1, x_2) \lambda^{(2)}(\mathrm{d}x_1 \mathrm{d}x_2) + (1-u) \int_{A_1 \cap A_2} \rho_1(x) \lambda(\mathrm{d}x) \tag{10}$$

holds for all $A_1, A_2 \in \mathscr{B}(\mathbb{X})$, some $u \in [0,1]$ and some symmetric probability density function ρ_2 on $(\mathbb{X}^2, \mathscr{B}(\mathbb{X})^2, \lambda^{(2)})$.

A structure lemma

Lemma (Thesis)

Then,

$$\mu_n(A_1 \times \cdots \times A_n) = \sum_{k=1}^{n} \sum_{(l_1,\ldots,l_k)\in(*)_{n,k}} u_n(l_1,\ldots,l_k) \int_{A_{l_1}\times\cdots\times A_{l_k}} \rho_k(x_1,\ldots,x_k)\lambda^{(k)}(\mathrm{d}x_1\ldots\mathrm{d}x_k)$$

for every $n \geq 3$, $A_1,\ldots,A_n \in \mathscr{B}(\mathbb{X})$, some constants $\{u_n(l_1,\ldots,l_k)\}_{n\geq 1,(l_1,\ldots,l_k)\in(*)_{n,k}}$, and some symmetric densities ρ_k on $(\mathbb{X}^k, \mathscr{B}(\mathbb{X})^k, \lambda^{(k)})$, with $A_{l_j} := \cap_{i\in l_j} A_i$. Moreover, $\{\rho_k\}_{k\geq 1}$ generates a new set $\{\nu_n\}_{n\geq 1}$ of p.m.'s, by $\mathrm{d}\nu_n := \rho_n \mathrm{d}\lambda^{(n)}$, which are symmetric and compatible. Finally, $\{u_n(l_1,\ldots,l_k)\}_{n\geq 1,(l_1,\ldots,l_k)\in(*)_{n,k}}$ are EPPF's in the sense of Kingman-Pitman.

Main results

Theorem

Consider $\{\mu_n\}_{n\geq 1}$ as in the structure Lemma, and suppose that

$$\rho_k(x_1,\ldots,x_k) = \int_{\mathbb{T}} \left[\prod_{i=1}^{k} g(x_i \mid \tau)\right] q(\mathrm{d}\tau)\ . \tag{11}$$

If $\mathbb{X} = \overline{\mathbb{U}}$, $\mathbb{U} \subseteq \mathbb{R}^d$ is an open subset with Lipschitz boundary, and the posterior distribution $q(\cdot \mid x)$ relative to the dominated model meets

$$\sup_{x_0 \in \mathbb{U}} \sup_{\substack{h:\mathbb{T}\to\mathbb{R} \\ \|h\|_{BL}\leq 1}} \left\| \int_{\mathbb{T}} h(\tau) q(\mathrm{d}\tau \mid x) \right\|_{\mathrm{W}_x^{1,p}(\mathbb{U})} < +\infty \tag{12}$$

for some $p > d$, then (3) holds with $\mathrm{d}_{\mathbb{M}} = \mathrm{d}_{\mathbb{M}}^{(FM)}$ and $\mathrm{d}_{\mathbb{P}} = \mathrm{d}_{\mathbb{P}}^{(FM)}$.

For the check of the hypotheses in the previous theorem, one can go on step by step. The hypotheses coming from the structure lemma can be checked by a direct inspection of the μ_n's. The hypothesis on the ρ_k's can be managed by using a theorem by Berti, Pratelli and Rigo (2004): a necessary and sufficient condition is that the sequence of predictive densities $\tilde{f}_n$ is P-a.s. uniformly integrable on compact subsets of $\mathbb{X}$. A simpler, but stronger, condition is $\sup_n \mathrm{E}\left[\int_{\mathbb{K}} \tilde{f}_n^p \mathrm{d}\lambda\right] < +\infty$ for some $p > 1$. Finally, the check of (12) can be carried on with the methods developed in the previous section, after noting that Lipschitz-continuity (w.r.t. to the Fortet-Mourier metric on $[\mathbb{T}]$) entails (12).

For the check of the hypotheses in the previous theorem, one can go on step by step. The hypotheses coming from the structure lemma can be checked by a direct inspection of the μ_n's. The hypothesis on the ρ_k's can be managed by using a theorem by Berti, Pratelli and Rigo (2004): a necessary and sufficient condition is that the sequence of predictive densities $\tilde{f}_n$ is P-a.s. uniformly integrable on compact subsets of $\mathbb{X}$. A simpler, but stronger, condition is $\sup_n \mathrm{E}\left[\int_{\mathbb{K}} \tilde{f}_n^p \mathrm{d}\lambda\right] < +\infty$ for some $p > 1$. Finally, the check of (12) can be carried on with the methods developed in the previous section, after noting that Lipschitz-continuity (w.r.t. to the Fortet-Mourier metric on $[\mathbb{T}]$) entails (12).

To illustrate the power of the method, one can see that the check of our conditions is almost immediate for priors of NRMII and Gibss type.

Main results

If the check of (12) is difficult, we provide a simpler condition, at the expense of further assumptions on $\mathbb{X}$, μ_1 and μ_2.

Theorem

Consider $\{\mu_n\}_{n\geq 1}$ as in the structure Lemma. Suppose that $\mathbb{X} = \overline{\mathbb{U}}$, where $\mathbb{U} \subseteq \mathbb{R}^d$ is a bounded open subset with Lipschitz boundary. If $\rho_1 \in \mathrm{W}^{1,p}(\mathbb{U})$ for some $p > d$, $\rho_1(x) \geq R > 0$ a.e. in $\mathbb{U}$ and the restriction of μ_2 to some open neighborhood $I(\delta)$ of the diagonal is a.c. w.r.t. $\lambda^{(2)}$ with local density $\rho_2\big|_{I(\delta)} \in \mathrm{H}^m(I(\delta))$ for some sufficiently large m, then (3) holds with $\mathrm{d}_{\mathbb{M}} = \mathrm{d}_{\mathbb{M}}^{(FM)}$ and $\mathrm{d}_{\mathbb{P}} = \mathrm{d}_{\mathbb{P}}^{(FM)}$.