Chilean Journal of Statistics
Vol. 1, No. 2, September 2010, 51–68

Special Issue "In Memory of Pilar Loreto Iglesias Zuazola"
Research Paper

Nonparametric Bayesian robustness

Fabrizio Ruggeri

CNR IMATI, Milano, Italy

(Received: 02 April 2010 · Accepted in final form: 06 August 2010)

Abstract

A new, nonparametric, approach to Bayesian robustness is presented. Whereas many studies in Bayesian robustness have dealt with a parametric sampling distribution, considering classes of prior distributions on the parameters, here we assume that the sampling distribution comes from a Dirichlet process with a parameter $\eta = \beta\alpha$, with $\beta > 0$ and α being a probability measure, specified with uncertainty.

Keywords: Bayesian robustness · Concentration function · Dirichlet process.

Mathematics Subject Classification: Primary 62F15 · Secondary 62F35.

1. Introduction

Parametric Bayesian robustness usually deals with uncertainty in the prior by modelling a class Γ of probability measures on the parameter space. Extensive reviews of the literature on the subject are presented, e.g., in Berger (1994) and Rios-Insua and Ruggeri (2000), and we refer to them for a discussion of the robust Bayesian viewpoint.

In this paper, we are going to present a new, nonparametric, approach to Bayesian robustness. In Bayesian nonparametrics, probability measures on the space $(\mathcal{X}, \mathcal{A})$ are chosen by a stochastic process, such as the Dirichlet process introduced by Ferguson (1973). We suppose that it is impossible to specify exactly the parameter η of a Dirichlet process P. As an example, considered in Ruggeri (1994a), let $\mathcal{X}$ be the real plane $\Re^2$ and η be proportional to a probability measure α for which only the marginals distributions can be specified, i.e., α is in a Fréchet class, without any knowledge about the joint distribution. Therefore, η being in a class Λ, we have a family of Dirichlet processes and, as in the robust parametric approach, we study the behaviour of some quantity of interest and compute its range as η varies in Λ. Some definitions and properties are provided in Section 2. Here, we are interested in the distance between Dirichlet processes (Section 3); in the probability of some subspace of the space of all probabilities on $(\mathcal{X}, \mathcal{A})$ (Section 4); in the probability that set probabilities take certain values (Section 5) and, finally, in some Bayes estimators (Section 6), e.g., of a random distribution function from P. The concentration function (Cifarelli and Regazzini, 1987; Fortini and Ruggeri, 1994, 1995) will be very useful in proving some results, especially in Section 7, where prior and posterior distances are compared. A short discussion concludes the paper in Section 8.

*Corresponding author. Fabrizio Ruggeri. CNR IMATI, Via Bassini 15, I-20133, Milano, Italy.
Email: fabrizio@mi.imati.cnr.it

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

2. Definitions

We first present some definitions and results about the Dirichlet processes, the Fréchet class in Bayesian nonparametrics and the concentration function.

Definition 2.1 Let $\mathcal{X}$ be a set and $\mathcal{A}$ be a σ-field of subsets of $\mathcal{X}$. Let η be a finite, nonnull, nonnegative, finite additive measure on $(\mathcal{X}, \mathcal{A})$. A random probability measure P on $(\mathcal{X}, \mathcal{A})$ is a Dirichlet process on $(\mathcal{X}, \mathcal{A})$ with parameter η, if for every $k = 1, 2, \ldots$, and measurable partition $B_1, \ldots, B_k$ of $\mathcal{X}$, the joint distribution of the random probabilities $(\mathrm{P}(B_1), \ldots, \mathrm{P}(B_k))$ is Dirichlet with parameters $(\eta(B_1), \ldots, \eta(B_k))$.

Ferguson (1973) proved the following result.

Theorem 2.2 Let P be a Dirichlet process on $(\Re^2, \mathcal{A})$ with parameter $\eta = \beta\alpha$ and $\alpha(\Re^2) = 1$. Let $Z_1, \ldots, Z_n$, be a sample of size n from P. Then, the conditional distribution of P, given $Z_1, \ldots, Z_n$, is a Dirichlet process with parameter $\eta^*(x, y) = \beta^* \alpha^*(x, y)$, where $\beta^* = \beta + n$ and

$$\alpha^*(x, y) = \frac{\beta\alpha(x, y) + \sum_{i=1}^{n} \delta_{Z_i}(x, y)}{\beta^*},$$

and $\alpha^*(\Re^2) = 1$.

Consider the space $\Re^2$ and the Borel σ-field $\mathcal{A}$ on it. Let F and G be two distributions on $\Re$ and let $\mathcal{B}$ denote the Borel σ-field on the real line.

Definition 2.3 The class of all bivariate distribution functions with given marginals F and G is called the Fréchet class $\Gamma(F, G)$.

Cifarelli and Regazzini (1987) introduced the concentration function of Π with respect to (w.r.t.) Π_0, where Π and Π_0 are two probability measures on the same measurable space $(\Theta, \mathcal{F})$. According to the Radon-Nikodym theorem, there is a partition $\{N, N^C\} \subset \mathcal{F}$ of Θ and a nonnegative function h on N^C such that, $\forall E \in \mathcal{F}$,

$$\Pi(E) = \int_{E \cap N^C} h(\theta)\Pi_0(d\theta) + \Pi_s(E \cap N),$$

$\Pi_0(N) = 0$, $\Pi_s(N) = \Pi_s(\Theta)$, where

$$\Pi_a(\cdot) = \int_{\cdot \cap N^C} h(\theta)\Pi_0(d\theta)$$

and Π_s denote the absolutely continuous and the singular part of Π w.r.t. Π_0, respectively. Take $h(\theta) = \infty$ over all N and

$$H(y) = \Pi_0\left(\{\theta \in \Theta \colon h(\theta) \le y\}\right), c_x = \inf\{y \in \Re \colon H(y) \ge x\}.$$

Finally, let $L_x = \{\theta \in \Theta \colon h(\theta) \le c_x\}$ and $L_x^- = \{\theta \in \Theta \colon h(\theta) < c_x\}$.

Definition 2.4 The function φ: $[0,1] \to [0,1]$ is said to be the concentration function (c.f.) of Π w.r.t. Π_0 if $\varphi(x) = \Pi(L_x^-) + c_x\{x - H(c_x^-)\}$, for $x \in (0,1)$, $\varphi(0) = 0$ and $\varphi(1) = \Pi_a(\Theta)$.

It is worth mentioning that the c.f. between finite measures, with equal total mass, could be defined in a similar way.

The following theorem, proved in Cifarelli and Regazzini (1987), states that $\varphi(x)$ substantially coincides with the minimum value of Π on the measurable subsets of Θ with Π_0-measure not smaller than x.

Theorem 2.5 If $A \in \mathcal{F}$ and $\Pi_0(A) = x$, then $\varphi(x) \le \Pi_a(A)$. Moreover, if $x \in [0,1]$ is adherent to the range of H, then B_x exists such that $\Pi_0(B_x) = x$ and

$$\varphi(x) = \Pi_a(B_x) = \min\{\Pi(A)\colon A \in \mathcal{F} \text{ and } \Pi_0(A) \ge x\}. \tag{1}$$

If Π_0 is nonatomic, then Equation (1) holds for any $x \in [0,1]$.

Theorem 2.5 is relevant in applying the c.f. to robust Bayesian analysis. In fact, given any $x \in [0,1]$, the probability, under Π, of all the subsets A with Π_0-measure x, is such that $\varphi(x) \le \Pi(A) \le 1 - \varphi(1-x)$.

A partial ordering of probability measures is possible by using the coefficients of divergence considered in Ali and Silvey (1966) and Csiszár (1967), and defined by

$$\rho(\Pi, g) = \int_{[0,\infty)} g(t) dH_\Pi(t) + \Pi_s(\Theta) \lim_{t\to\infty} \frac{g(t)}{t},$$

where g: $[0,\infty) \to \Re$ is continuous and convex, whereas H_Π and Π_s are defined as before, for any $\Pi \in \mathcal{P}$, w.r.t. a fixed $\Pi_0 \in \mathcal{P}$.

Gini's concentration ratio (Gini, 1914) given by

$$C(\Pi) = 2\int_0^1 \{x - \varphi(x)\}\, dx$$

and the index

$$G(\Pi) = \sup_{x\in[0,1]} \{x - \varphi(x)\}$$

proposed by Pietra (1915), which equals twice the variational distance

$$\sup_{A\in\mathcal{F}} |\Pi(A) - \Pi_0(A)|,$$

are obtained as particular cases of $\rho(\Pi, g)$, taking, respectively,

$$g(t) = \frac{1}{2} \int_{\Re} |t-u| dH_\Pi(u) \; + \; \frac{1}{2}\, \Pi_S(\Theta) \quad \text{and} \quad g(t) = |t-1|.$$

In addition, when Π is absolutely continuous w.r.t. Π_0, the Kullback-Leibler index and the χ^2 divergence are obtained from $\rho(\Pi, g)$ setting $g(t) = t\log(t)$ and $g(t) = (t-1)^2$.

3. Distance Between Dirichlet Process

As the parameter η of a Dirichlet process varies in a class, a family of Dirichlet processes is obtained and a sensitivity analysis could be performed about how "far" these processes are from one another. The idea of measuring the distance between stochastic processes is, of course, not new in literature. We simply mention Vajda (1990), where the Renyi distance and the Kullback-Leibler divergence, both based on Hellinger integrals, were used to evaluate the distances between distributions of regular Markov processes. Here, we consider, as a distance between two Dirichlet processes, both the maximum Hellinger distance and the maximum Kullback-Leibler divergence between the distributions, under the two Dirichlet processes, of the probability of any subset $A \in \mathcal{A}$. The interest in such distributions follows, quite naturally, from the definition of a Dirichlet process in terms of the distribution of random probabilities of partitions of $\mathcal{A}$; see Definition 2.1.

3.1 Hellinger distance

Definition 3.1 Given the Dirichlet processes P and Q on $(\mathcal{X}, \mathcal{A})$, their distance is given by

$$d_{\mathrm{H}}(P, Q) = \sup_{A \in \mathcal{A}} d(P(A), Q(A)),$$

where $d(X, Y)$ denotes the Hellinger distance between two random variables whose distributions have densities p and q w.r.t. a dominating measure μ, i.e.,

$$d(X, Y) = \left\{ \int (\sqrt{p} - \sqrt{q})^2 \, d\mu \right\}^{1/2}.$$

Here, d_{H} is actually a distance. In fact, it is symmetric and nonnegative. In addition, $d_{\mathrm{H}}(P, Q) = 0$ if and only if the processes have the same parameters a.e. From Definition 2.1, it follows that, for any $A \in \mathcal{A}$, $P(A)$ and $Q(A)$ are Beta distributed, with parameters $(\eta_1(A), \beta - \eta_1(A))$ and $(\eta_2(A), \beta - \eta_2(A))$. The condition $d_{\mathrm{H}}(P, Q) = 0$ implies $d(P(A), Q(A)) = 0$, for all $A \in \mathcal{A}$, so that the two Beta distributions coincide and $\eta_1(A) = \eta_2(A)$, for all $A \in \mathcal{A}$ and the two Dirichlet processes have the same parameter η. Viceversa, two Dirichlet processes with the same parameter are such that $d_{\mathrm{H}}(P, Q) = 0$. Finally, given the processes P, Q and R, the triangle inequality is proved by

$$d_{\mathrm{H}}(P, Q) \leq \sup_{A \in \mathcal{A}} \{d(P(A), R(A)) + d(R(A), Q(A))\} \leq d_H(P, R) + d_H(R, Q).$$

The concise notation $f_{\eta,\beta}$ denotes, for an arbitrary, but fixed $A \in \mathcal{A}$, the density of a Beta distributed random variable with parameters $(\eta, \beta - \eta)$, where $\eta = \eta(A)$ and $\beta = \eta(\mathcal{X})$). When possible, the subscript β will be omitted, whereas a similar notation will be used later for a Dirichlet distribution. Given two measures η_1 and η_2, such that $\eta_1(\mathcal{X}) = \beta_1$ and $\eta_2(\mathcal{X}) = \beta_2$, respectively, it can be easily shown that

$$d(P(A), Q(A)) = \sqrt{2} \, (1 - Y(\eta_1(A), \eta_2(A)))^{1/2},$$

where

$$Y(\eta_1,\eta_2) = \int_0^1 \sqrt{f_{\eta_1,\beta_1}(y) f_{\eta_2,\beta_2}(y)}dy$$
$$= \frac{\Gamma((\eta_1+\eta_2)/2)\Gamma((\beta_1+\beta_2)/2-(\eta_1+\eta_2)/2)}{\sqrt{\Gamma(\eta_1)\Gamma(\beta_1-\eta_1)\Gamma(\eta_2)\Gamma(\beta_2-\eta_2)}} \times \frac{\sqrt{\Gamma(\beta_1)\Gamma(\beta_2)}}{\Gamma((\beta_1+\beta_2)/2)}.$$

The distance $d_{\rm H}(P,Q)$ can be expressed by means of the concentration function described in Section 2.

Without loss of generality, suppose that $\beta_2 \leq \beta_1$. We consider the c.f. φ_1 of η_2 w.r.t. η_1, but we should notice that would have obtained the same result by considering the c.f. φ_2 of η_1 w.r.t. η_2. It can be shown that it is false, in general, that $\varphi_1(t) = \varphi_2(t), \forall t \in [0,1]$, but here it is possible to prove that $Y(x,\varphi_1(x)) = Y(\varphi_1(x), \beta_2 - \varphi_2(\beta_1 - \varphi_1(x)))$. Furthermore, it can be proved that $Y(x, \beta-\varphi(\beta-x)) = Y(\beta-x, \varphi(\beta-x))$, so that it is sufficient to consider $Y(x,\varphi(x))$.

Theorem 3.2 We have that

$$d_{\rm H}(P,Q) = \sqrt{2}\left(1 - \inf_{0\leq x\leq\beta_1} Y(x,\varphi(x))\right)^{1/2},$$

where φ is the c.f. of η_2 w.r.t. η_1.

Proof Consider $\mathcal{A}_x = \{A \in \mathcal{A}\colon \eta_1(A) = x\}$, for $0 \leq x \leq \beta_1$. We want to find $A \in \mathcal{A}_x$, which minimises $Y(x,\eta_2(A))$. Take

$$Z_x(\eta_2(A)) = Y(x,\eta_2(A)) \frac{\sqrt{\Gamma(\eta_1(A))\,\Gamma(\beta_1-\eta_1(A))}\,\Gamma((\beta_1+\beta_2)/2)}{\sqrt{\Gamma(\beta_1)\,\Gamma(\beta_2)}}.$$

It follows that

$$\frac{\partial Z_x}{\partial \eta_2(A)} = \frac{Z_x(\eta_2(A))}{2} \times T_x(\eta_2(A)),$$

where $\Psi(x) = \partial \log(\Gamma(x))/\partial x$ and

$$T_x(\eta_2(A)) = \Psi\left(\frac{\eta_1(A)+\eta_2(A)}{2}\right) - \Psi\left(\frac{\beta_1+\beta_2}{2} - \frac{\eta_1(A)+\eta_2(A)}{2}\right)$$
$$-\Psi(\eta_2(A)) + \Psi(\beta_2 - \eta_2(A)).$$

It can be shown that

$$\lim_{\eta_2(A)\to 0} \frac{\partial Z_x}{\partial \eta_2(A)} = +\infty$$

and

$$\lim_{\eta_2(A)\to \beta_2} \frac{\partial Z_x}{\partial \eta_2(A)} = -\infty.$$

From Abramowitz and Stegun (1972, pp. 259-260), it follows that

$$\Psi(z) = \int_0^\infty \left[\frac{e^{-t}}{t} - \frac{e^{-zt}}{1-e^{-t}} \right] dt \quad \text{and} \quad \Psi^{'}(z) = \int_0^\infty \frac{te^{-zt}}{1-e^{-t}} 1 \, dt.$$

Therefore

$$\Psi\left(\frac{\eta_1(A)+\eta_2(A)}{2}\right) - \Psi(\eta_2(A)) = \int_0^\infty \frac{e^{-\eta_2(A)t}\{1 - e^{-\frac{\eta_1(A)-\eta_2(A)}{2}}\}}{1-e^{-t}} dt$$

and

$$\Psi(\beta_2 - \eta_2(A)) - \Psi\left(\frac{\beta_1+\beta_2}{2} - \frac{\eta_1(A)+\eta_2(A)}{2}\right) =$$

$$\int_0^\infty \frac{e^{-(\beta_2-\eta_2(A))t}\left\{e^{-((\beta_1-\eta_1(A))/2-(\beta_2-\eta_2(A))/2)t} - 1\right\}}{1-e^{-t}} dt.$$

Looking for the above quantities within brackets to be positive for all t, it follows that $\partial Z_x/\partial\eta_2(A)$ is positive for $\eta_2(A) < \beta_2 - \beta_1 + \eta_1(A)$ and negative for $\eta_2(A) > \eta_1(A)$. When $\beta_2 - \beta_1 + \eta_1(A) \leq \eta_2(A) \leq \eta_1(A)$, then it can be shown that

$$\begin{aligned}\frac{\partial T_x}{\partial\eta_2(A)} &= \left\{\frac{1}{2}\Psi^{'}\left(\frac{\eta_1(A)+\eta_2(A)}{2}\right) - \Psi^{'}(\eta_2(A))\right\} \\ &+ \left\{\frac{1}{2}\Psi^{'}\left(\frac{\beta_1+\beta_2}{2} - \frac{\eta_1(A)+\eta_2(A)}{2}\right) - \Psi^{'}(\beta_2 - \eta_2(A))\right\}\end{aligned}$$

is negative. Then, $\partial Z_x/\partial\eta_2(A) = 0$ at a unique point, which is the unique maximum for $Z_x(\eta_2(A))$ (e.g., $\eta_2(A) = x$ if $\beta_1 = \beta_2$) and, since Z_x is a continuous function, it reaches a minimum at either $\inf_{A\in\mathcal{A}_x} \eta_2(A)$ or $\sup_{A\in\mathcal{A}_x} \eta_2(A)$, i.e., as discussed in Section 2, at either $\varphi(x)$ or $\beta_2 - \varphi(\beta_1 - x)$. The same argument can be repeated for any $x \in [0, \beta_1]$. ■

Therefore, the distance between two processes is found by computing the c.f. $\varphi(x)$ and then minimising the function $Y(x, \varphi(x))$, as in the following examples.

Example 3.3 Let P and Q be Dirichlet processes on $(\Re, \mathcal{B})$, whose parameters have densities η_1 and η_2 such that $\eta_2(x) = \gamma\eta_1(x)$ on a subset A with measure 0.5 under η_1 and $\eta_2(x) = (2-\gamma)\eta_1(x)$ on A^C, with $0 \leq \gamma \leq 1$ and $\eta_1(\Re) = 1 = \eta_2(\Re)$. Therefore, the c.f. of η_2 w.r.t. η_1 is given by

$$\varphi(x) = \begin{cases} \gamma x, & 0 \leq x \leq 0.5; \\ (2-\gamma)x + \gamma - 1, & 0.5 \leq x < 1. \end{cases}$$

By proving that the expected value $\mathcal{E}_x L = \Psi(x) - \Psi(\beta)$ and assuming for simplicity $\beta = 1$, then, from the well-known result $\Psi(x) - \Psi(1-x) = -\pi\cot\pi x$, and some, not trivial, algebraic manipulations, it follows that $\partial Y(x, \varphi(x))/\partial x$ is negative (positive) for $x \leq 0.5$ ($x > 0.5$), so that $Y(x, \varphi(x))$ achieves its minimum value at $x = 0.5$. Finally, it can be shown that $\lim_{\gamma\to 0} Y(x, \varphi(x)) = 0$, so that $d_{\text{H}}(P, Q) = \sqrt{2}$.

Distances are computed for different values of γ and shown in Table 1.

Table 1. Hellinger distance.

γ	0.0001	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Distance	1.40	0.98	0.79	0.65	0.53	0.42	0.33	0.24	0.16	0.08	0.0

Example 3.4 Let P and Q be Dirichlet processes whose parameters η_1 and η_2 have distributions $\mathcal{E}(1)$ and $\mathcal{G}(2,1)$, respectively. It can be proved that the c.f. of η_2 w.r.t. η_1 is given by $\varphi(x) = x + (1-x)\log(1-x)$, and numerically shown that $d_{\mathrm{H}}(P,Q) \approx \sqrt{2}$. The result is not surprising since the c.f. tells us that there exist subsets whose probability is very small under η_1 and very large under η_2.

The most interesting application is about classes Π of Dirichlet processes Q determined by their parameters η being in a class Γ. Let η_0 be the parameter of a baseline Dirichlet process P_0. Let Λ the class of all c.f.'s of the parameters $\eta \in \Gamma$ w.r.t. η_0. The following theorem simplifies the search of the distance between the process P_0 and the other processes Q. Its proof is omitted because it is very similar to that of Theorem 3.2.

Theorem 3.5 For any $x \in [0,1]$, define $\hat{\varphi}(x) = \inf_{\varphi\in\Lambda} \varphi(x)$. Then, it follows that

$$\sup_{Q\in\Pi} d_{\mathrm{H}}(P_0,Q) = \sqrt{2}\left(1 - \inf_{0\le x\le\beta_1} Y(x,\hat{\varphi}(x))\right)^{1/2}.$$

The class Γ could be one of those described by Fortini and Ruggeri (1994). That is, particular cases of ε-contaminations, total variation neighbourhood, density ratio and density bounded. There, such neighbourhoods of a given probability measure η_0 were defined by considering all the probability measures whose c.f. w.r.t. η_0 was not below a given monotone nondecreasing, continuous, convex function $g(x)$ with $g(0) = 0$ and $g(1) \le 1$. It is even possible to consider measures with finite mass β, providing that $g(\beta) \le \beta$.

Example 3.6 Let η be in the ε-contaminated class $\Gamma_\varepsilon = \{\eta\colon \eta = (1-\varepsilon)\eta_0 + \varepsilon\gamma, \gamma \in \mathcal{M}\}$, where $\mathcal{M}$ is the class of all measures with mass β. Such a class is a neighbourhood of η_0 described by Fortini and Ruggeri (1994) by taking $g(x) = (1-\varepsilon)x$. Here,

$$Y(x,(1-\varepsilon)x) = \frac{\Gamma((1-\varepsilon/2)x)\Gamma(\beta-(1-\varepsilon/2)x)}{\sqrt{\Gamma(x)\Gamma(\beta-x)\Gamma((1-\varepsilon)x)\Gamma(\beta-(1-\varepsilon)x)}},$$

so that $\lim_{x\to 0}\Gamma(x) = \infty$ implies that $\lim_{x\to\beta} Y(x,(1-\varepsilon)x) = 0$ and, therefore, there exists Q such that $d_{\mathrm{H}}(P_0,Q) = \sqrt{2}$.

Example 3.7 Let η be in the total variation neighbourhood of η_0, which corresponds to

$$g(x) = \begin{cases} 0, & 0 \le x \le \varepsilon; \\ x-\varepsilon, & \varepsilon < x \le \beta. \end{cases}$$

Given $0 < x < \varepsilon$, then it follows that $\lim_{t\to 0} Y(x,t) = 0$ implies that there exists Q such that $d_{\mathrm{H}}(P_0,Q) = \sqrt{2}$.

The results are not surprising because the maximum distance is achieved by considering a probability measure Q, which is not absolutely continuous w.r.t. P_0, as shown by $g(1) < \beta$; see Fortini and Ruggeri (1995), for details. A lesser distance is obtained when we consider the class of all probability measures whose c.f.'s are not below the one described in Example 3.3. The same distances in Table 1 are now the maximum distances in the class of the Dirichlet processes.

Finally, it is worth mentioning the following result about the distance between processes when considering random vectors $P(A_1), \ldots, P(A_k)$ and $Q(A_1), \ldots, Q(A_k)$, which shows that the distance between processes increases as we consider finer partitions. Let $f_{\eta_1,\ldots,\eta_k}$ be the density of a Dirichlet distributed random variable with parameters $(\eta_1, \ldots, \eta_k)$. Let $\{B_1, \ldots, B_k\}$ be a measurable partition of $\mathcal{A}$ and $\{B_{10}, B_{11}\}$ be a measurable partition of B_1. It follows that $P(B_1), \ldots, P(B_k)$ and $Q(B_1), \ldots, Q(B_k)$ are Dirichlet distributed with densities $f_{\eta_1,\ldots,\eta_k}$ and $f_{\gamma_1,\ldots,\gamma_k}$, respectively.

Theorem 3.8 We have that

$$\sup_{\{B_1,\ldots,B_k\}} d(\{P(B_1), \ldots, P(B_k)\}, \{Q(B_1), \ldots, Q(B_k)\})$$

is a nondecreasing function of k.

Proof It follows from

$$\frac{\int_0^1 \sqrt{f_{\eta_{10},\eta_{11},\ldots,\eta_k} f_{\gamma_{10},\gamma_{11},\ldots,\gamma_k} dy}}{\int_0^1 \sqrt{f_{\eta_1,\ldots,\eta_k} f_{\gamma_1,\ldots,\gamma_k} dy}} = \int_0^1 \sqrt{f_{\eta_{10},\eta_{11}} f_{\gamma_{10},\gamma_{11}} dy} \leq 1.$$

■

3.2 Kullback-Leibler divergence

The distance between Dirichlet processes could be measured by means of other indices, like the coefficients of divergence. In particular, we consider now the Kullback-Leibler divergence and we show that the results are not very different from the previous ones. Such an index is not a proper distance because it is not symmetric, but it can be useful when interested in measuring distances of measures from a given one.

Definition 3.9 Given the Dirichlet processes P and Q on $(\mathcal{X}, \mathcal{A})$, their distance is given by

$$d_{\mathrm{KL}}(P, Q) = \sup_{A \in \mathcal{A}} d(P(A), Q(A)),$$

where $d(X, Y)$ denotes the Kullback-Leibler divergence between two random variables whose distributions have densities p and q w.r.t. a dominating measure μ, i.e.,

$$d(X, Y) = \int p \log\left(\frac{p}{q}\right) d\mu.$$

The quantity $d(P(A), Q(A)) = Y(\eta_1(A), \eta_2(A))$ equals

$$\int_0^1 f_{\eta_1(A)}(y) \log\left(\frac{f_{\eta_1(A)}(y)}{f_{\eta_2(A)}(y)}\right) dy = \int_0^1 \frac{y^{\eta_1(A)-1}(1-y)^{\beta-\eta_1(A)-1}\Gamma(\beta)}{\Gamma(\eta_1(A))\Gamma(\beta-\eta_1(A))}$$
$$\times \log\left(\frac{y^{\eta_1(A)-1}(1-y)^{\beta-\eta_1(A)-1}\Gamma(\beta)\Gamma(\eta_2(A))\Gamma(\beta-\eta_2(A))}{y^{\eta_2(A)-1}(1-y)^{\beta-\eta_2(A)-1}\Gamma(\eta_1(A))\Gamma(\beta-\eta_1(A))\Gamma(\beta)}\right) dy$$
$$= -\log(\Gamma(\eta_1(A))) - \log(\Gamma(\beta-\eta_1(A))) + \log(\Gamma(\eta_2(A)))$$
$$+ \log(\Gamma(\beta-\eta_2(A))) + (\eta_1(A) - \eta_2(A))$$
$$\times\{\Psi(\eta_1(A)) - \Psi(\beta-\eta_1(A))\}.$$

We now compute the distance between any Dirichlet process Q with parameter η_2 from a baseline Dirichlet process P with parameter η_1.

Theorem 3.10 We have that

$$d_{\mathrm{KL}}(P, Q) = \sup_{0\le x\le\beta} Y(x, \varphi(x)),$$

where φ is the c.f. of η_2 w.r.t. η_1.

Proof Consider $\mathcal{A}_x = \{A \in \mathcal{A} : \eta_1(A) = x\}$, $0 \le x \le \beta$. We want to find $A \in \mathcal{A}_x$, which maximises $Y(x, \eta_2(A))$. It follows that

$$\frac{\partial Y}{\partial \eta_2(A)} = \frac{\Gamma'(\eta_2(A))}{\Gamma(\eta_2(A))} - \frac{\Gamma'(\beta-\eta_2(A))}{\Gamma(\beta-\eta_2(A))} - \mathcal{E}_{\eta_1(A)}L,$$

where $\mathcal{E}_\eta L$ denotes the expected value of $L = \log(T/(1-T))$, when T is beta distributed with parameters $(\eta, \beta-\eta)$. Furthermore,

$$\frac{\partial^2 Y}{\partial \eta^2(A)} = \frac{\Gamma''(\eta_2(A))\Gamma(\eta_2(A)) - \left(\Gamma'(\eta_2(A))\right)^2}{(\Gamma(\eta_2(A)))^2} -$$
$$- \frac{-\Gamma''(\beta-\eta_2(A))\Gamma(\beta-\eta_2(A)) + \left(\Gamma'(\beta-\eta_2(A))\right)^2}{(\Gamma(\eta_2(A)))^2}$$
$$= \mathrm{Var}_{\eta_2(A)} \log(T) + \mathrm{Var}_{\beta-\eta_2(A)} \log(T) > 0.$$

The convexity of Y implies that Y takes its maximum value at either $\inf_{A\in\mathcal{A}_x} \eta_2(A)$ or $\sup_{A\in\mathcal{A}_x} \eta_2(A)$, i.e., as discussed in Section 2, at either $\varphi(x)$ or $\beta - \varphi(\beta - x)$. The same argument can be repeated for any $x \in [0, \beta]$. Since it can be proved that $Y(x, \beta - \varphi(\beta - x)) = Y(\beta - x, \varphi(\beta - x))$, it follows that it is sufficient to consider $Y(x, \varphi(x))$. ∎

Therefore, the distance between two processes is found by computing the c.f. $\varphi(x)$ and then maximising the function $Y(x, \varphi(x))$, as in the following examples.

Example 3.11 (Example 3.3 continued) For $x \le 0.5$ and $\beta = 1$, it follows that

$$Z_x(\gamma) = \frac{\partial Y(x, \varphi(x))}{\partial x} = -\gamma\pi\cot(\gamma\pi x) + \gamma\pi\cot(\pi x) + \frac{(1-\gamma)\pi^2 x}{\sin^2(\pi x)},$$

so that $Z_x(1) = 0$ for all $x \in [0,1]$. In addition,

$$\frac{\partial Z_x(\gamma)}{\partial \gamma} = -\pi \cot(\gamma\pi x) + \frac{\gamma\pi^2 x}{\sin^2(\gamma\pi x)} + \pi\cot(\pi x) - \frac{\pi^2 x}{\sin^2(\pi x)},$$

which equals 0 at $\gamma = 1$. The quantity within brackets in

$$\frac{\partial^2 Z_x(\gamma)}{\partial \gamma^2} = \frac{2\pi^2 x}{\sin^3(\gamma\pi x)} \{\sin(\gamma\pi x) - \gamma\pi x\cos(\gamma\pi x)\}$$

is always positive (obvious for $\gamma x > 1/2$, whereas otherwise we should look at $t < \tan(t)$, for $0 \le t < \pi/2$).

Therefore, $\partial^2 Z_x(\gamma)/\partial\gamma^2 > 0$ implies that $\partial Z_x(\gamma)/\partial\gamma$ is increasing and, because of $\{\partial Z_x(\gamma)/\partial\gamma\}_{\gamma=1} = 0$, that it negative for $\gamma < 1$. As a consequence, then Z_x is decreasing as a function of γ and, because of $Z_x(1) = 0$, it is positive for any $\gamma \in (0,1)$, given any $x \in [0,1]$. Thus, $\partial Y(x,\varphi(x))/\partial x$ is positive for $x \le 0.5$, while it can be similarly proved that it is negative for $x > 0.5$, so that $Y(x,\varphi(x))$ achieves its maximum value at $x = 0.5$. Finally, it can be shown that $\lim_{\gamma\to 0} Y(x,\varphi(x)) = \infty$, so that $d_{\mathrm{KL}}(P,Q) = \infty$.

Distances are computed for different values of γ and shown in Table 2.

Table 2. Kullback-Leibler divergence.

γ	0.0001	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Distance	8.76	1.9	1.2	0.79	0.53	0.35	0.21	0.12	0.05	0.01	0.0

Example 3.12 (Example 3.4 continued) It can be shown that $d_{\mathrm{KL}}(P,Q)$ is unbounded also in this case.

As for Hellinger distance, we can prove a theorem similar to Theorem 3.5, which simplifies the computation of the distance when considering the class Λ of all c.f.'s of the parameters $\eta \in \Gamma$ w.r.t. η_0.

Theorem 3.13 For any $x \in [0,1]$, define $\hat\varphi(x) = \inf_{\varphi\in\Lambda} \varphi(x)$. Then, it follows that

$$\sup_{Q\in\Pi} d_{\mathrm{KL}}(P_0,Q) = \sup_{0\le x\le\beta} Y(x,\hat\varphi(x)).$$

As before about the Hellinger distance, the class Γ could be described by means of the c.f.

Example 3.14 (Example 3.6 continued) For the ε-contaminated class with $\beta = 1$, we obtain that $\lim_{x\to\beta} Y(x,(1-\varepsilon)x) = \infty$ and therefore there exists Q such that $d_{\mathrm{H}}(P_0,Q) = \infty$ (the proof is the same as in Example 3.3).

Example 3.15 (Example 3.7 continued) In the total variation neighbourhood of η_0, it can be proved that, given $0 < x < \varepsilon$, then $\lim_{t\to 0} Y(x,t) = +\infty$, so that there exists Q, such that $d_{\mathrm{KL}}(P_0,Q) = \infty$.

4. Distance Between Distributions of Probabilities

Whereas the parametric robust Bayesian approach is often interested in finding upper and lower bounds on posterior set probabilities and expectations in the parameter space, the nonparametric approach must deal with the changes in the distribution of probabilities and expectations in the sample space. In this paper, we focus only on set probabilities, simply mentioning that ranges of distributions could be found for all random functionals, e.g., for those for which Cifarelli and Regazzini (1990) have computed the distribution functions.

Lemma 4.1 Let $0 < \eta < \beta$. Then, the function

$$Y(\eta) = \frac{\int_0^x y^{\eta-1}(1-y)^{\beta-\eta-1}dy}{\int_0^1 y^{\eta-1}(1-y)^{\beta-\eta-1}dy}$$

is strictly decreasing in η.

Proof Let $L(y) = \log(y/(1-y))$ and $f_\eta(y) = y^{\eta-1}(1-y)^{\beta-\eta-1}$. Then,

$$Y^{'}(\eta) = \frac{Z_\eta(x)}{\left(\int_0^1 f_\eta(y)dy\right)^2},$$

with

$$Z_\eta(x) = \int_0^x L(y)f_\eta(y)\,dy \int_x^1 f_\eta(y)\,dy - \int_x^1 L(y)f_\eta(y)\,dy \int_0^x f_\eta(y)\,dy.$$

For any $\eta \in (0,\beta)$, it follows that $\partial Z_\eta(x)/\partial x = 0$ inside $(0,1)$ at the unique point $\hat{x}_\eta$ such that $L(\hat{x}_\eta) = \mathcal{E}_\eta L$, being $\hat{x}_\eta$ a minimum, since $Z_\eta(1/2) < Z_\eta(0) = Z_\eta(1) = 0$. It follows that, for all η's, Z_η is negative, except for $x = 0, 1$, so that $Y(\eta)$ is strictly decreasing in η. ■

Lemma 4.2 Let $0 < \eta_1, \eta_2 < \beta$ and P and Q be Beta distributions with parameters $(\eta_1, \beta-\eta_1)$ and $(\eta_2, \beta-\eta_2)$, respectively, and densities f_{η_1} and f_{η_2}. Then, the c.f. of Q w.r.t. P can be computed, for any $y \in [0,1]$, as

$$x = \int_0^y f_{\eta_1}(t)dt,\ \varphi(x) = \int_0^y f_{\eta_2}(t)dt\ \eta_2 \geq \eta_1,$$
$$x = \int_y^1 f_{\eta_1}(t)dt,\ \varphi(x) = \int_y^1 f_{\eta_2}(t)dt\ \eta_2 \leq \eta_1.$$

Proof It can be easily shown that the likelihood ratio is given by

$$h(\theta) = \frac{f_{\eta_2}(\theta)}{f_{\eta_1}(\theta)} = K\left(\frac{\theta}{1-\theta}\right)^{\eta_2-\eta_1},$$

where K is a constant independent of θ. As $h(\theta)$ is increasing (decreasing) for $\eta_2 > \eta_1$ ($\eta_2 < \eta_1$), it follows, using the same notation as in Section 2, that L_x has the form $[0, y_x]$ ($[y_x, 1]$), for all $x \in [0,1]$. ■

We can now consider the distribution function of the probability $P(A)$ of $A \in \mathcal{A}$, when P is chosen from a Dirichlet process. Suppose there exists a baseline process P_0 with parameter η_0 and a class of processes Q with parameters η in Γ and $\eta_0 \in \Gamma$. We compare the distributions of $P(A)$ under the processes Q with the one under P_0 by using their c.f.'s, as presented in Fortini and Ruggeri (1994). For each $x \in [0,1]$, we look for the lowest c.f. $\hat{\varphi}(x)$, i.e., for the minimum probability, under the distributions of $Q(A)$, of all subsets having probability x under the distribution of $P_0(A)$.

Theorem 4.3 Let Q_1 and Q_2 be the Dirichlet processes with parameters η_1 and η_2, respectively, such that $\eta_1(A) = \inf_{\eta\in\Gamma} \eta(A)$ and $\eta_2(A) = \sup_{\eta\in\Gamma} \eta(A)$. Let φ_i be the c.f. of $Q_i(A)$ w.r.t. $P_0(A)$, $i = 1, 2$, then, for all $x \in [0,1]$, $\hat{\varphi}(x) = \min\{\varphi_1(x), \varphi_2(x)\}$.

Proof Consider any $\eta \in \Gamma$ such that $\eta(A) > \eta_0(A)$. Let Q be the corresponding Dirichlet process, then it follows from Lemma 4.2 that the c.f. of $Q(A)$ w.r.t $P_0(A)$ is given by

$$x = \int_0^y f_{\eta_0}(t)dt \quad \text{and} \quad \varphi(x) = \int_0^y f_\eta(t)dt.$$

From Lemma 4.1, it follows that $\int_0^y f_\eta(t)dt$ is decreasing in η, so that it is minimised by $\eta_2(A)$. A similar argument can be applied to η's such that $\eta(A) < \eta_0(A)$. ■

Example 4.4 Take the ε-contaminated class $\Gamma = \{\eta = (1-\varepsilon)\eta_0 + \varepsilon Q, Q \in \mathcal{M}_\beta\}$, where $\mathcal{M}_\beta$ is the class of all finite measures Q such that $Q(\mathcal{X}) = \eta(\mathcal{X}) = \beta$. In this case, Q_1 and Q_2 are such that $\eta_1(A) = 0$ and $\eta_2(A) = \beta$, respectively. It follows from Theorem 2.2 that $Q_1(A) = 0$ a.s. and $Q_2(A) = \beta$ a.s., i.e., $Q_1(A)$ and $Q_2(A)$ are Dirac measures concentrated at 0 and β, respectively. Therefore, their c.f.'s w.r.t. $P_0(A)$ are such that $\varphi_1(x) = \varphi_2(x) = \hat{\varphi}(x) = 0$, for all $x \in [0,1]$.

The same $\hat{\varphi}$ is obtained if we consider the parameters η and η_0 updated after observing a sample Z_1 of size 1 (extension to larger sample is trivial). If $Z_1 \in A$, then the posterior $Q_2^*(A)$ is a Dirac measure concentrated at $\beta+1$. Otherwise, the posterior $Q_1^*(A)$ is a Dirac measure concentrated at 0. In both cases, the corresponding c.f. w.r.t. the updated η_0^* is equal to zero everywhere.

Example 4.5 Consider all the parameters $\eta = \beta\gamma$, for $\beta > 0$, with the probability measure γ in the Fréchet class $\Gamma(F, G)$ and, as a baseline parameter, any $\eta_0 = \beta\gamma_0$ with γ_0 in Γ, e.g., the independent one with joint distribution $F(x)\,G(y)$. Take the subset $A = (-\infty, x] \times (-\infty, y]$, then it follows that Q_1 and Q_2 are such that $\eta_1(A) = \beta W(x,y)$ and $\eta_2(A) = \beta M(x,y)$. Update η and η_0 after observing a sample of size n, then the updated Q_1^* and Q_2^* are such that $\eta_1^*(A) = (\beta+n)\hat{W}(x,y)$ and $\eta_2^*(A) = (\beta+n)\hat{M}(x,y)$.

5. Distance Between Sets of Probability Measures

In this section, we compare the probabilities given by a class of Dirichlet processes to some subspaces of the space $\mathcal{P}$ of all the probability measures defined on the space $(\mathcal{X}, \mathcal{A})$. Given the subsets $A \in \mathcal{A}$ and $B \in \mathcal{B}$, where $\mathcal{B}$ is the Borel σ-field on $[0,1]$, consider the subspace $\Gamma = \{\Pi \in \mathcal{P}\colon \Pi(A) \in B\}$. Let P be a Dirichlet process on $(\mathcal{X}, \mathcal{A})$ with parameter η, then the random probability $\mathrm{P}(A)$ is beta distributed with parameters $(\eta(A), \beta - \eta(A))$ and density $f_{\eta(A)}$. The following result can therefore be easily proved.

Theorem 5.1 We have that

$$\mathcal{P}(\Gamma) = \int_B f_{\eta(A)}(x)dx.$$

More generally, it is possible to consider a measurable partition $A_1, \ldots, A_k$ of $\mathcal{X}$, and the measurable subsets $B_1, \ldots, B_k$ in $\mathcal{B}$, and define the class

$$\Gamma = \{\Pi \in \mathcal{P}\colon \Pi(A_i) \in B_i, \quad i = 1, \ldots, k\}.$$

From Definition 2.1, it follows that the joint distribution of the random probabilities $(\mathrm{P}(A_1), \ldots, \mathrm{P}(A_k))$ is Dirichlet with parameters $(\eta(A_1), \ldots, \eta(A_k))$ and density $f_{\eta(A_1),\ldots,\eta(A_k)}$. The following result can therefore be easily proved.

Theorem 5.2 We have that

$$\mathcal{P}(\Gamma) = \int_{B_1\times\ldots\times B_k} f_{\eta(A_1),\ldots,\eta(A_k)}(x_1, \ldots, x_k)\, dx_1 \ldots dx_k.$$

It should be observed that Theorem 5.1 is a special case of Theorem 5.2, when $k = 2$, but it has been presented separately to have a handy reference in the forthcoming examples.

Subsets of probability measures can be defined by means of the random functionals considered by Cifarelli and Regazzini (1990), who provide the distribution $\mathcal{M}$ of the random functional

$$Y_\psi = \int_{\Re} \psi(x)\, \mathrm{P}(dx).$$

We can now consider the class $\Gamma = \{\Pi \in \mathcal{P}\colon Y_\psi \in B\}$ and we get

$$\mathcal{P}(\Gamma) = \int_B d\mathcal{M}(x). \tag{2}$$

Consider now a family $\mathcal{R}$ of Dirichlet processes, where the parameter η belongs to a class Δ. Like in the parametric Bayesian robust analysis, the sensitivity to the changes in η is measured by considering upper and lower bounds on the probabilities of subsets of the space of all probabilities. In particular, it is possible to consider α in the Fréchet class $\Gamma(F, G)$, observing that Equation (2) does not apply to such case because it holds only for measures on $\Re$. Because of the nature of the Dirichlet process, it is possible to define some classes of probability measures $\Pi = \{\mathrm{P}\colon \mathrm{P}(A) \in B\}$, where $A \in \mathcal{A}$ and B is a Lebesgue measurable subset in $[0, 1]$. Similar classes could be defined by asking either that the random probabilities of a finer partition of $\mathcal{X}$ belong to some subset or the random functional, considered by Cifarelli and Regazzini (1990), takes value on some subset. Being $\mathrm{P}(A)$ beta distributed, it is easy to compute the probability of Π. As in parametric robustness, it is worthwhile to compute upper and lower bounds on the probability $\mathcal{P}(\Pi)$.

As an example, we could take $\Pi = \{\mathrm{P}\colon \mathrm{P}(A) \le x\}$ and the parameter η of the Dirichlet process in a class Γ, so that

$$\mathcal{P}(\Pi) = \frac{\int_0^x y^{\eta(A)-1}(1-y)^{\beta-\eta(A)-1}\, dy}{\int_0^1 y^{\eta(A)-1}(1-y)^{\beta-\eta(A)-1}\, dy}.$$

Applying Lemma 4.1, it follows that upper and lower bounds on $\mathcal{P}(\Pi)$ are achieved for $\eta(A)$ equal to $\inf_{\eta\in\Gamma} \eta(A)$ and $\sup_{\eta\in\Gamma} \eta(A)$, respectively. Another interesting case is given by $A = (\infty, y)$, so that we get a class of random distribution functions $\{F\colon F(y) < x\}$. For example, for $x = 1/2$, we have the class of all the distributions whose median is greater than y. Bounds can be easily computed for the Fréchet class described in Example 4.5, too.

6. Distance Between Bayes estimators

In the literature on Bayesian nonparametrics, many Bayes estimators of quantities of interest have been presented. It could be attractive to investigate how much they change as the parameter η of a Dirichlet process varies in a class Γ. This problem can be reduced to a usual one in parametric robustness. Consider the estimation of the mean and the estimation of a distribution function, both under squared loss, solved in Ferguson (1973). We consider the no-sample problem, being similar to the one with data. The Bayes estimator of the mean is given by $\int_{\Re} x\, d\alpha(x)$ and η could be in most of the classes considered in the parametric literature, for which methods for computing bounds are well known. As an example, let the density η be bounded within the densities l and u. Then, the upper bound on the Bayes estimator is given by $\hat{\eta} \equiv l$ on $(-\infty, x]$ and $\hat{\eta} \equiv u$ on (x, ∞), with x being determined by $\hat{\eta}(\Re) = \eta(\Re)$. In the Fréchet class about $Z = (X, Y)$, the Bayes estimators of variance and means of X and Y are constant; see Ruggeri (1994a, Section 4.1.2). We are thus facing the nice situation of having a large class, say $\Gamma(F, G)$, in which robustness is achieved. Upper and lower bounds on the covariance have been found too in Ruggeri (1994a, Section 4.1.2).

Following a suggestion in Ruggeri (1994b), it could be checked if the Bayes estimator of the distribution, say $\eta(-\infty, x)/\eta(\Re)$, is within a prespecified band, maybe around a baseline distribution function. In the Fréchet class, bounds are given by $\hat{W}$ and $\hat{M}$.

7. Comparing Prior and Posterior Distances

So far, we have measured the distance between quantities of interest, presenting results which are valid both before and after observing a sample. Now, we want to check if the sample influences the above distances, e.g., reducing them *a posteriori*.

First of all, it is worth considering the coefficient of divergence

$$\rho(\eta, g) = \int_{[0,\infty)} g(t)\, dH_\eta(t) + \eta_s(\Theta) \lim_{t \to \infty} \frac{g(t)}{t}$$

described in Section 2, when applied to compare the distance between a parameter $\eta \in \Gamma$ and a baseline η_0. It follows that prior and posterior distances coincide for most of the well-known indices (Kullback-Leibler, χ^2-divergence and Pietra) but not for Gini's one, that is for all indices for which $g(1) = 0$, i.e., which compare probability measures giving no weight where the two measures coincide.

Theorem 7.1 Let $\rho(\eta, g)$ be the coefficient of divergence of η w.r.t. η_0 and $\rho(\eta^*, g)$ be the coefficient of divergence of η^* w.r.t. η_0^*, updated parameters after observing a sample of size n. Then, it follows that

$$\rho(\eta^*, g) = \rho(\eta, g) + ng(1).$$

Proof Using the notations described in Section 2, we know that

$$H(y) = \eta_0\left(\{\theta \in \Theta \colon h(\theta) \le y\}\right) \text{ and } c_x = \inf\{y \in \Re \colon H(y) \ge x\},$$

whereas

$$H^*(y) = \eta_0^*\left(\{\theta \in \Theta : h(\theta) \leq y\}\right) = \begin{cases} H(y), & 0 \leq x < 1; \\ H(y) + n, & x \geq 1; \end{cases}$$

and

$$c_x^*(x) = \inf\{y \in \Re\colon H^*(y) \geq x\} = \begin{cases} c(x), & x < H(1); \\ 1, & H(1) \leq x \leq H(1) + n; \\ c(x-n), & x > H(1) + n. \end{cases}$$

The result is proved by observing that

$$\eta_s(\Theta) = \eta_s^*(\Theta) \quad \text{and} \quad \int_{[0,\infty)} g(t)dH^*(t) = \int_{[0,\infty)} g(t)dH(t) + ng(1).$$

■

Consider now the distance between Dirichlet processes as described in Section 5. As expected, any sample decreases such a distance. In the following, we suppose that data come from the "true" distribution $\tilde{\eta}$ and that the support of $\tilde{\eta}$ contains the support of the parameters of the Dirichlet processes.

Theorem 7.2 We have that $d_{\mathrm{H}}(P^*, Q^*) < d_{\mathrm{H}}(P, Q)$, where P^* and Q^* are the Dirichlet processes obtained by updating the processes P and Q after observing a sample of size n.

Proof We consider a sample of size 1 because the case n is obtained by reiteratively applying the proof for the case $n = 1$. We do not consider the case $\eta_1(A) = \eta_2(A)$, which gives a null distance, both a priori and *a posteriori*. We prove, first of all, that

$$Y(\eta_1(A) + \delta_{Z_1}(A), \eta_2(A) + \delta_{Z_1}(A)) > Y(\eta_1(A), \eta_2(A)),$$

where $Y(\eta_1(A), \eta_2(A)) = \int_0^1 \sqrt{f_{\eta_1(A)}(y) f_{\eta_2(A)}(y)}dy$. Note that, in fact,

$$\begin{aligned} \frac{Y(\eta_1(A) + \delta_{Z_1}(A), \eta_2(A) + \delta_{Z_1}(A))}{Y(\eta_1(A), \eta_2(A))} &= \frac{\Gamma((\eta_1(A) + \eta_2(A))/2 + \delta_{Z_1}(A))}{\Gamma((\eta_1(A) + \eta_2(A))/2)} \\ &\quad \times \frac{\Gamma(\beta + 1 - (\eta_1(A) + \eta_2(A))/2 - \delta_{Z_1}(A))}{\Gamma(\beta - (\eta_1(A) + \eta_2(A))/2)} \\ &\quad \times \sqrt{\frac{\Gamma(\eta_1(A))}{\Gamma(\eta_1(A) + \delta_{Z_1}(A))} \frac{\Gamma(\beta - \eta_1(A))}{\Gamma(\beta + 1 - \eta_1(A) - \delta_{Z_1}(A))} \frac{\Gamma(\eta_2(A))}{\Gamma(\eta_2(A) + \delta_{Z_1}(A))}} \\ &\quad \times \sqrt{\frac{\Gamma(\beta - \eta_2(A))}{\Gamma(\beta + 1 - \eta_2(A) - \delta_{Z_1}(A))}}. \end{aligned}$$

The above quantity equals

$$\frac{\beta - (\eta_1(A) + \eta_2(A))/2}{\sqrt{(\beta - \eta_1(A))(\beta - \eta_2(A))}},$$

when $Z_1 \notin A$ and

$$\frac{(\eta_1(A)+\eta_2(A))/2}{\sqrt{\eta_1(A)\eta_2(A))}},$$

when $Z_1 \in A$. Both quantities are greater than one if and only if $(\eta_1(A)-\eta_2(A))^2 > 0$, which is always true.

To complete the proof, observe that the function Y achieves its infimum *a posteriori* for a subset A^* such that $\eta_0^*(A^*) = x$ and $\eta^*(A^*) = \varphi(x)$. Since the prior Y is strictly less than the posterior one, its prior infimum is lower than the prior Y evaluated at A^*, so it is lower than the posterior infimum. It follows that the distance is strictly decreasing when a sample is given. ■

Theorem 7.3 We have that, a.s.,

$$\lim_{n\to\infty} d_{\mathrm{H}}(P,Q) = 0.$$

Proof Given a sample Z_i and a subset $A \in \mathcal{A}$, then $\delta_{Z_i}(A)$ can be seen as a Bernoulli random variable having mean $\tilde{\eta}(A)$, the "true" probability of A (to avoid triviality, we suppose that $0 < \tilde{\eta}(A) < 1$). Because of the strong law of large numbers, it follows that

$$\sum_{i=1}^{n} \frac{\delta_{Z_i}(A)}{n} \to \tilde{\eta}(A) \quad \text{a.s.},$$

so that $\sum_{i=1}^n \delta_{Z_i}(A)$ is unbounded (and similarly $n - \sum_{i=1}^n \delta_{Z_i}(A)$).

Applying the asymptotic formula presented in Abramowitz and Stegun (1972, p. 257) given by

$$\Gamma(az+b) \sim \sqrt{2\pi} e^{-az}(az)^{az+b-1/2}, \quad z\to\infty,$$

it follows that

$$Y(\eta_1(A) + \sum_{i=1}^{n}\delta_{Z_i}(A), \eta_2(A) + \sum_{i=1}^{n}\delta_{Z_i}(A)) \sim 1 \quad \text{a.s.},$$

completing the proof. ■

Theorem 7.4 We have that

$$d_{\mathrm{KL}}(P^*,Q^*) < d_{\mathrm{KL}}(P,Q),$$

where P^* and Q^* are the Dirichlet processes obtained by updating the processes P and Q after observing a sample of size n.

Proof We consider a sample of size 1 because the case n is obtained by reiteratively applying the proof for the case $n = 1$. We do not consider the case $\eta_1(A) = \eta_2(A)$, which gives a null distance, both *a priori* and *a posteriori*. We prove, first of all, that

$$e^{Y(\eta_1(A)+\delta_{Z_1}(A),\eta_2(A)+\delta_{Z_1}(A))} < e^{Y(\eta_1(A),\eta_2(A))},$$

where

$$Y(\eta_1(A), \eta_2(A)) = \int_0^1 f_{\eta_1(A)}(y) \log\left(\frac{f_{\eta_1(A)}(y)}{f_{\eta_2(A)}(y)}\,dy\right).$$

Such inequality is always satisfied because simple computations show that proving it is equivalent to proving, when $Z_1 \notin A$, that

$$\frac{\beta - \eta_2(A)}{\beta - \eta_1(A)}\, e^{(\eta_2(A)-\eta_1(A))/(\beta-\eta_1(A))} < 1$$

and, when $Z_1 \in A$, that

$$\frac{\eta_2(A)}{\eta_1(A)}\, e^{(\eta_1(A)-\eta_2(A))/\eta_1(A)} < 1.$$

To complete the proof, observe that the function Y achieves its supremum *a posteriori* for a subset A^* such that $\eta_0^*(A^*) = x$ and $\eta^*(A^*) = \varphi(x)$. Since the prior Y is strictly greater than the posterior one, its prior supremum is greater than the prior Y evaluated at A^*. So, it is greater than the posterior supremum. It follows that the distance is strictly decreasing whence a sample is given. ■

Theorem 7.5 We have that, a.s.,

$$\lim_{n\to\infty} d_{\mathrm{KL}}(P, Q) = 0.$$

Proof Like in the proof of Theorem 7.3, we see that

$$Y(\eta_1(A) + \sum_{i=1}^{n} \delta_{Z_i}(A), \eta_2(A) + \sum_{i=1}^{n} \delta_{Z_i}(A)) = o(1) \quad \text{a.s.}$$

Applying both the same asymptotic formula and the following one:

$$\Psi(z) \sim \log(z) - \frac{1}{2z} - \sum_{n=1}^{\infty} \frac{B_{2n}}{2nz^{2n}}, \quad z \to \infty,$$

where B_n are the Bernoulli numbers; see Abramowitz and Stegun (1972, p.259). ■

8. Discussion

In this paper, we have presented some results about performing sensitivity analysis when the probability measure is chosen by a Dirichlet process, whose parameter is specified with uncertainty. Further research could deal with different processes, such as Pólya trees, which could choose, almost surely, probability measures absolutely continuous with respect to the Lebesgue measure. Also, different distances between processes could be treated, such as Prohorov. Finally, the approach taken in this paper could be applied to distributions of probabilities of finer partitions or random functionals.

Acknowledgements

The author is deeply indebted to Michael Lavine for his comments.

References

Abramowitz, M., Stegun I.A., 1972. Handbook of Mathematical Functions. Dover, New York.

Ali, S.M., Silvey, S.D., 1966. A general class of coefficients of divergence of one distribution from another. Journal of The Royal Statistical Society Series B—Statistical Methodology, 28, 131-142.

Berger, J., 1994. An overview of robust Bayesian analysis. Test, 3, 5-58.

Cifarelli, D.M., Regazzini, E., 1987. On a general definition of concentration function. Sankhyā, 49, 307-319.

Cifarelli, D.M., Regazzini, E., 1990. Distribution functions of means of Dirichlet process. Annals of Statistics, 18, 429-442.

Csiszár, I., 1967. Information-type measures of difference of probability distributions and indirect observations. Studia Scientiarum Mathematicarum Hungarica Stud, 2, 299-318.

Ferguson, T.S., 1973. A Bayesian analysis of some nonparametric problems. Annals of Statistics, 1, 209-230.

Fortini, S., Ruggeri, F., 1994. Concentration functions and Bayesian robustness. Journal of Statistical Planning and Inference, 40, 205-220.

Fortini, S., Ruggeri, F., 1995. On defining neighbourhoods of measures through the concentration function. Sankhyā, 56, 444-457.

Gini, C., 1914. Sulla misura della concentrazione della variabilità dei caratteri. Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti, 73, 1203-1248.

Pietra, G., 1915. Delle relazioni tra gli indici di variabilità. Atti del Reale Istituto Veneto di S.L.A. A.A., 74, 775-792.

Rios-Insua, D., Ruggeri, F., 2000. Robust Bayesian Analysis. Springer-Verlag, New York.

Ruggeri, F., 1994a. Bayesian nonparametrics, robustness and Fréchet classes. Ph.D. Dissertation. Duke University, Durham.

Ruggeri, F., 1994b. Discussion of "An overview of robust Bayesian analysis" by J. Berger. Test, 3, 108-109.

Vajda, I., 1990. Distances and discrimination rates for stochastic processes. Stochastic Processes and their Applications, 35, 47-57.