INTERNATIONAL CENTRE FOR ECONOMIC RESEARCH

WORKING PAPER SERIES

Stephen G. Walker, Antonio Lijoi, and Igor Prünster

CONTRIBUTIONS TO THE UNDERSTANDING OF BAYESIAN CONSISTENCY

Working Paper no. 13/2004
November 2004

APPLIED MATHEMATICS
WORKING PAPER SERIES

Contributions to the understanding of Bayesian consistency

Stephen G. Walker[1], Antonio Lijoi[2] and Igor Prünster[2,3]

[1] Institute of Mathematics, Statistics and Actuarial Science, University of Kent, Kent CT2 7NZ, UK. E-mail: S.G.Walker@kent.ac.uk

[2] Dipartimento Economia Politica e Metodi Quantitatavi, Università degli Studi di Pavia, via San Felice 5, 27100 Pavia, Italy. E-mail: lijoi@unipv.it, igor.pruenster@unipv.it

[3] ICER, Villa Gualino, viale Settimio Severo 63, 10133 Torino, Italy.

Abstract. Consistency of Bayesian nonparametric procedures has been the focus of a considerable amount of research. Here we deal with strong consistency for Bayesian density estimation. An awkward consequence of inconsistency is pointed out. We investigate reasons for inconsistency and precisely identify the notion of "data tracking". Specific examples in which this phenomenon can not occur are discussed. When it can happen, we show how and where things can go wrong, in particular the type of sets where the posterior can put mass.

Key words: Bayesian consistency; Density estimation; Hellinger distance; Weak neighborhood.

1. Introduction.

A first formulation of the issue of consistency of Bayesian inferential procedures is given in Doob (1949). It states that if there exists a consistent sequence of estimators of the unknown parameter, then the posterior estimates are consistent in the sense that the posterior distribution converges to a point mass at the unknown parameter outside a set of prior mass zero. A drawback of such an approach is that the null sets on which convergence fails could be relevant. In this case, the problem can be circumvented by resorting to a "frequentist" notion of consistency which gives rise to the "what if" method adopted by Diaconis & Freedman (1986). The idea consists in generating independent data from a "true" fixed distribution f_0 and checking whether the posterior accumulates in (suitably defined) neighborhoods of f_0. This corresponds to requiring the data to eventually swamp the prior.

An early use of the "what if" method can be found in Freedman (1963), where it is shown that weak consistency does not necessarily hold for priors supported by discrete distributions on a countable set of states. However, if the number of states is finite, consistency is achieved and the result extends to the countable case by introducing an additional entropy condition. A sufficient condition for weak consistency with more general priors is suggested in Schwartz (1965). This is solely a support condition. Further examples of inconsistency, involving mixtures, are illustrated in Diaconis & Freedman (1986).

When considering problems of density estimation, it is natural to ask for the strong consistency of Bayesian procedures. An early contribution in this area is the 1988 unpublished University of Illinois technical report by A.R. Barron, which is based on the use of uniformly consistent tests. Later developments, combining well–established techniques in the theory of empirical processes with ideas from Barron's report, provide sufficient conditions for strong consistency in terms of metric entropies. For instance, Barron, Schervish & Wasserman (1999) specify bracketing entropy conditions for strong consistency to hold true and apply their results to a number of commonly used priors in Bayesian nonparametric inference. Following the same lines, Ghosal, Ghosh & Ramamoorthi (1999) provide slightly weaker sufficient conditions for strong consistency in terms of the L_1–metric entropy and deal with mixtures of a Dirichlet process. This approach has also been employed for verifying strong consistency of specific priors in Bayesian Nonparametrics. See, for example, Petrone & Wasserman (2002). New ideas for solving consistency issues are given in Walker (2003; 2004), where a simple sufficient condition for strong consistency is represented by the finiteness of a suitable sum of square roots of prior probabilities.

The present paper aims at providing an understanding of the main issues that arise when dealing with consistency of Bayesian procedures. An argument which motivates the interest in consistency can be based on a notion of merging which differs from the classical one introduced by Blackwell & Dubins (1962). Indeed, we consider the case of two Bayesians sharing the same prior but collecting two independent data sets from the same density f_0. It turns out that if the prior is inconsistent at f_0 then the two Bayesians disagree even if more and more data are collected. This is quite an unpleasant feature. Given this, it is even more important to determine possible sources of (strong) inconsistency. In order to develop such an analysis, we still preserve the support condition introduced by Schwartz (1965). We illustrate how consistency at some density f_0 depends on the prior mass assigned to the "pathological" set of those densities that are close to f_0, in a weak sense, and far apart from f_0, in the L_1–metric. If the prior does not put mass on such sets, then (strong) consistency is achieved at f_0. Many priors of common use meet such a requirement. We do provide some related illustration. If the prior mass on such sets is positive, one has to take care about densities that track the data, a notion to be made precise later on. In order to get rid of the data tracking phenomenon, one has to look for sufficient conditions which avoid it. Regarding this aspect, we reconsider and slightly generalize a result of Walker (2004), which is given in terms of a summability condition of prior probabilities. We provide an interpretation and show that this sufficient condition is not necessary. Finally a new sufficient condition is provided. When applied to the prior of the counter–example in Barron et al. (1999) it nicely shows the reason of its inconsistency.

According to these guidelines, the structure of the paper is as follows. In Section 2 we introduce some notation and preliminary definitions. Section 3 starts with a description of our variation of the merging problem and investigates reasons for (strong) inconsistency. In subsection 3.1 we consider the case in which the above mentioned "pathological" set of densities has zero prior probability and show that some of the commonly used priors do not allow for inconsistency. Afterwards, in subsection 3.2 we look at a more general setting with a more detailed description of the data tracking behaviour. Then we consider and slightly generalize the simple sufficient condition given in Walker (2004) and provide an alternative condition which guarantees strong consistency of the prior. Section 4 connects ideas and provides further understanding concerning the sets of densities where inconsistency can occur.

2. Notation and some basic facts.

We consider a sequence of observations $(X_n)_{n\geq 1}$ each taking values in some metric, complete and separable space $\mathbb{X}$ endowed with a σ-algebra which we agree to denote by $\mathscr{X}$. If $\mathbb{F}$ indicates the space of probability density functions with respect to some measure λ on $\mathbb{X}$, the

$$d_H(f,g) = \left\{ \int_{\mathbb{X}} \left(\sqrt{f(x)} - \sqrt{g(x)} \right)^2 \lambda(\mathrm{d}x) \right\}^{1/2}$$

for any f and g in $\mathbb{F}$, and set $\mathscr{F}$ to be the Borel σ-algebra of $\mathbb{F}$. Suppose Π stands for a prior distribution on $(\mathbb{F},\mathscr{F})$. In this case, we assume that, given a density f drawn from Π, the observations are i.i.d. with common density f, that is

$$\mathrm{pr}\{(X_1,\ldots,X_n) \in A\} = \int_A \int_{\mathbb{F}} \left\{ \prod_{i=1}^n f(x_i) \right\} \Pi(\mathrm{d}f)\, \lambda(\mathrm{d}x_1) \cdots \lambda(\mathrm{d}x_n)$$

for each $n \geq 1$ and A in $\mathscr{X}^n$. The posterior distribution on $(\mathbb{F},\mathscr{F})$, given the observations $(X_1,\ldots,X_n)$, coincides with

$$\Pi_n(B) = \frac{\int_B \prod_{i=1}^n f(X_i)\, \Pi(\mathrm{d}f)}{\int_{\mathbb{F}} \prod_{i=1}^n f(X_i)\, \Pi(\mathrm{d}f)}$$

for all B in $\mathscr{F}$. The frequentist approach to Bayesian consistency is based on the idea of fixing a density f_0 as the "true" density from which the data are independently sampled and check whether the posterior accumulates in any Hellinger neighborhood of f_0. Hence Π is *strongly consistent* or, equivalently, *Hellinger–consistent* at f_0 if, for any $\epsilon > 0$,

$$\Pi_n(A_\epsilon) \to 1 \qquad \text{a.s. } [P_0^\infty]$$

as $n \to +\infty$, where $A_\epsilon = \{f \in \mathbb{F} : d_H(f, f_0) < \epsilon\}$, P_0 is the probability distribution whose density coincides with f_0 and P_0^∞ is the infinite product measure on $\mathbb{X}^\infty$. An alternative less stringent notion of consistency can be given by referring to the space $\mathbb{P}$ of probability distributions on $(\mathbb{X}, \mathscr{X})$, equipped with the weak topology. A weak neighborhood of any probability distribution P^* in $\mathbb{P}$ is the set

$$W_\epsilon = \left\{P \in \mathbb{P} : \left|\int \phi_i \, \mathrm{d}P - \int \phi_i \, \mathrm{d}P^*\right| < \epsilon \quad i = 1, \ldots, k\right\}$$

for a k-tuple of continuous and bounded real-valued functions ϕ_i defined on $\mathbb{X}$. In this case, we say that a prior Π is *weakly consistent* at f_0 if, for any $\epsilon > 0$,

$$\Pi_n(W_\epsilon) \to 1 \qquad \text{a.s. } [P_0^\infty],$$

as $n \to +\infty$, where W_ϵ stands for a weak neighborhood of P_0. Recall that the weak topology is coarser than the one induced by d_H, the latter being equivalent to the total variation topology on $\mathbb{P}$.

Another notion we need to consider is that of support of a prior. We say that P_0 is in the support of Π if any neighborhood of P_0 has positive Π–probability. According to the topology defined on $\mathbb{P}$, we distinguish weak and Hellinger support of Π which will be denoted by $S_W(\Pi)$ and $S_H(\Pi)$, respectively. One can reasonably think that P_0 being in $S_W(\Pi)$ would imply weak consistency of Π at P_0. Such a guess is wrong as shown, for example, by the counter–example in Diaconis & Freedman (1986) for mixtures of Dirichlet processes. Hence, one needs to impose a stronger support condition in order to achieve weak consistency at P_0. To this end, consider two probability distributions P and Q such that P is absolutely continuous with respect to Q and define the *Kullback-Leibler divergence* between P and Q as

$$D_K(P, Q) = \int \log\left(\mathrm{d}P/\mathrm{d}Q\right) \, \mathrm{d}P. \tag{1}$$

If $\mathbb{P}^*$ is a subset of $\mathbb{P}$ formed by all probability distributions dominated by a common σ–finite measure λ, (1) reduces to

$$d_K(f_P, f_Q) = \int f_P \, \log(f_P/f_Q),$$

where $f_P = \mathrm{d}P/\mathrm{d}\lambda$ and $f_Q = \mathrm{d}Q/\mathrm{d}\lambda$ are the densities of P and Q, respectively, with respect to λ for any $P, Q \in \mathbb{P}^*$. Hence, d_K can be seen as a measure of divergence on the corresponding space of densities $\mathbb{F}$. If $K_\epsilon = \{P \in \mathbb{P}^* : d_K(f_{P_0}, f_P) < \epsilon\}$ is a neighborhood of P_0 with respect to d_K, the probability distribution P_0 is in the Kullback-Leibler support of Π, $S_K(\Pi)$, if $\Pi(K_\epsilon) > 0$ for any $\epsilon > 0$. One may notice that $S_W(\Pi) \supset S_H(\Pi) \supset S_K(\Pi)$. A fundamental sufficient condition for obtaining weak consistency is due to Schwartz (1965): if P_0 is in $S_K(\Pi)$, then Π is weakly consistent at P_0.

When dealing with density estimation, it is more natural to ask for strong consistency and one might hope that a Kullback-Leibler support condition still suffices. However, as it has been shown in Barron et al. (1999), this does not happen without any further condition. All the contributions in this area aim at giving simple sufficient conditions for strong consistency and preserve the Kullback-Leibler support condition. In the following sections we attempt at understanding the deep reasons of possible strong inconsistencies in cases in which weak consistency holds true.

3. Inconsistency and possible solutions.

It is commonly agreed that consistency is an important property of statistical procedures. This is true in a Bayesian setting as well. Indeed, lack of consistency might yield unpleasant consequences of the type we are going to describe. Before proceeding with the illustration, it is worth recalling that a lot of attention in the literature has focused on the so–called merging of opinions. It essentially consists of a situation in which two Bayesians assess different priors and one is interested in checking whether their posterior inferences tend to coincide as long as more data are collected. Original work on this issue can be found in Blackwell & Dubins (1962) where it is proved that, under a condition of absolute continuity of one prior with respect to the other, merging of opinion occurs in the sense that the L_1–distance between predictive distributions becomes negligible as the sample size increases. Later discussions are provided, among others, in Diaconis & Freedman (1986) and in Ghosal et al. (1999). Essentially, the merging of opinion, or agreement, for large samples boils down to consistency. That is, posterior distributions accumulate around the same, and correct, density function. See, for example, Barron et al. (1999).

Here we consider a different setup which, to our knowledge, has not been investigated before. Suppose that two Bayesians are conducting the same experiment which naturally leads to deal with two independent samples. This involves the idea of replication of experiments, which is crucial for scientific advancement. Hence, independent and identically distributed samples are drawn from the same probability distribution, P_0, which has density with respect to the Lebesgue measure given by f_0. Both Bayesians agree on using the same prior distribution Π on the space of density functions. Naturally, they collect independent data sets from P_0, say $X_1^{(1)}, X_2^{(1)}, \ldots$ for the first Bayesian and $X_1^{(2)}, X_2^{(2)}, \ldots$ for the second Bayesian. In such a case, one would reasonably expect that for large samples the two Bayesians will agree with each other. However, we show that it is possible to construct priors, even ones which have f_0 in the Kullback-Leibler support of Π, for which agreement is not achieved. Let us first introduce some notation. Define $g : \mathbb{X}^n \times \mathbb{X}^n \to \mathbb{R}$ as a measurable function of the n–dimensional independent samples $X_1^{(j)}, \ldots, X_n^{(j)}$, for $j = 1, 2$. Moreover, denote with $E_0^{(j)}(g)$ the expectation of g with respect to the j–th sample $X_1^{(j)}, \ldots, X_n^{(j)}$ keeping fixed $X_1^{(l)}, \ldots, X_n^{(l)}$, where $l \neq j$.

Theorem 1 *Assume f_0 is in $S_K(\Pi)$. Then, if Π is not Hellinger–consistent at f_0,*

$$E_0^{(j)}\left\{D_K(\Pi_n^{(1)},\Pi_n^{(2)})\right\} > n\delta \qquad a.s.[P_0^\infty]$$

infinitely often, for some $\delta > 0$, having denoted with $\Pi_n^{(j)}$ the posterior distribution based on data set $X_1^{(j)},\ldots,X_n^{(j)}$, with $j = 1, 2$.

For the proof, the reader can refer to the Appendix.

Such an outcome is certainly startling for two Bayesians using the same prior and sampling from the same density. In particular, there is no merging of information (see for example Barron's technical report of 1988). Hence, identification of consistent priors and investigation of possible sources of inconsistency are important issues.

We first consider the latter issue and try to understand why the Kullback–Leibler support condition is sufficient for weak, but not for strong consistency. It is clear, indeed, that inconsistency at f_0 may be caused by sequences of densities that convergence weakly, but not in L_1, to f_0. An example of such a behavior is associated with the sequence of densities $f_n(x) = 1+\sin(2\pi nx)$ for x in $[0, 1]$. The corresponding sequence of distributions converges weakly to the uniform on $[0, 1]$, whereas f_n oscillates ever more wildly and does not converge to anything. The oscillating behavior, together with high peaks at the maxima of the f_n's, causes the undesirable phenomenon of "tracking the data". In other terms, data corresponding to these peaks remarkably increase the likelihood and thus may lead the posterior not swamping mass from the rough densities. Hence, one has to focus attention on the set

$$V_{\delta,\epsilon} = W_\delta \cap A_\epsilon^c$$

where W_δ and A_ϵ denote weak and Hellinger neighborhoods, respectively, of f_0.

Since by weak consistency the posterior Π_n will accumulate in W_δ, for any $\delta > 0$, and $V_{\delta,\epsilon}$ shrinks as δ goes to 0, the first issue to face is whether for all small enough δ the prior is prevented to put mass on $V_{\delta,\epsilon}$. Intuitively, one can figure this constraint as being allowed to track the data up to a finite number of observations. In the following subsections we first deal with the case in which $\Pi(V_{\delta,\epsilon}) = 0$, for any δ less than some fixed $\delta^* > 0$ and we then consider cases in which such a condition is not met.

3.1. Consistency with $\Pi(V_{\delta,\epsilon}) = 0$**.** Having identified in $V_{\delta,\epsilon}$ the set that might give rise to inconsistency, the first issue to face is to look for priors that satisfy $\Pi(V_{\delta,\epsilon}) = 0$. Consistency is automatically achieved in this case. Indeed, it turns out that some of the commonly used priors satisfy this condition. Here we provide an illustration by considering some noteworthy examples.

1. *Monotone decreasing densities.* Here we consider the case in which the prior is concentrated on monotone decreasing densities on $\mathbb{R}^+$. Bayesian nonparametric inference

with such priors is considered in Hansen & Lauritzen (2002). Besides dealing with theoretical and computational issues associated with Bayesian estimation in this setting, they point out consistency as an interesting aspect to investigate. To this end, let us first recall that, if F is a probability distribution function corresponding to some monotone decreasing density f, then the following representation holds true

$$F(x) = \int_{\mathbb{R}^+} F(x;\theta)\,\mathrm{d}G(\theta)$$

where G is a distribution function, $F(x;\theta) = \theta^{-1}\min\{x,\theta\}$ if $\theta > 0$ and $F(x;0)$ is degenerate at 0. Moreover G is uniquely determined by

$$G(\theta) = F(\theta) - \theta\, f(\theta).$$

We now verify that Hellinger consistency holds true for any $f_0(x) = \int_x^{+\infty} \frac{1}{\theta}\,\mathrm{d}G_0(\theta)$. Let W_δ be a δ–weak neighborhood of G_0. Then $G \in W_\delta$ implies the f is in an ϵ–Hellinger neighborhood, A_ϵ, of f_0. Assume G converges weakly to G_0, so that, for any $x > 0$,

$$\int_x^{+\infty} \frac{1}{\theta}\mathrm{d}G(\theta) \to \int_x^{+\infty} \frac{1}{\theta}\,\mathrm{d}G_0(\theta).$$

Consequently, by Scheffe's theorem, one has that $\int |f(x) - f_0(x)|\,\mathrm{d}x \to 0$, equivalently that $f \to f_0$ in Hellinger distance. To show the converse, we prove that $G \in W_\delta^c$ implies $f \in A_\epsilon^c$. Define a weak neighbourhood of G_0 as

$$W_\delta = \left\{ G : \left| \int_0^{+\infty} \frac{(\theta - y)\mathbb{I}_{(y,+\infty)}(\theta)}{\theta}\,\mathrm{d}G(\theta) - \int_0^{+\infty} \frac{(\theta - y)\mathbb{I}_{(y,+\infty)}(\theta)}{\theta}\,\mathrm{d}G_0(\theta) \right| < \delta \right\}.$$

for a fixed $y > 0$. As usual, $\mathbb{I}_A$ is the indicator function of set A. If $G \notin W_\delta$, then $|F(y) - F_0(y)| > \delta$, which yields $f \in A_\epsilon^c$ for $\epsilon < \delta$. Thus, one has that $\Pi(V_{\delta,\epsilon}) = 0$ and Hellinger consistency holds without any further assumption.

2. Mixture models. Consider the mixture model

$$f(x) = \int \phi_h(x - \theta)\,\mathrm{d}Q(\theta),$$

where $\phi_h(x-\theta)$ is the normal density function with mean θ and variance h^2. Moreover, Q has a nonparametric prior and μ is the prior distribution for h. This model is considered by Ghosal et al. (1999). It is assumed that

$$f_0(x) = \int \phi_{h_0}(x - \theta)\,\mathrm{d}Q_0(\theta)$$

is the true density function. Note that h can cause trouble by getting arbitrarily close to 0.

Let, first, $h_0 > \tau > 0$ and define neighbourhoods of (h_0, Q_0) as the set

$$W_{\delta,\tau} = \{(h, Q) : |h - h_0| < \tau, \;\; Q \in W_\delta\}.$$

If $(h, Q) \to (h_0, Q_0)$, then from

$$\begin{aligned} \left| \int \phi_h(x-\theta)\, \mathrm{d}Q(\theta) - \int \phi_{h_0}(x-\theta)\, \mathrm{d}Q_0(\theta) \right| &\le \int |\phi_h(x-\theta) - \phi_{h_0}(x-\theta)|\, \mathrm{d}Q_0(\theta) \\ &+ \left| \int \phi_h(x-\theta) \mathrm{d}\left(Q(\theta) - Q_0(\theta)\right) \right|. \end{aligned}$$

one has $f(x) \to f_0(x)$ pointwise for all x and Scheffe's theorem implies $f \to f_0$ in the Hellinger distance. This means that $(h, Q) \in W_{\delta,\tau}$ implies $f \in A_\epsilon$.

Now consider

$$|P(B) - P_0(B)| = \left| \int \left(\Phi_h(B;\theta) - \Phi_{h_0}(B;\theta)\right) \mathrm{d}Q_0(\theta) + \int \Phi_h(B;\theta) \mathrm{d}\left(Q(\theta) - Q_0(\theta)\right) \right|,$$

where $\Phi_h(B;\theta) = \int_B \phi_h(x-\theta)\mathrm{d}x$. Again, for h bounded away from 0, and excluding the case when $(h, Q) = (h_0, Q_0)$, we can always find a set B to make this positive. Now, consider the case in which h gets arbitrarily close to 0. It is easy to see that $|P(B) - P_0(B)| \to |Q(B) - P_0(B)|$. Hence, consistency can fail when the prior puts positive mass on h in a neighbourhood of 0 and positive mass on Q in Hellinger neighbourhoods of P_0. This problem can be circumvented by requiring Q and P to have different supports. For example, take P and Q with supports coinciding with the real line and with $[-a, a]$ for some finite and positive a, respectively. This ensures that the prior for Q puts zero mass in Hellinger neighbourhoods of P_0. Note that Ghosal et al. (1999) prove an analogous result by different arguments.

3. Finite-dimensional parametric family. Here we consider sampling models $\{f(x;\theta) : \theta \in \Theta\}$, where Θ is a finite–dimensional parameter space. Such families, provided the support condition is met by the prior, lead to consistency. The point is that $\Pi(V_{\delta,\epsilon}) = 0$ for some $\delta > 0$ and for all $\epsilon > 0$. For $f \in V_{\delta,\epsilon}$ for all $\delta > 0$ it is required that the density f be oscillating, the number of oscillations increasing to ∞ as $\delta \downarrow 0$. This just can not happen if f is based on a finite dimensional parameter.

To formalise this, we have the following simple conditions which should be easily verifiable for any particular $f(\cdot;\theta)$. If $f(\cdot;\theta_k) \to f(\cdot;\theta_0)$ weakly, that is

$$\int g(x)\, f(x;\theta_k)\, \mathrm{d}x \to \int g(x)\, f(x;\theta_0)\, \mathrm{d}x$$

for all bounded and continuous g, implies $|\theta_k - \theta_0| \to 0$. If $\theta \mapsto f(x;\theta)$ is continuous almost everywhere with respect to the Lebesgue measure, this in turn implies

$$f(x;\theta_k) \to f(x;\theta_0)$$

pointwise almost everywhere. Then weak neighbourhoods of $f_0(\cdot) \equiv f(\cdot;\theta_0)$ are equivalent to Hellinger neighbourhoods and so for every $\epsilon > 0$ there exists a $\delta > 0$ such that $\Pi(V_{\delta,\epsilon}) = 0$.

4. Discrete model. Here we consider the case when observations take values in a countable set, e.g. $\mathbb{X} = \{1, 2, 3, \ldots\}$. Denote by $f^{(k)}$ the random mass assigned to the integer k by f. Suppose Π is concentrated on all discrete probability distributions on $\mathbb{X}$. Let f_0 be any distribution in the support of Π and indicate with $f_0^{(k)}$ the true mass assigned to k. Then P_{f_n} converges weakly to P_{f_0} if and only if $f_n^{(k)} \to f_0^{(k)}$ for all k, which also implies that f_n converges in L_1 to f_0.

3.2. Consistency with $\Pi(V_{\delta,\epsilon}) > 0$**.** The general case we consider in the present subsection is the one which involves more technical issues to deal with. As mentioned in the introduction, it has been the focus of many papers in the literature, wherein conditions on the prior are specified in terms of metric entropy and do not admit an easy interpretation. The first result we provide allows for a natural identification of the data tracking behavior. Indeed it shows that the posterior mass concentrated on densities that are away from f_0 in a Hellinger sense and do not track the data vanishes as the sample size increases with $[P_0^\infty]$ probability 1.

Define the data tracking set as a random set of the type $B_{n,\gamma}^c := \{f : R_n(f) \geq \mathrm{e}^{n\gamma}\}$, for any $\gamma > 0$. Recall that $A_\epsilon^c = \{f : d_H(f_0, f) > \epsilon\}$ and $R_n(f) = \prod_{i=1}^n f(X_i)/f_0(X_i)$.

Theorem 2 *Let f_0 be in the Kullback–Leibler support of Π. Then we have*

$$\Pi_n(A_\epsilon^c \cap B_{n,\gamma}) \to 0 \qquad a.s.[P_0^\infty]$$

for any $\gamma < -2\log(1-\epsilon)$.

Proof. Note that for any f in $B_{n,\gamma}$ one has $R_n(f)\,\mathrm{e}^{-n\gamma} < 1$ which yields

$$\mathrm{e}^{-n\gamma} \int_{B_{n,\gamma}\cap A_\epsilon^c} R_n(f)\,\Pi(\mathrm{d}f) < \mathrm{e}^{-\frac{n\gamma}{2}} \int_{B_{n,\gamma}\cap A_\epsilon^c} [R_n(f)]^{\frac{1}{2}}\,\Pi(\mathrm{d}f). \qquad (2)$$

If we observe that

$$\mathrm{E}_0\left\{\int_{B_{n,\gamma}\cap A_\epsilon^c} [R_n(f)]^{1/2}\,\Pi(\mathrm{d}f)\right\} < (1-\epsilon)^n\,\Pi(A_\epsilon^c),$$

where E_0 denotes the expected value with respect to P_0^∞, and apply the Markov inequality, then

$$P_0^\infty\left\{\int_{B_{n,\gamma}\cap A_\epsilon^c} [R_n(f)]^{1/2}\,\Pi(\mathrm{d}f) > \mathrm{e}^{-n\delta}\right\} \leq \Pi(A_\epsilon^c)\,\mathrm{e}^{-n(\log(1-\epsilon)-\delta)}.$$

where $\delta > 0$ is chosen in such a way that $\gamma/2 < \delta < -\log(1-\epsilon)$. Hence, the Borel–Cantelli lemma leads to

$$P_0^\infty\left(\bigcup_{N\geq 1}\bigcap_{n\geq N}\left\{\int_{B_{n,\gamma}\cap A_\epsilon^c}[R_n(f)]^{1/2}\,\Pi(\mathrm{d}f)\leq \mathrm{e}^{-n\delta}\right\}\right)=1$$

which, combined with (2), implies that for all but a finite number of n's

$$\int_{B_{n,\gamma}\cap A_\epsilon^c} R_n(f)\,\Pi(\mathrm{d}f) < \exp\left\{-n\left(\delta-\frac{\gamma}{2}\right)\right\} \qquad \text{a.s.}[P_0^\infty]$$

Since $f_0 \in S_K(\Pi)$, one has that, for any $\beta > 0$ and for all but a finite number of n's,

$$I_n = \int_{\mathbb{F}} R_n(f)\,\Pi(\mathrm{d}f) > \mathrm{e}^{-n\beta} \qquad \text{a.s.}[P_0^\infty]$$

If we fix $\beta < \delta - \gamma/2$, then $\Pi_n(B_{n,\gamma}\cap A_\epsilon^c) \to 0$. □

By Theorem 2 problems might arise because of the sets $A_\epsilon^c \cap B_{n,\gamma}^c$ and we are, then, interested in finding sufficient conditions for which

$$\Pi_n(A_\epsilon^c \cap B_{n,\gamma}^c) \to 0 \qquad \text{a.s.}[P_0^\infty]. \tag{3}$$

Let us first focus on a prior Π concentrating masses $\Pi_1, \Pi_2, \ldots$ on at most a countable number of densities such that $\sum \Pi_k = 1$. Note that (3) is equivalent to

$$J_n = \sum_{\{k: f_k \in A_\epsilon^c\cap B_{n,\gamma}^c\}} R_{nk}\,\Pi_k < \exp(-n\delta) \qquad \text{a.s.}[P_0^\infty]$$

for all large n for some $\delta > 0$, having denoted by $R_{nk} = R_n(f_k)$. Since

$$I_n \geq J_n > \Pi(A_\epsilon^c \cap B_{n,\gamma}^c)\,\exp(n\gamma)$$

where $I_n := \sum_k R_{nk}\,\Pi_k < \exp(n\beta)$ (almost surely) for all large n for any $\beta > 0$, we have that

$$\Pi(A_\epsilon^c \cap B_{n,\gamma}^c) < \exp(-n\eta) \qquad \text{a.s.}[P_0^\infty]$$

for all large n for any $\eta < \gamma - \beta$, where we can fix $\beta < \gamma$. Consequently, the Cauchy-Schwarz inequality yields

$$J_n = \sum_{\{k: f_k\in A_\epsilon^c\}} \mathbb{I}_{B_{n,\gamma}^c}(f_k)\,R_{nk}\,\Pi_k \leq \sum_{\{k: f_k\in A_\epsilon^c\}} \left\{R_{nk}^2\,\Pi_k\right\}^{1/2}\,\left\{\Pi(A_\epsilon^c\cap B_{n,\gamma}^c)\right\}^{1/2}.$$

Since

$$\sum_{\{k: f_k\in A_\epsilon^c\}} R_{nk}^2\,\Pi_k < \left(\sum_k R_{nk}\,\Pi_k^{1/2}\right)^2$$

a sufficient condition for (3) to hold true is

$$\sum_{\{k: f_k \in A_\epsilon^c\}} R_{nk} \, \Pi_k^{1/2} < \exp(n\eta') \qquad \text{a.s.}[P_0^\infty]$$

for all large n. Recall that $R_{nk} < \exp(n\eta')$ almost surely, with respect to P_0^∞, for all large n and for all $\eta' > 0$. Thus we can conclude that $\sum_k \Pi_k^{1/2} < \infty$ is sufficient for consistency. See Walker (2004) for different derivations of this result. As a matter of fact, in Walker (2004) it has been shown that a similar condition is sufficient in a more general setting as well.

Let us now consider a general prior Π, not necessarily discrete. Let f_0 be fixed and take A_ϵ^c to be the complement of an ϵ–Hellinger neighborhood of f_0. By separability of $\mathbb{F}$, such a set can be covered by a countable union of disjoint sets B_j, where $B_j \subseteq B_j^* := \{f : d_H(f, f_j) < \eta\}$, f_j are densities in A_ϵ^c and η is any number in $(0, \epsilon)$. If f_0 is in the Kullback-Leibler support of the prior Π and

$$\sum_{j \geq 1} \sqrt{\Pi(B_j)} < +\infty$$

then Π is Hellinger–consistent at f_0. By virtue of the arguments illustrated at the beginning of the present section, this result can be refined by confining oneself to the determination of a covering of $V_{\delta,\epsilon} \subseteq A_\epsilon^c$. Moreover, by mimicking the proof in Walker (2004), one can state that Hellinger–consistency holds true at $f_0 \in S_K(\Pi)$ if, for some $\alpha \in (0, 1)$,

$$\sum_{j \geq 1} \Pi^\alpha(V_j) < +\infty \tag{4}$$

where the sets V_j have diameter $\eta < \epsilon$ and form a (countable) partition of $V_{\delta,\epsilon}$.

At this stage, one might wonder whether (4) is also necessary for consistency to hold true. The answer to such a question is, in general, negative and can be motivated by an argument which shows that violation of (4) does not imply inconsistency. Assume that Π is not consistent at $f_0 \in S_K(\Pi)$ and that, for some $\epsilon > 0$, $A_\epsilon \subset S_H(\Pi)$. Hence, there exists α in $(0, 1)$ such that

$$\sum_{j \geq 1} \Pi^\alpha(V_j) = +\infty$$

for any covering of $V_{\delta,\epsilon}$. Now take $\tilde{f}$ in $A_{\epsilon/2}$ and denote by $\tilde{V}_j$ the disjoint sets of diameter $\eta < \epsilon/2$ by means of which $\tilde{V}_{\delta,\epsilon/2}$ can be covered where $\tilde{V}_{\delta,\epsilon/2} = \tilde{W}_\delta \cap \tilde{A}_{\epsilon/2}$, $\tilde{W}_\delta$ and $\tilde{A}_{\epsilon/2}$ being, respectively, a weak and Hellinger neighbourhood of $\tilde{f}$. Note that $\tilde{V}_{\delta,\epsilon/2} \supseteq V_{\delta,\epsilon}$ and that any covering of $V_{\delta,\epsilon}$ can be extended to a covering of $\tilde{V}_{\delta,\epsilon/2}$. Thus

$$\sum_{j \geq 1} \Pi^\alpha(\tilde{V}_j) = +\infty$$

must hold. Since $\tilde{f}$ is arbitrary, consistency would fail at each density in $A_{\epsilon/2}$ thus contradicting Doob's theorem. See Lijoi et al. (2004). Hence, Π cannot be inconsistent at all densities in $A_{\epsilon/2}$, even if for each such densities the series of the Π^{α}–probabilities diverges, for some α in $(0, 1)$.

One can alternatively face the issue of establishing the validity of (3), and thus of consistency relying upon the construction of a suitable covering of the random set $B_{n,\gamma}^c$. In the following, $(c_k)_{k\geq 1}$ is an increasing sequence of positive numbers such that $0 < \sup_k(c_{k+1} - c_k) = c^* < +\infty$ and, given $\eta > 0$, we set $\delta \geq c^* + \eta$. Moreover, let

$$C_{n,k} := \{f : \mathrm{e}^{nc_k} \leq R_n(f) < \mathrm{e}^{nc_{k+1}}\}.$$

Theorem 3 *Let f_0 be in the Kullback–Leibler support of Π. Assume that for all $k \geq 1$ there exists a positive integer $n_0 = n_0(k)$ and $\xi_k > 0$, with $\sum \xi_k < +\infty$, such that for all $n \geq n_0$,*

$$P_0^{\infty}\{\Pi(C_{n,k}) < \xi_k \exp\{-n(\delta + c_k)\}\} = 1,$$

for some sequence $(c_k)_{k\geq 1}$ of the type defined above and $\delta \geq c^* + \eta$. Then Π is Hellinger–consistent at f_0.

Proof. Let

$$\frac{\int_{A_\epsilon^c} R_n(f)\,\Pi(\mathrm{d}f)}{\int_{\mathbb{F}} R_n(f)\,\Pi(\mathrm{d}f)} = \Pi_n(A_\epsilon^c) = \Pi_n(A_\epsilon^c \cap B_{n,\gamma}) + \Pi_n(A_\epsilon^c \cap B_{n,\gamma}^c).$$

By Theorem 2, the first summand tend to 0, almost surely with respect to P_0^{∞}. Thus, we focus attention on $\Pi_n(A_\epsilon^c \cap B_{n,\gamma}^c)$. If $(c_k)_{k\geq 1}$ is the sequence described above, with $c_1 = \gamma$, then

$$\Pi_n(A_\epsilon^c \cap B_{n,\gamma}^c) = \frac{J_n}{I_n} = \frac{\sum_k \int_{C_{n,k}\cap A_\epsilon^c} R_n(f)\,\Pi(\mathrm{d}f)}{\int_{\mathbb{F}} R_n(f)\,\Pi(\mathrm{d}f)}$$

and

$$I_n > J_n > \sum_k \mathrm{e}^{nc_k}\,\Pi(C_{n,k}).$$

According to Lemma 1 in Barron et al. (1999), the fact that f_0 is in $S_K(\Pi)$ implies that, for each n, $I_n \notin \{0, \infty\}$ almost surely. By virtue of the hypothesis on Π, one has

$$J_n < \sum_k \mathrm{e}^{nc_{k+1}}\,\Pi(C_{n,k}) < \sum_k \xi_k\,\mathrm{e}^{-n(\delta - c_{k+1} + c_k)} < \mathrm{e}^{-n\eta}\sum_k \xi_k \qquad \text{a.s.}[P_0^{\infty}].$$

Hence, since $I_n < \exp(n\beta)$ for any $\beta > 0$, choosing $\beta < \eta$ leads to $\Pi_n(A_\epsilon^c \cap B_{n,\gamma}^c) \to 0$, a.s.$[P_0^{\infty}]$. □

Special attention is required for Π, based on knowledge of f_0, in order to contradict the assumption of the theorem; bearing in mind that

$$\Pi(C_{n,k}) < \exp(-nc_k)$$

a.s. and

$$\sum_k \exp(-nc_k)\,\Pi(C_{n,k}) \to 0$$

for all choices of $\{c_k\}$.

It is therefore interesting to investigate why the prior suggested in the counter–example of Barron et al. (1999) does not meet the condition given in the above Theorem 3. Their prior assigns positive masses to single densities for which $R_n(f) = 2^n$, with f_0 being the uniform density on $[0,1]$, thus explaining the phenomenon of tracking the data. More precisely, for any sample $X_1,\ldots,X_n$ and $\gamma < \log 2$, one has that $\Pi(B^c_{n,\gamma}) \geq \exp(-2)2^{-n}/(2c_0 n)$ where c_0 is some positive constant. Let, for any sequence $(c_k)_{k\geq 1}$ defined as above, $k^* = k^*(n)$ be such that $2^n \in C_{n,k^*}$. This means that $\Pi(C_{n,k^*}) = \Pi(B^c_{n,\gamma})$ and $\exp(-nc_{k^*+1}) > 2^{-n}$. Hence, for any $\delta \geq c^* + \eta$

$$\exp(-n\delta - nc_{k^*}) < \exp(-n\eta)\,2^{-n}$$

and, since $\eta > 0$, the following inequality must hold true for all n large enough

$$\Pi(C_{n,k^*}) \geq \frac{\exp(-2)2^{-n}}{2c_0 n} > \exp(-n\eta)\,2^{-n} > \xi_{k^*}\exp\{-n(\delta + c_{k^*})\}$$

for any sample $X_1,\ldots,X_n$.

4. Connecting ideas.

The purpose of this section is to bring together the various pieces in the case when $\Pi(V_{\delta,\epsilon}) > 0$ and to further understand (4), assuming without loss of generality, $\alpha = 1/2$. It provides an understanding of the sets $V_{\delta,\epsilon}$.

For (4) to be satisfied with $\alpha = 1/2$ we can, when the prior Π does not put mass on single densities, achieve Hellinger neighbourhoods of size no greater than $\epsilon > 0$ from a dense set $\{f_k\}$, say $B_k = N_{e_k}(f_k)$, such that $e_k \leq \epsilon$ for all k and $\Pi(N_{e_k}(f_k)) \leq M/k^{2+r}$ for some finite M and $r > 0$. We can pick the $\{e_k\}$ to make this hold; if $\Pi(N_\epsilon(f_k)) < M/k^{2+r}$ then we take $e_k = \epsilon$ else we take $e_k < \epsilon$ such that $\Pi(N_{e_k}(f_k)) = M/k^{2+r}$.

Hence, with this we have

$$\sum_k \sqrt{\Pi(B_k)} < +\infty$$

ensuring that

$$\Pi_n\left(\bigcup_k B_{k,M} \setminus N_\epsilon(f_0)\right) \to 0 \ \text{ a.s.}$$

We may not have however that $\bigcup_k B_k$ covers the space of densities $\mathbb{F}$. To investigate what may be left out, consider $\mathbb{F}^* = \bigcup_k B_k$ and let $S = \mathbb{F} \setminus \mathbb{F}^*$. We can state immediately that if it turns out to be that $e_k > e > 0$ for all large k then we have that $S = \emptyset$ and consistency holds. If this is not the case, then we must have that S is closed since $\mathbb{F}^*$ is open. Consequently, S is *nowhere dense*, since $\mathbb{F} \setminus S$ is dense in $\mathbb{F}$. So S is where the posterior could put mass; a nowhere dense, closed (and so has empty interior) subset of $\mathbb{F}$. For inconsistency to occur it must be that $S \bigcap V_{\delta,\epsilon} \neq \emptyset$ for all $\delta > 0$ and $\epsilon > 0$; that is S must contain a sequence of densities which converge weakly to f_0 but not in a strong sense.

Let us now use subscript M to denote the dependence on M, which can be arbitrarily large. We will establish that $\Pi(S_M) \to 0$ as $M \to \infty$. Now, for any $\delta > 0$ and any $\epsilon > 0$ there exists an $L < +\infty$ such that

$$\Pi\left(\bigcup_{k=1}^{L} N_\epsilon(f_k)\right) > 1 - \delta.$$

If we choose M big enough so that $e_k > \epsilon$ for all $k = 1, \ldots, L$, then clearly

$$\Pi\left(\bigcup_{k} N_{e_k}(f_k)\right) > 1 - \delta$$

as well. We can do this by taking M such that

$$M/L^{2+r} > \max_{k \in 1 \ldots L} \Pi\left(N_\epsilon(f_k)\right)$$

and note that

$$\Pi\left(N_{e_k}(f_k)\right) \geq M/L^{2+r}$$

for all $k = 1, \ldots L$. Therefore, $\Pi(\mathbb{F}^*_M) \to 1$ as $M \to \infty$. In fact, $S_M \downarrow S'$ for some set S' with $\Pi(S') = 0$ and

$$S' = \bigcap_{M} S_M.$$

So we have that, for inconsistency, Π_n must put mass into $\Delta_M = S_M \setminus S'$, for each M, and $\Delta_M \downarrow \emptyset$.

Summarising, we know that for inconsistency the posterior must put mass into a subset of $V_{\delta,\epsilon}$. Now we know what this subset is like; it is closed, nowhere dense and, based on the arbitrariness of M, it can be made arbitrarily close to the empty set.

5. Summary.

In most cases, such as parametric models and many nonparametric ones, it is that $\Pi(V_{\delta,\epsilon}) = 0$ for small enough δ and ϵ. If not, for inconsistency a number of "unnatural" connections between f_0 and Π need to exist, the most startling of which is that

$$\sum_j \Pi^\alpha(\tilde{V}_j) = +\infty$$

for all $\alpha < 1$. When inconsistency does occur, we have identified precisely where the posterior is putting mass, namely a nowhere dense, closed subset of $V_{\delta,\epsilon}$ for all $\delta > 0$ and $\epsilon > 0$.

Acknowledgments. A. Lijoi and I. Prünster were partially supported by the Italian Ministry of University and Research (MIUR), research project "Bayesian nonparametric methods and their applications". The work of S.G. Walker is financed by an EPSRC Advanced Research Fellowship.

Appendix

In order to prove Theorem 1 we need to introduce

$$d_{K,n}(f_0, f) = \frac{1}{n}\sum_{i=1}^{n} \log\{f_0(X_i)/f(X_i)\}$$

to be the sample Kullback-Leibler divergence between f_0 and f.

Proof of Theorem 1

We let $I_n^{(j)} = \int R_n^{(j)}(f)\,\Pi(\mathrm{d}f)$, where

$$R_n^{(j)}(f) = \prod_{i=1}^{n} \frac{f(X_i^{(j)})}{f_0(X_i^{(j)})} \qquad j = 1, 2.$$

Since f_0 is in the Kullback-Leibler support of the prior, from Schwartz (1965) one has

$$\frac{1}{n}\log I_n \to 0 \qquad \text{a.s.}[P_0^\infty];$$

and, from Barron (1988), $E_0(n^{-1}\log I_n) \to 0$ a.s.$[P_0^\infty]$. Moreover, the following identity

$$-\frac{1}{n}\log I_n = \frac{1}{n}D_K(\mu, \Pi) - \frac{1}{n}D_K(\mu, \Pi_n) + \int d_{K,n}(f_0, f)\,\mu(\mathrm{d}f) \tag{5}$$

holds true for any measure μ which is absolutely continuous with respect to Π. Indeed, we have

$$\begin{aligned} -\log I_n &= \int \log\left\{\frac{R_n(f)\,\Pi(\mathrm{d}f)}{I_n\,\Pi(\mathrm{d}f)}\right\}\mu(\mathrm{d}f) - \int \log R_n(f)\,\mu(\mathrm{d}f) \\ &= \int \log\left(\mathrm{d}\Pi_n/\mathrm{d}\Pi\right)\,\mathrm{d}\mu + n\int d_{K,n}(f_0, f)\,\mu(\mathrm{d}f). \end{aligned}$$

If in (5) we set $\Pi_n = \Pi_n^{(1)}$ and $\mu = \Pi_n^{(2)}$, we have

$$-\frac{1}{n}\log I_n^{(1)} = \frac{1}{n}D_K(\Pi_n^{(2)}, \Pi) - \frac{1}{n}D_K(\Pi_n^{(2)}, \Pi_n^{(1)}) + \int d_{K,n}^{(1)}(f_0, f)\,\Pi_n^{(2)}(\mathrm{d}f),$$

where $d_{K,n}^{(1)}(f_0,f) := (1/n)\sum_{i=1}^n \log(f_0(X_i^{(1)})/f(X_i^{(1)}))$. Let us now take expectations with respect to $X_1^{(1)},\ldots,X_n^{(1)}$ and keep $X_1^{(2)},\ldots,X_n^{(2)}$ fixed. Then, using the fact that $E_0^{(1)}(n^{-1}\log I_n^{(1)}) \to 0$, we have

$$E_0^{(1)}\left\{\frac{1}{n}D_K(\Pi_n^{(2)},\Pi_n^{(1)})\right\} \sim \frac{1}{n}D_K(\Pi_n^{(2)},\Pi) + \int d_K(f_0,f)\,\Pi_n^{(2)}(\mathrm{d}f).$$

Finally, the inconsistency of Π, applied to $\Pi_n^{(2)}$ yields

$$\limsup_n \int d_K(f_0,f)\,\Pi_n^{(2)}(\mathrm{d}f) > 0 \quad \text{a.s.}[P_0^\infty]$$

and so

$$\limsup_n E_0^{(1)}\left\{\frac{1}{n}D_K(\Pi_n^{(2)},\Pi_n^{(1)})\right\} > 0 \qquad \text{a.s.}[P_0^\infty]$$

and the result follows. □

REFERENCES

Barron, A., Schervish, M.J. & Wasserman, L. (1999). The consistency of distributions in nonparametric problems. *Ann. Statist.* **27**, 536–561.

Blackwell, D. & Dubins, L. (1962). Merging of opinions with increasing information. *Ann. Math. Statist.* **33**, 882–886.

Diaconis, P. & Freedman, D. (1986). On the consistency of Bayes estimates. *Ann. Statist.* **14**, 1–26.

Doob, J.L. (1949). Application of the theory of martingales. In *Le Calcul des Probabilités et ses Applications* 23–27. Colloques Internationaux du Centre National de la Recherche Scientifique, Paris.

Freedman, D.A. (1963). On the asymptotic behavior of Bayes' estimates in the discrete case. *Ann. Math. Statist.* **34**, 1386–1403.

Ghosal, S., Ghosh, J.K. & Ramamoorthi, R.V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.* **27**, 143–158.

Hansen, M.B. & Lauritzen, S.L. (2002). Nonparametric Bayes inference for concave distribution functions. *Statist. Neerlandica* **56**, 110–127.

Lijoi, A., Prünster, I. & Walker, S.G. (2004). Extending Doob's consistency theorem to nonparametric densities. *Bernoulli* **10**, 651–663.

Petrone, S. & Wasserman, L. (2002). Consistency of Bernstein polynomial posteriors. *J. Roy. Stat. Soc. B* **64**, 79–100.

Schwartz, L. (1965). On Bayes procedures. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **4**, 10–26.

Walker, S.G. (2003). On sufficient conditions for Bayesian consistency. *Biometrika* **90**, 482–488.

Walker, S.G. (2004). New approaches to Bayesian consistency. *Ann. Statist.* **32**, 2028–2043.

1. Luigi Montrucchio and Fabio Privileggi, "On Fragility of Bubbles in Equilibrium Asset Pricing Models of Lucas-Type," *Journal of Economic Theory* 101, 158-188, 2001 (ICER WP 2001/5).
2. Massimo Marinacci, "Probabilistic Sophistication and Multiple Priors," *Econometrica* 70, 755-764, 2002 (ICER WP 2001/8).
3. Massimo Marinacci and Luigi Montrucchio, "Subcalculus for Set Functions and Cores of TU Games," *Journal of Mathematical Economics* 39, 1-25, 2003 (ICER WP 2001/9).
4. Juan Dubra, Fabio Maccheroni, and Efe Ok, "Expected Utility Theory without the Completeness Axiom," *Journal of Economic Theory*, forthcoming (ICER WP 2001/11).
5. Adriana Castaldo and Massimo Marinacci, "Random Correspondences as Bundles of Random Variables," April 2001 (ICER WP 2001/12).
6. Paolo Ghirardato, Fabio Maccheroni, Massimo Marinacci, and Marciano Siniscalchi, "A Subjective Spin on Roulette Wheels," *Econometrica*, forthcoming (ICER WP 2001/17).
7. Domenico Menicucci, "Optimal Two-Object Auctions with Synergies," July 2001 (ICER WP 2001/18).
8. Paolo Ghirardato and Massimo Marinacci, "Risk, Ambiguity, and the Separation of Tastes and Beliefs," *Mathematics of Operations Research* 26, 864-890, 2001 (ICER WP 2001/21).
9. Andrea Roncoroni, "Change of Numeraire for Affine Arbitrage Pricing Models Driven By Multifactor Market Point Processes," September 2001 (ICER WP 2001/22).
10. Maitreesh Ghatak, Massimo Morelli, and Tomas Sjoström, "Credit Rationing, Wealth Inequality, and Allocation of Talent", September 2001 (ICER WP 2001/23).
11. Fabio Maccheroni and William H. Ruckle, "BV as a Dual Space," *Rendiconti del Seminario Matematico dell'Università di Padova*, 107, 101-109, 2002 (ICER WP 2001/29).
12. Fabio Maccheroni, "Yaari Dual Theory without the Completeness Axiom," *Economic Theory*, forthcoming (ICER WP 2001/30).
13. Umberto Cherubini and Elisa Luciano, "Multivariate Option Pricing with Copulas," January 2002 (ICER WP 2002/5).
14. Umberto Cherubini and Elisa Luciano, "Pricing Vulnerable Options with Copulas," January 2002 (ICER WP 2002/6).
15. Steven Haberman and Elena Vigna, "Optimal Investment Strategies and Risk Measures in Defined Contribution Pension Schemes," *Insurance: Mathematics and Economics* 31, 35-69, 2002 (ICER WP 2002/10).
16. Enrico Diecidue and Fabio Maccheroni, "Coherence without Additivity," *Journal of Mathematical Psychology*, forthcoming (ICER WP 2002/11).
17. Paolo Ghirardato, Fabio Maccheroni, and Massimo Marinacci, "Ambiguity from the Differential Viewpoint," April 2002 (ICER WP 2002/17).

18. Massimo Marinacci and Luigi Montrucchio, "A Characterization of the Core of Convex Games through Gateaux Derivatives," *Journal of Economic Theory*, forthcoming (ICER WP 2002/18).
19. Fabio Maccheroni and Massimo Marinacci, "How to Cut a Pizza Fairly: Fair Division with Decreasing Marginal Evaluations," *Social Choice and Welfare*, 20, 457-465, 2003 (ICER WP 2002/23).
20. Erio Castagnoli, Fabio Maccheroni and Massimo Marinacci, "Insurance Premia Consistent with the Market," *Insurance: Mathematics and Economics* 31, 267-284, 2002 (ICER WP 2002/24).
21. Fabio Privileggi and Guido Cozzi, "Wealth Polarization and Pulverization in Fractal Societies," September 2002 (ICER WP 2002/39).
22. Paolo Ghirardato, Fabio Maccheroni, and Massimo Marinacci, "Certainty Independence and the Separation of Utility and Beliefs," December 2002 (ICER WP 2002/40).
23. Salvatore Modica and Marco Scarsini, "The Convexity-Cone Approach to Comparative Risk and Downside Risk", January 2003 (ICER WP 2003/1).
24. Claudio Mattalia, "Existence of Solutions and Asset Pricing Bubbles in General Equilibrium Models", January 2003 (ICER WP 2003/2).
25. Massimo Marinacci and Luigi Montrucchio, "Cores and Stable Sets of Finite Dimensional Games", March 2003 (ICER WP 2003/7).
26. Jerome Renault, Sergio Scarlatti, and Marco Scarsini, "A Folk Theorem for Minority Games", April 2003 (ICER WP 2003/10).
27. Peter Klibanoff, Massimo Marinacci, and Sujoy Mukerji, "A Smooth Model of Decision Making under Ambiguity", April 2003 (ICER WP 2003/11).
28. Massimo Marinacci and Luigi Montrucchio, "Ultramodular Functions", June 2003 (ICER WP 2003/13).
29. Erio Castagnoli, Fabio Maccheroni, and Massimo Marinacci, "Choquet Insurance Pricing: a Caveat", June 2003 (ICER WP 2003/14).
30. Thibault Gajdos and Eric Maurin, "Unequal Uncertainties and Uncertain Inequalities: an Axiomatic Approach, June 2003 (ICER WP 2003/15).
31. Thibault Gajdos and John A. Weymark, "Multidimensional Generalized Gini Indices", June 2003 (ICER WP 2003/16).
32. Thibault Gajdos, Jean-Marc Tallon, and Jean-Christophe Vergnaud, "Decision Making with Imprecise Probabilistic Information", June 2003 (ICER WP 2003/18).
33. Alfred Müller and Marco Scarsini, "Archimedean Copulae and Positive Dependence", July 2003 (ICER WP 2003/25).
34. Bruno Bassan, Olivier Gossner, Marco Scarsini, and Shmuel Zamir, "Positive Value of Information in Games", *International Journal of Game Theory*, forthcoming (ICER WP 2003/26).
35. Marco Dall'Aglio and Marco Scarsini, "Zonoids, Linear Dependence, and Size-Biased Distributions on the Simplex", *Advances in Applied Probability* forthcoming (ICER WP 2003/27).
36. Taizhong Hu, Alfred Muller, and Marco Scarsini, "Some Counterexamples in Positive Dependence", *Journal of Statistical Planning and Inference*, forthcoming (ICER WP 2003/28).
37. Massimo Marinacci, Fabio Maccheroni, Alain Chateauneuf, and Jean-Marc Tallon, "Monotone Continuous Multiple Priors" August 2003 (ICER WP 2003/30).

38. Massimo Marinacci and Luigi Montrucchio, "Introduction to the Mathematics of Ambiguity", October 2003 (ICER WP 2003/34).
39. Fabio Maccheroni, Massimo Marinacci and Aldo Rustichini, "Variational representation of preferences under ambiguity", June 2004 (ICER WP 2004/5).
40. Antonio Lijoi, Ramsés H. Mena, and Igor Prünster, "Hierarchical mixture modelling with normalized Gaussian priors", November 2004 (ICER WP 2004/12).
41. Stephen G. Walker, Antonio Lijoi, and Igor Prünster, "Contributions to the understanding of Bayesian consistency", November 2004 (ICER WP 2004/13).
42. Antonio Lijoi, Igor Prünster, and Stephen G. Walker, "On consistency of nonparametric normal mixtures", October 2004 (ICER WP 2004/23).
43. Stephen G. Walker, Antonio Lijoi, and Igor Prünster, "On rates of convergence for posterior distributions", October 2004 (ICER WP 2004/24).