Quaderni di Statistica
Vol. 2, 2000

On parametric models for invariant probability measures

Patrizia Berti
Dipartimento di Matematica Pura ed Applicata "G. Vitali", Università di Modena
E-mail: berti.patrizia@unimo.it
Sandra Fortini
Istituto di Metodi Quantitativi, Università "L. Bocconi"
E-mail: sandra.fortini@uni-bocconi.it
Lucia Ladelli
Dipartimento di Matematica, Politecnico di Milano
E-mail: lucia.ladelli@mate.polimi.it
Eugenio Regazzini
Dipartimento di Matematica "F. Casorati", Università di Pavia
E-mail: eugenio@dimat.unipv.it
Pietro Rigo
Dipartimento di Statistica "G. Parenti", Università di Firenze
E-mail: rigo@ds.unifi.it

Summary: Let $\widetilde{x} = (\widetilde{x}_n : n \in \mathbb{N})$ be a sequence of random variables with values in a Polish space X. If $\widetilde{x}$ is exchangeable (stationary), then $\widetilde{x}$ is i.i.d. (ergodic) conditionally on some random probability measure $\widetilde{p}$ on $\mathcal{B}(X)$ ($\widetilde{q}$ on $\mathcal{B}(X^\infty)$). In the exchangeable case, two necessary and sufficient conditions are given for $\widetilde{x}$ to be i.i.d. conditionally on $\widetilde{t}(\widetilde{p})$, where $\widetilde{t}$ is a given transformation. Essentially the same conditions apply when $\widetilde{x}$ is stationary, apart from "i.i.d." is replaced by "ergodic" and $\widetilde{p}$ by $\widetilde{q}$. The basic difference between the exchangeable and the stationary case lies in the empirical measure to be used. Up to a proper choice of the latter, the two conditions work in a large class of invariant distributions for $\widetilde{x}$. Finally, the set of ergodic probability measures is shown to be a Borel set, and almost sure weak convergence of a certain sequence of empirical measures is proved in the stationary case.

Keywords: Empirical distribution, Exchangeability, Invariant probability measure, Predictive distribution, Predictive sufficient statistic, Stationarity.

1. Introduction

Let $\tilde{x} = (\tilde{x}_n : n \in \mathbb{N})$ be a sequence of random variables with values in a Polish space X. Suppose $\tilde{x}$ is exchangeable or, equivalently, $\tilde{x}$ is independent and identically distributed (i.i.d.) conditionally on some random probability measure $\tilde{p}$ on the Borel σ-field on X, $\mathcal{B}(X)$. In Theorem (7.2) of Fortini, Ladelli and Regazzini (2000) – referred to as FLR in the sequel – conditions are given for $\tilde{x}$ to be i.i.d. conditionally on $\tilde{t}(\tilde{p})$, where $\tilde{t}$ is a given transformation. In fact, the hypotheses of Theorem (7.2) concern sequences of predictive sufficient statistics, and are conceived to have some statistical content within a Bayesian predictive framework.

In so far as the *sole* representation part is concerned – i.e., finding conditions under which $\tilde{x}$ is i.i.d. given $\tilde{t}(\tilde{p})$, where $\tilde{t}$ is an assigned transformation – the hypotheses of Theorem (7.2) are sufficient but not necessary.

The aim of this paper is to take up again the representation part to provide necessary and sufficient conditions. From a statistical point of view, such conditions are useful for investigating existence of underlying parametric models for the distribution of $\tilde{x}$.

In particular, two conditions are given. The first one (that will be denoted as condition (b)) is of the abstract type. Nevertheless, by using it, various other sufficient conditions are easily obtained, included the one proposed in FLR. Moreover, it allows a shorter and more direct proof of Theorem (7.2). The second condition (condition (c)) applies to the special case where $\tilde{t}$ is continuous (an assumption made also in Theorem (7.2)). In this case, however, it has some statistical meaning.

The problem studied in this paper is different from the one tacled in Olshen (1974). In the latter, among other things, an exchangeable sequence $\tilde{x}$ is shown to be i.i.d. conditionally on *some* real random variable M. Such result, clearly, is based on the Borel isomorphism theorem. In our case, instead, we investigate whether $\tilde{x}$ is i.i.d. conditionally on $\tilde{t}(\tilde{p})$, where $\tilde{t}$ is a *given* transformation (possibly, suggested by a sufficient statistic) and $\tilde{p}$ is the almost sure weak limit of the empirical measures.

A further and minor goal of this paper is to start, in a particular case, an analysis of Bayesian nonparametric problems for non exchangeable data. Suppose $\tilde{x}$ is stationary or, equivalently, $\tilde{x}$ is ergodic conditionally on some random probability measure $\tilde{q}$ on $\mathcal{B}(X^\infty)$. Then, conditions (b) and (c) are still working, apart from "i.i.d." is replaced by "ergodic" and $\tilde{p}$ by $\tilde{q}$. In this framework, incidentally, the set of ergodic probability measures is shown to be a Borel set, and almost sure weak convergence of a certain sequence of empirical measures is proved.

The basic difference between the exchangeable and the stationary case lies in the empirical measure to be used. Up to a proper choice of the latter, conditions (b) and (c) work for a large class of invariant distributions for $\tilde{x}$.

The paper is organized as follows. Section 2 is devoted to preliminaries and notation, Section 3 includes the statement of Theorem (7.2), while Sections 4 and 5 contain the main results, in the exchangeable and stationary case, respectively.

2. Preliminaries and notation

The basic notation is that of FLR, with some slight adaptions in view of the last Section 5.

Let S be any set. Then, S^∞ is the space of all sequences $x = (x_1, x_2, ...)$ of elements of S, and $\tilde{x}_n$ is the n-th coordinate map on S^∞, that is,

$$\tilde{x}_n(x) = x_n \quad \text{for all } x = (x_1, x_2, ...) \in S^\infty \text{ and } n \in \mathbb{N}.$$

When S is a topological space, $\mathcal{B}(S)$ denotes the Borel σ-field on S, and S^∞ is equipped with the product topology. By a *Polish space*, it is meant a topologically complete separable space. If S is Polish, then S^∞ is Polish, too, and $\mathcal{B}(S^\infty) = \mathcal{B}(S)^\infty$.

Given a σ-field $\mathcal{E}$ on S, let $M = M(\mathcal{E})$ be the set of probability measures (p.m.'s) on $\mathcal{E}$, and let $\mathcal{P} = \mathcal{P}(\mathcal{E})$ be the σ-field on M

generated by the sets $\{p \in M : p(A) \in B\}$, for $A \in \mathcal{E}$ and $B \in \mathcal{B}([0,1])$. Any measurable function $f : (\Omega, \mathcal{A}) \to (M, \mathcal{P})$, where $(\Omega, \mathcal{A})$ is any measurable space, is called a random p.m.. Thus, a function f on $(\Omega, \mathcal{A})$ is a random p.m. whenever $f(\omega)$ is a p.m. on $\mathcal{E}$ for $\omega \in \Omega$, and $\omega \mapsto f(\omega)(A)$ is $\mathcal{A}$-measurable for $A \in \mathcal{E}$.

We will consider a sequence of observable random elements taking values in a Polish space X. For our purposes, it is convenient to work in the coordinate space $(X^\infty, \mathcal{B}(X^\infty))$ and to identify the sequence of observables with $\widetilde{x} := (\widetilde{x}_1, \widetilde{x}_2, ...)$. Moreover, we let $M_1 = M(\mathcal{B}(X))$ and $M_2 = M(\mathcal{B}(X^\infty))$, i.e., M_1 and M_2 denote the sets of p.m.'s on $\mathcal{B}(X)$ and on $\mathcal{B}(X^\infty)$, respectively. Both M_1 and M_2 are equipped with the topology of weak convergence of p.m.'s. Since X is Polish, M_1 and M_2 are Polish, too, and $\mathcal{B}(M_1)$ and $\mathcal{B}(M_2)$ coincide with the σ-fields $\mathcal{P}(\mathcal{B}(X))$ and $\mathcal{P}(\mathcal{B}(X^\infty))$defined earlier.

For every $n \in \mathbb{N}$, the *empirical measure* e_n associated to $(\widetilde{x}_1, ..., \widetilde{x}_n)$ is defined as

$$e_n = \frac{1}{n}\sum_{k=1}^{n}\delta_{\widetilde{x}_k}$$

where δ_a is the unit mass at a. Clearly, $e_n(x)$ is a p.m. on $\mathcal{B}(X)$ for $x \in X^\infty$, and $x \mapsto e_n(x)(A)$ is Borel measurable for $A \in \mathcal{B}(X)$, i.e., e_n is a random p.m. from $(X^\infty, \mathcal{B}(X^\infty))$ into $(M_1, \mathcal{B}(M_1))$.

A p.m. P on $\mathcal{B}(X^\infty)$ is *exchangeable* if it is invariant under all finite permutations of X^∞, that is, if $P = P \circ \pi^{-1}$ for all functions $\pi : X^\infty \to X^\infty$ of the form

$$\pi(x_1, ..., x_n, x_{n+1}, ...) = (x_{j_1}, ..., x_{j_n}, x_{n+1}, ...)$$

for some $n \in \mathbb{N}$ and some permutation $(j_1, ..., j_n)$ of $(1, ..., n)$. As usual, $P \circ \pi^{-1}$ denotes the p.m. on $\mathcal{B}(X^\infty)$ given by $P \circ \pi^{-1}(A) = P(\pi^{-1}(A))$. When P is exchangeable, we will also say that $\widetilde{x} = (\widetilde{x}_1, \widetilde{x}_2, ...)$ is exchangeable under P.

Given $p \in M_1$, let p^∞ denote the corresponding product p.m., i.e., p^∞ is the p.m. on $\mathcal{B}(X^\infty)$ which makes the coordinate random variables $\widetilde{x}_1, \widetilde{x}_2, \ldots$ i.i.d. according to p.

We are now in a position to state de Finetti's representation theorem for (infinite) exchangeable sequences.

Theorem 1 (de Finetti's representation theorem)

Let P be a p.m. *on $\mathcal{B}(X^\infty)$ where X is Polish. Then, the following statements are equivalent:*

(i) *P is exchangeable;*

(ii) *There is a unique* p.m. *ν on $\mathcal{B}(M_1)$ such that*

$$P(A) = \int p^\infty(A)\nu(dp) \quad \textit{for all } A \in \mathcal{B}(X^\infty);$$

(iii) *There is a random* p.m. *$\widetilde{p}$ (i.e., a Borel function $\widetilde{p} : X^\infty \to M_1$) such that,*

for each $A \in \mathcal{B}(X^\infty)$, $\widetilde{p}^\infty(A)$ is a version of $P(\widetilde{x} \in A|\widetilde{p})$.

Moreover, when P is exchangeable, one has $\nu(B) = P(\widetilde{p} \in B)$ for all $B \in \mathcal{B}(M_1)$ and

$$e_n \Rightarrow \widetilde{p} \quad P\text{-a.s.},$$

where " $\Rightarrow$ " stands for weak convergence of p.m.'s.

We close this section with the notion of statistic. By definition, a statistic is a measurable function of the data. In our case, the data are the n-tuple $(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ for some n. Under the exchangeability assumption, the order in which observations appear is not relevant, and $(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ can be summarized by the empirical measure e_n.

Let

$$D = \{e_n(x) : x \in X^\infty, n \in \mathbb{N}\}$$

be the union of the ranges of all the random p.m.'s e_n, and let M_1^* be a Borel subset of M_1 such that $M_1^* \supset D$. Further, let T be a Polish space and $\widetilde{t} : M_1^* \to T$ a Borel function. Then, for each n, $\widetilde{t} \circ e_n$ is Borel measurable on X^∞ and is a summary of the data $(\widetilde{x}_1, ..., \widetilde{x}_n)$. Accordingly, in Sections 3 and 4 a statistic is meant as the restriction to D, $\widetilde{t}|D$, of any Borel function $\widetilde{t} : M_1^* \to T$, where $M_1^* \in \mathcal{B}(M_1)$ and $M_1^* \supset D$.

In last Section 5, a slightly different (but conceptually equivalent) notion of statistic is used.

3. Statement of Theorem (7.2) of FLR

Let P be an exchangeable p.m. on $\mathcal{B}(X^\infty)$ and let $\widetilde{x} = (\widetilde{x}_n : n \in \mathbb{N})$ be the sequence of coordinate random variables on X^∞. According to Theorem 1, conditionally on some random p.m. $\widetilde{p}$, $\widetilde{x}$ is i.i.d. according to $\widetilde{p}$. Fix a Borel function $\widetilde{t} : M_1^* \to T$, where T is Polish, $M_1^* \in \mathcal{B}(M_1)$ and $M_1^* \supset D$, and suppose that $\nu(M_1^*) = 1$ where ν is as in Theorem 1.

Under these conditions one question is whether, conditionally on $\widetilde{t}(\widetilde{p})$, $\widetilde{x}$ is again i.i.d. according to $\widetilde{p}$. Answering this question is useful, for instance, for investigating existence of an underlying parametric model for the distribution of $\widetilde{x}$. Suppose in fact that, conditionally on $\widetilde{t}(\widetilde{p})$, $\widetilde{x}$ is i.i.d. according to $\widetilde{p}$. Then, $\widetilde{p}$ and $\widetilde{t}(\widetilde{p})$ play essentially the same role, and a random parameter $\widetilde{\theta}$ can be defined as $\widetilde{\theta} = \widetilde{t}(\widetilde{p})$. In other terms, the "original random parameter" $\widetilde{p}$, whose existence is granted by de Finetti's theorem, can be reduced through $\widetilde{t}$. As far as $\widetilde{t}$ is concerned, in what follows it should be viewed as any given transformation. However, in a particular statistical problem, $\widetilde{t}$ is typically suggested by some sufficient statistic.

In any case, a positive answer to the earlier question occurs in case

(a) $\widetilde{p} \in M_1^*$ and $g(\widetilde{t}(\widetilde{p})) = \widetilde{p}$, P-a.s.,
for some function $g : T \to M_1$ such that $\sigma(g) \subset \overline{\mathcal{B}}(T)$

where $\overline{\mathcal{B}}(T)$ is the completion of $\mathcal{B}(T)$ with respect to the distribution of $\widetilde{t}\,(\widetilde{p})$.

Indeed, under (a), $\widetilde{x}$ is i.i.d. (according to $\widetilde{p}$) conditionally on $\widetilde{t}\,(\widetilde{p})$. In particular, one has

$$P(A) = \int g(t)^{\infty}(A)\, \overline{\nu^*}(dt) \quad \text{for all } A \in \mathcal{B}(X^{\infty})$$

where ν^* is the distribution of $\widetilde{t}\,(\widetilde{p})$ and $\overline{\nu^*}$ is the completion of ν^*.

Theorem (7.2) of FLR, quoted in Section 1, just gives conditions for (a). To state it, various definitions are needed.

First, let P be any p.m. on $\mathcal{B}(X^{\infty})$. For each $n \in \mathbb{N}$, $P[(\widetilde{x}_j : j > n) \in \cdot \,|\widetilde{x}_1, ..., \widetilde{x}_n]$ is called the *predictive distribution* (at time n). A statistic $\widetilde{t}\,|D$ is *predictive sufficient* if, for each $n \in \mathbb{N}$ and $A \in \mathcal{B}(X^{\infty})$, one has

$$P[(\widetilde{x}_j : j > n) \in A|\, \widetilde{t} \circ e_n] = P[(\widetilde{x}_j : j > n) \in A|\widetilde{x}_1, ..., \widetilde{x}_n] \quad P\text{-a.s.}.$$

So, the informal idea is that $\widetilde{t}\,|D$ is predictive sufficient provided predictive distributions depend on data only through it. We send back to FLR for more on predictive sufficiency and for some references therein.

Next, let $\mathcal{G} \subset \mathcal{B}(X)$ be a countable π-class such that $\sigma(\mathcal{G}) = \mathcal{B}(X)$ and $P_1(\partial A) = 0$ for all $A \in \mathcal{G}$, where P_1 is the marginal of P on the first coordinate.

Further, for each n, fix a function Q_n on $T \times \mathcal{B}(X)$ such that:

– $Q_n(t, \,\cdot\,)$ is a p.m. on $\mathcal{B}(X)$ for $t \in T$,

– $Q_n(\,\cdot\,, A)$ is $\mathcal{B}(T)$-measurable for $A \in \mathcal{B}(X)$,

– $\int_B Q_n(t, A)\mu_n(dt) = P(\widetilde{x}_{n+1} \in A, \widetilde{t} \circ e_n \in B)$

for all $A \in \mathcal{B}(X)$ and $B \in \mathcal{B}(T)$, where μ_n denotes the distribution of $\widetilde{t} \circ e_n$.

Clearly, such Q_n exists since X is Polish.

Finally, consider the following condition; cf. condition (7.1) of FLR.

(*) There is a set $S' \in \mathcal{B}(T^\infty)$ satisfying:

– $P((\widetilde{t} \circ e_n) \in S') = 1$;
– Each $(t_n) \in S'$ is a convergent sequence;
– For each $(t_n) \in S'$, $A \in \mathcal{G}$ and $\epsilon > 0$, there is $\delta > 0$ such that

$$|Q_n(s_n, A) - Q_n(t_n, A)| < \epsilon \quad \text{for all sufficiently large } n$$

whenever $(s_n) \in S'$ and

$$d_T(s_n, t_n) < \delta \quad \text{for all sufficiently large } n,$$

d_T being a metric for T.

We are now able to state Theorem (7.2).

Theorem (7.2) of FLR
Let X, T be Polish spaces, P an exchangeable p.m. *on $\mathcal{B}(X^\infty)$, and $\widetilde{t} : M_1^* \to T$ a Borel function. If $\widetilde{t}|D$ is predictive sufficient, $\widetilde{t}$ is continuous on M_1^*, $\nu(M_1^*) = 1$, and condition* (*) *is satisfied, then condition* (a) *holds.*

Finally we note that, if $\widetilde{t}|D$ is predictive sufficient, $\widetilde{t}$ is continuous on M_1^* and $\nu(M_1^*) = 1$, then condition (*) is necessary in order that (a) holds for a *continuous* g; cf. Theorem (7.4) of FLR.

4. The exchangeable case

In this section, P is an exchangeable p.m. on $\mathcal{B}(X^\infty)$ and $\widetilde{p}$ and ν are as in Theorem 1. For definiteness, setting

$$E = \{x \in X^\infty : e_n(x) \text{ converges weakly as } n \to \infty\},$$

it is supposed that

$$\widetilde{p}(x) = \text{weak} \lim_n e_n(x) \quad \text{for all } x \in E$$

and $\widetilde{p}(x) = p_0$ for all $x \notin E$, where p_0 is any fixed element of M_1.

Given any Borel function $\widetilde{t} : M_1^* \to T$, with $M_1^* \in \mathcal{B}(M_1)$ and $M_1^* \supset D$, let us consider the condition:

(b) $\widetilde{t}$ is injective on B, for some $B \in \mathcal{B}(M_1^*)$ such that $\nu(B) = 1$.

Before any comments on (b), we prove that it amounts to (a).

Theorem 2

Let X, T be Polish spaces, P an exchangeable p.m. *on $\mathcal{B}(X^\infty)$, and $\widetilde{t} : M_1^* \to T$ a Borel function. Then, conditions* (a) *and* (b) *are equivalent. Moreover, under* (a) *or* (b), *the function g involved in* (a) *can be taken Borel measurable.*

Proof

(a) $\Rightarrow$ (b). Under (a), there is a set $A \in \mathcal{B}(X^\infty)$, $P(A) = 1$, such that $\widetilde{p}(x) \in M_1^*$ and $g \circ \widetilde{t} \circ \widetilde{p}(x) = \widetilde{p}(x)$ for all $x \in A$. Then, $\nu(C) = 1$ whenever $C \in \mathcal{B}(M_1)$ and $C \supset \widetilde{p}(A)$, i.e., $\widetilde{p}(A)$ has ν-outer measure 1. Further, $\widetilde{p}(A)$ is an analytic set. Hence, there is $B \in \mathcal{B}(M_1)$ with $B \subset \widetilde{p}(A)$ and $\nu(B) = 1$. Since $\widetilde{t}$ is injective on $B \subset M_1^*$, condition (b) holds.

(b) $\Rightarrow$ (a). Since $\widetilde{t}$ is injective on B, for each $t \in \widetilde{t}(B)$ there is precisely one $p_t \in B$ such that $\widetilde{t}(p_t) = t$. Fix $p_0 \in M_1$, and define $g(t) = p_t$ for $t \in \widetilde{t}(B)$, and $g(t) = p_0$ for $t \in T \backslash \widetilde{t}(B)$. Then, $g(\widetilde{t}(p)) = p$ for all $p \in B$. Further, $g : T \to M_1$ is Borel measurable. In fact, $\widetilde{t}(B) \in \mathcal{B}(T)$, due to $\widetilde{t}$ is Borel and injective on B, and for the same reason one has

$$g^{-1}(H) \cap \widetilde{t}(B) = \{t \in \widetilde{t}(B) : p_t \in H\} = \widetilde{t}(B \cap H) \in \mathcal{B}(T)$$

for all $H \in \mathcal{B}(M_1)$.

Finally, let $A = \widetilde{p}^{-1}(B)$. Then, $P(A) = \nu(B) = 1$ and, for all $x \in A$, one has $\widetilde{p}(x) \in M_1^*$ and $g \circ \widetilde{t} \circ \widetilde{p}(x) = \widetilde{p}(x)$. Hence, (a) holds for a Borel function g. □

A first remark on Theorem 2, even if not essential, is that the function g in condition (a) can always be taken Borel measurable. Apart from this fact, two other points need to be discussed. One is the possible meaning of (b) and the other is that, taking (b) as a starting point, various other sufficient conditions for (a) can be obtained.

4.1. Meaning of condition (b)

From a statistical point of view, to get condition (a) (which is our goal, as explained in Section 3), it would be desirable to ask conditions only on the way $\widetilde{t}$ summarizes data. More precisely, it would be desirable to ask conditions on $\widetilde{t}|D$, in particular on its connections with predictive distributions, but not on the behaviour of $\widetilde{t}$ on $M_1^* \setminus D$.

Strictly speaking, this is not possible in general. Suppose in fact that $\nu(D) = 0$, choose any Borel function ϕ on D, and take $M_1^* \supset D$ to be any Borel set with $\nu(M_1^*) = 1$. Then, apart from trivial cases, ϕ admits two Borel extensions to M_1^*, say $\widetilde{t}_1$ and $\widetilde{t}_2$, such that (b) holds for $\widetilde{t}_1$ and fails for $\widetilde{t}_2$. In view of Theorem 2, condition (a) holds for $\widetilde{t}_1$ but fails for $\widetilde{t}_2$, even if $\widetilde{t}_1|D = \phi = \widetilde{t}_2|D$.

In Theorem (7.2) of FLR, for instance, continuity of $\widetilde{t}$ is asked on all M_1^*. However, all other conditions of Theorem (7.2) concern $\widetilde{t}|D$ only, and also the continuity assumption looks admissible. Indeed, when $\widetilde{t}$ is continuous on D and admits a continuous extension to M_1^* (e.g., when $\widetilde{t}$ is uniformly continuous on D), it is natural to take $\widetilde{t}$ on M_1^* as such continuous extension.

Generally, condition (b) deals with the behaviour of $\widetilde{t}$ on $M_1^* \setminus D$, and in this sense it does not have an intuitive statistical content. Nevertheless, (b) is also necessary, and thus one can think in term of it without any real loss of generality. In addition, (b) makes clear which properties are requested to $\widetilde{t}$ in order that $\widetilde{p}$ can be reduced through $\widetilde{t}$:

over a set B of ν-probability 1, $\widetilde{t}$ must be able to distinguish between two different weak limits of empirical measures.

4.2. Other sufficient conditions for (a)

By using Theorem 2, sufficient conditions for (a) can be obtained. Moreover, it is possible to give a shorter (and more direct) proof of Theorem (7.2). We begin with the latter point.

Proof of Theorem (7.2) of FLR

Since $\mathcal{G}$ is countable and $P_1(\partial A) = 0$ for all $A \in \mathcal{G}$, there is $F_1 \in \mathcal{B}(X^\infty)$ with $P(F_1) = 1$ and $\widetilde{p}(x)(\partial A) = 0$ for all $x \in F_1$ and $A \in \mathcal{G}$. Moreover, $Q_n(\widetilde{t} \circ e_n, \cdot\,) \Rightarrow \widetilde{p}$, P-a.s., where the Q_n are as in condition (*).

Let

$$F = E \cap F_1 \cap \{(\widetilde{t} \circ e_n) \in S'\} \cap \{Q_n(\widetilde{t} \circ e_n, \cdot\,) \Rightarrow \widetilde{p}\}.$$

Since $\widetilde{p}(F)$ is an analytic set and $P(F) = 1$, there is $H \in \mathcal{B}(M_1)$ with $H \subset \widetilde{p}(F)$ and $\nu(H) = 1$. Let $B = H \cap M_1^*$. By definition of F, continuity of $\widetilde{t}$ and condition (*), it follows that $\widetilde{p}(x) = \widetilde{p}(y)$ whenever $x, y \in F$ and $\widetilde{t} \circ \widetilde{p}(x) = \widetilde{t} \circ \widetilde{p}(y)$. Hence, $\widetilde{t}$ is injective on B, and since $\nu(B) = \nu(M_1^*) = 1$, condition (b) holds. By Theorem 2, this concludes the proof. □

Let us turn now to some other sufficient conditions for (a). If assumptions on the behaviour of $\widetilde{t}$ on $M_1^* \backslash D$ are allowed, then, by Theorem 2, a plenty of conditions are available. Because of the remarks in Subsection 4.1, however, we focus on conditions concerning $\widetilde{t}|D$ only, apart from continuity of $\widetilde{t}$ which is asked on all M_1^*. Then, one possible candidate is:

(c) There is a set $A \in \mathcal{B}(X^\infty)$, $P(A) = 1$, such that

$$lim\, sup_n\, d_T\, (\widetilde{t} \circ e_n(x), \widetilde{t} \circ e_n(y)) > 0$$

whenever $x, y \in A$ and

$$lim\, inf_n\, \rho(e_n(x), e_n(y)) > 0$$

where d_T is a metric for T and ρ is Prohorov metric on M_1.

We recall that Prohorov metric ρ on M_1 is given by

$$\rho(p, q) = inf\, \{\epsilon > 0 : p(A) \leq q(A^\epsilon) + \epsilon \text{ for all } A \in \mathcal{B}(X)\}$$

where $A^\epsilon = \{u : d_X(u, A) < \epsilon\}$, d_X being a metric for X.

Theorem 3

Let X, T be Polish spaces, P an exchangeable p.m. *on $\mathcal{B}(X^\infty)$, and $\widetilde{t} : M_1^* \to T$ a Borel function. If $\widetilde{t}$ is continuous on M_1^* and $\nu(M_1^*) = 1$, then conditions* (a), (b) *and* (c) *are equivalent.*

Proof

Let $\widetilde{t}$ be continuous on M_1^* with $\nu(M_1^*) = 1$. By Theorem 2, it is enough to prove that (b) and (c) are equivalent.

(c) $\Rightarrow$ (b). Let E be defined as at the beginning of Section 4. Since $\widetilde{p}(A \cap E)$ is analytic and has ν-outer measure 1, there is $H \in \mathcal{B}(M_1)$ with $H \subset \widetilde{p}(A \cap E)$ and $\nu(H) = 1$. Then, (b) holds with $B = H \cap M_1^*$. To see this, the only non trivial fact is injectivity of $\widetilde{t}$ on B. Fix $p_1, p_2 \in B$ such that $p_1 \neq p_2$, and take $x_i \in A \cap E$ with $p_i = \widetilde{p}(x_i)$ for $i = 1, 2$. Then, $\lim_n \rho(e_n(x_1), e_n(x_2)) = \rho(p_1, p_2) > 0$, due to $x_1, x_2 \in E$ and $p_1 \neq p_2$, and thus continuity of $\widetilde{t}$ and condition (c) yield

$$d_T(\widetilde{t}(p_1), \widetilde{t}(p_2)) = \lim_n d_T(\widetilde{t} \circ e_n(x_1), \widetilde{t} \circ e_n(x_2)) > 0.$$

(b) $\Rightarrow$ (c). Define $A = E \cap \widetilde{p}^{-1}(B)$, note that $P(A) = 1$, and fix $x, y \in A$ such that $lim\,inf_n\, \rho(e_n(x), e_n(y)) > 0$. Then, $\widetilde{p}(x)$, $\widetilde{p}(y) \in B$ and, since $x, y \in E$,

$$\rho(\widetilde{p}(x), \widetilde{p}(y)) = \lim_n \rho(e_n(x), e_n(y)) > 0.$$

Thus, continuity and injectivity of $\widetilde{t}$ on B yield

$$\lim_n d_T(\widetilde{t} \circ e_n(x), \widetilde{t} \circ e_n(y)) = d_T(\widetilde{t} \circ \widetilde{p}(x), \widetilde{t} \circ \widetilde{p}(y)) > 0. \;\square$$

By Theorem 3, in the relevant case where $\widetilde{t}$ is continuous on M_1^* and $\nu(M_1^*) = 1$, (c) is equivalent to (a). Moreover, in line with the remarks in Subsection 4.1, condition (c) deals with $\widetilde{t}|D$ only. Hence, Theorem 3 is an improvement of Theorem (7.2).

Next, given a function $f : S \rightarrow S'$ with S and S' metric spaces (with distances d and d'), let us call f "uniformly injective" in case: For each $\epsilon > 0$ there is $\delta > 0$ such that $d'(f(a), f(b)) \geq \delta$ whenever $a, b \in S$ and $d(a, b) \geq \epsilon$. If $\widetilde{t}$ is uniformly injective on D, then condition (c) trivially holds. Thus, by Theorem 3, it is enough for condition (a) that $\widetilde{t}$ is uniformly injective on D and continuous on M_1^*, with $\nu(M_1^*) = 1$.

We close this section by noting that condition (c) becomes more meaningful if $\rho(e_n(x), e_n(y))$ is translated into a sort of distance between $(x_1, ..., x_n)$ and $(y_1, ..., y_n)$. To this end, it is convenient to replace Prohorov distance with some other equivalent metric. Let L be the set of real valued functions f on X such that, for all $a, b \in X$,

$$|f(a) - f(b)| \leq 1 \wedge d_X(a, b).$$

The so called bounded Lipschitz metric on M_1 is defined as

$$d_{BL}(p, q) = sup_{f \in L} \left|\int f dp - \int f dq\right|,$$

and it can be shown to satisfy $\rho^2 \leq d_{BL} \leq 2\rho$ (see, e.g., Huber, 1981, Corollary 4.3, p. 33). Thus, by using d_{BL} instead of ρ, condition (c) can be equivalently written as:

(c) There is a set $A \in \mathcal{B}(X^\infty)$, $P(A) = 1$, such that

$$lim\, sup_n\, d_T\, (\widetilde{t} \circ e_n(x), \widetilde{t} \circ e_n(y)) > 0$$

whenever $x, y \in A$ and

$$lim\, inf_n\, sup_{f \in L}\, |\tfrac{1}{n}\sum_{i=1}^{n} f(x_i) - \tfrac{1}{n}\sum_{i=1}^{n} f(y_i)| > 0.$$

5. The stationary case

This section includes versions of Theorems 2 and 3 and a convergence result for the case where P is stationary. Some remarks on Bayesian nonparametric inference for stationary data, or more generally for data with invariant distribution, are also given. We begin with a result asserting that, under general conditions, every invariant p.m. is a unique integral mixture of extreme points.

Given a measurable space $(\Omega, \mathcal{A})$ and a class $\mathbb{F}$ of measurable functions $f : (\Omega, \mathcal{A}) \to (\Omega, \mathcal{A})$, let $\mathcal{I}$ be the set of those p.m.'s P on $\mathcal{A}$ which are $\mathbb{F}$-invariant, i.e., $P = P \circ f^{-1}$ for all $f \in \mathbb{F}$. Further, let $ext\, \mathcal{I}$ be the set of extreme points of $\mathcal{I}$, and let $ext\, \mathcal{I}$ be equipped with the trace σ-field $(ext\, \mathcal{I}) \cap \mathcal{P}(\mathcal{A})$. According to Section 2, $(ext\, \mathcal{I}) \cap \mathcal{P}(\mathcal{A})$ is generated by the maps $Q \mapsto Q(A), A \in \mathcal{A}$, from $ext\, \mathcal{I}$ into the reals. We also recall that a p.m. P on $\mathcal{A}$ is *perfect* if, for each $\mathcal{A}$-measurable function $f : \Omega \to \mathbb{R}$, there is $B \in \mathcal{B}(\mathbb{R})$ such that $B \subset f(\Omega)$ and $P(f \in B) = 1$. Next Theorem 4, due to Maitra (1977), unifies results of Bogoliouboff, de Finetti, Farrel, Kryloff and Varadarayan.

Theorem 4 (Maitra)

Suppose $\mathbb{F}$ *is countable,* $\mathcal{A}$ *is countably generated and includes the singletons, and every* p.m. *on* $\mathcal{A}$ *is perfect. Then, for each* $P \in \mathcal{I}$ *there is a unique* p.m. μ *on* $(ext\,\mathcal{I}) \cap \mathcal{P}(\mathcal{A})$ *such that* $P(A) = \int_{ext\,\mathcal{I}} Q(A)\mu(dQ)$ for all $A \in \mathcal{A}$.

In our case, $(\Omega, \mathcal{A}) = (X^\infty, \mathcal{B}(X^\infty))$ where X is Polish, so that $(\Omega, \mathcal{A})$ meets the conditions of Theorem 4 and the σ-field $(ext\,\mathcal{I}) \cap \mathcal{P}(\mathcal{A})$ reduces to $\mathcal{B}(ext\,\mathcal{I})$. If $\mathbb{F}$ is the class of all finite permutations of X^∞, then $\mathcal{I}$ is the set of exchangeable p.m.'s and $ext\,\mathcal{I}$ the set of product p.m.'s, i.e., $ext\,\mathcal{I} = \{p^\infty : p \in M_1\}$. Hence, the equivalence between (i) and (ii) in Theorem 1 follows from Theorem 4, after noting that μ and ν are connected by the relation $\mu = \nu \circ \phi^{-1}$, where $\phi(p) = p^\infty$ for $p \in M_1$. In other terms, at least formally, de Finetti's theorem can be embedded into a more general result on invariant p.m.'s.

Since de Finetti's theorem is fundamental in Bayesian nonparametric inference for exchangeable data, one could hope that, taking Theorem 4 as a starting point, a relevant part of the usual theory can be extended to the invariant case. In principle, this is possibly true. However, moving from the exchangeable to the invariant case, the problem becomes technically much more intricated. So, developing a nonparametric theory for invariant data, analogous to the usual one for exchangeable data, seems to be very hard. In particular, it looks hard to get usable statistical procedures. On the other hand, it would be interesting to investigate which part, if any, of the usual Bayesian nonparametric theory can be extended to invariant data.

In the sequel, as a significant example, we discuss the stationary case. Let $s : X^\infty \to X^\infty$ be the shift transformation: $s(x_1, x_2, \ldots) = (x_2, x_3, \ldots)$. A p.m. P on $\mathcal{B}(X^\infty)$ is *stationary* if $(\widetilde{x}_1, \widetilde{x}_2, \ldots)$ and $(\widetilde{x}_2, \widetilde{x}_3, \ldots)$ have the same distribution under P, or equivalently if $P = P \circ s^{-1}$. Clearly, P is stationary in case is exchangeable, but the converse is not true. If P is stationary and $P(A) \in \{0, 1\}$ for each Borel set A with $A = s^{-1}A$, then P is said to be *ergodic*. When P is stationary or ergodic, we will also say that $\widetilde{x}$ is stationary or ergodic under P. Let C be the set of ergodic p.m.'s and M_2 the set of all p.m.'s on $\mathcal{B}(X^\infty)$; cf. Section 2. By relying on an argument of Maitra (1977), we now prove that C is a Borel subset of M_2.

Lemma 5

If X is a separable metric space, then $C \in \mathcal{B}(M_2)$.

Proof

Let: $\mathbb{F} = \{s\}$, $\mathcal{I}$ the set of stationary p.m.'s, $\mathcal{N} = \{A \in \mathcal{B}(X^\infty) : P(A) = 0$ for all $P \in \mathcal{I}\}$, and $\mathcal{U} = \{A \in \mathcal{B}(X^\infty) : P(A\Delta(s^{-1}A)) = 0$ for all $P \in \mathcal{I}\}$. By Lemma 4 of Maitra (1977), the σ-field $\mathcal{U}$ is sufficient (in the classical sense) for $\mathcal{I}$. Since $\mathcal{B}(X^\infty)$ is countably generated, sufficiency of $\mathcal{U}$ implies existence of a sufficient and countably generated σ-field $\mathcal{U}_0$ such that $\mathcal{U}_0 \subset \mathcal{U} \subset \sigma(\mathcal{U}_0 \cup \mathcal{N})$; cf. Burkholder (1961, Theorem 1). Fix countable fields $\mathcal{H}_0$ and $\mathcal{H}$ such that $\mathcal{U}_0 = \sigma(\mathcal{H}_0)$ and $\mathcal{B}(X^\infty) = \sigma(\mathcal{H})$, and define

$$B = \{P \in M_2 : P(A) = P(s^{-1}A) \text{ for all } A \in \mathcal{H}, \text{ and } P(A) \in \{0,1\} \text{ for all } A \in \mathcal{H}_0\}.$$

Since $\mathcal{H}_0$ and $\mathcal{H}$ are countable, B is Borel, and since $\mathcal{B}(X^\infty) = \sigma(\mathcal{H})$, one has $B \subset \mathcal{I}$. Let $P \in B$. Since P is degenerate on the π-class $\mathcal{H}_0 \cup \mathcal{N}$, then P is also degenerate on $\sigma(\mathcal{H}_0 \cup \mathcal{N})$. Since $\mathcal{U} \subset \sigma(\mathcal{U}_0 \cup \mathcal{N}) = \sigma(\mathcal{H}_0 \cup \mathcal{N})$, it follows that $P \in C$. Hence, $B \subset C$, while it is clear that $B \supset C$. To sum up, $C = B \in \mathcal{B}(M_2)$. □

Setting $\mathbb{F} = \{s\}$, Theorem 4 applies to stationary p.m.'s, and the set of extreme points of stationary p.m.'s coincides with C. Hence, each stationary P admits the representation $P(\cdot) = \int Q(\cdot)\mu(dQ)$ for some unique p.m. μ on $\mathcal{B}(C)$. Furthermore, just as in the exchangeable case, μ is the probability distribution of $\tilde{q}$, for some Borel function $\tilde{q} : X^\infty \to C$ such that $\tilde{q}(A)$ is a version of $P(\tilde{x} \in A|\tilde{q})$ for all $A \in \mathcal{B}(X^\infty)$.

To realize the program sketched above, i.e., to develop a Bayesian nonparametric theory for stationary data, one has to assess priors on C. Precisely, one should "propose" some reasonable class of priors μ on $\mathcal{B}(C)$, and calculate the corresponding posterior and predictive distributions. Such priors should have large support, so as to obtain a real nonparametric theory. Further, they should cover a broad range of potential beliefs, and the posterior and predictive distributions should be not too difficult to evaluate. Clearly, it is not easy to put together all these requisites.

As a preliminary step, we investigate, for stationary data, the same problem of Section 4, i.e., existence of underlying parametric models.

A different kind of empirical measure is to be used. Given $k \in \mathbb{N} \cup \{0\}$, define:

$$\widetilde{y}_j = (\widetilde{x}_j, ..., \widetilde{x}_{j+k}) \quad \text{for } j \in \mathbb{N},$$

$$f_{n,k} = \tfrac{1}{n-k} \sum_{j=1}^{n-k} \delta_{\widetilde{y}_j} \quad \text{for } n > k.$$

For fixed n, k and $x \in X^\infty$, $f_{n,k}(x)$ is a p.m. on $\mathcal{B}(X^{k+1})$. To obtain a p.m. on $\mathcal{B}(X^\infty)$, we fix any $\gamma \in M_2$ and we refer to $f_{n,k} \times \gamma$ instead of $f_{n,k}$. (For each p.m. α on $\mathcal{B}(X^{k+1})$, $\alpha \times \gamma$ denotes the p.m. on $\mathcal{B}(X^\infty)$ under which $\widetilde{y}_1$ has distribution α, $(\widetilde{x}_{k+2}, \widetilde{x}_{k+3}, ...)$ has distribution γ, and $\widetilde{y}_1$ is independent of $(\widetilde{x}_{k+2}, \widetilde{x}_{k+3}, ...)$). Clearly, this is only a rough device to transform $f_{n,k}$ into a p.m. on $\mathcal{B}(X^\infty)$, and γ will not play any essential role. Next, call k_n the integer part of $n/2$ and define

$$f_n = f_{n,k_n} \times \gamma.$$

Thus, each $f_n : X^\infty \to M_2$ is a Borel function, and it is a summary of the data $(\widetilde{x}_1, ..., \widetilde{x}_n)$. When P is stationary, we will use the f_n as empirical measures. One reason is the following.

Theorem 6

If X is a Polish space and P a stationary p.m. *on $\mathcal{B}(X^\infty)$, then*

$$f_n \Rightarrow \widetilde{q} \quad P\text{-a.s.}$$

where " $\Rightarrow$ " stands for weak convergence of p.m.'s.

Proof

Conditionally on $\widetilde{q}$, $\widetilde{x}$ is ergodic with distribution $\widetilde{q}$. Hence, it is enough to show that, if P is ergodic then $f_n \Rightarrow P$, P-a.s.. Suppose that P is ergodic, and fix $k \in \mathbb{N}$. Let ϕ_k be the canonical projection of X^∞ onto X^{k+1}. Since the sequence of X^{k+1}-valued random variables $(\widetilde{y}_n : n \in \mathbb{N}) = ((\widetilde{x}_n, ..., \widetilde{x}_{n+k}) : n \in \mathbb{N})$ is ergodic under P, one has $f_{n,k} \Rightarrow P \circ \phi_k^{-1}$ P-a.s. as $n \to \infty$. Further, for all $B \in \mathcal{B}(X^{k+1})$ and $n > 2k+1$, a direct calculation shows that

$$f_n \circ \phi_k^{-1}(B) = f_{n,k_n}(B \times X^{k_n - k}) = f_{n-k_n+k,k}(B).$$

Thus, $f_n \circ \phi_k^{-1} \Rightarrow P \circ \phi_k^{-1}$, P-a.s. as $n \to \infty$, and this concludes the proof. □

Since f_n (and not e_n) is now used as empirical measure, the notion of statistic is to be slightly modified, too. Let

$$G = \{f_n(x) : x \in X^\infty, n \in \mathbb{N}\}$$

be the union of the ranges of all the f_n, and let $M_2^* \in \mathcal{B}(M_2)$ be such that $M_2^* \supset G$. In what follows, in line with the notion adopted so far, a statistic is meant as the restriction to G, $\widetilde{t}|G$, of any Borel function $\widetilde{t} : M_2^* \to T$.

At this stage, the argument proceeds essentially as in Sections 3 and 4. The first part of Section 3 remains unaffected, apart from "exchangeable" is to be replaced by "stationary", "i.i.d." by "ergodic", ν by μ, and $\widetilde{p}$ by $\widetilde{q}$. In particular, given a stationary P and a Borel function $\widetilde{t}: M_2^* \to T$, where $M_2^* \in \mathcal{B}(M_2)$ and $M_2^* \supset G$, condition (a) simply becomes

(a$_0$) $\widetilde{q} \in M_2^*$ and $g(\widetilde{t}(\widetilde{q})) = \widetilde{q}$, P-a.s., for some Borel function $g: T \to M_2$.

A slight difference between (a$_0$) and (a), suggested by Theorem 2, is that g is now asked to be a Borel function, and not merely a $\overline{\mathcal{B}}(T)$-measurable function. In any case, if (a$_0$) holds then, conditionally on $\widetilde{t}(\widetilde{q})$, $\widetilde{x}$ is ergodic with distribution $\widetilde{q}$. So, under (a$_0$), the "original random parameter" $\widetilde{q}$ can be reduced through $\widetilde{t}$, i.e., the random parameter can be taken to be $\widetilde{\theta} = \widetilde{t}(\widetilde{q})$. Next, conditions (b) and (c) turn into:

(b$_0$) $\widetilde{t}$ is injective on B, for some $B \in \mathcal{B}(M_2^*)$ such that $\mu(B) = 1$;

(c$_0$) There is a set $A \in \mathcal{B}(X^\infty)$, $P(A) = 1$, such that

$$lim\, sup_n\, d_T\, (\widetilde{t} \circ f_n(x), \widetilde{t} \circ f_n(y)) > 0$$

whenever $x, y \in A$ and

$$lim\, inf_n\, \rho(f_n(x), f_n(y)) > 0;$$

where ρ is now Prohorov metric on M_2. As in Section 4, condition (c$_0$) is perhaps more expressive if ρ is replaced by some other equivalent metric, like the bounded Lipschitz metric on M_2.

Finally, the arguments for proving Theorems 2 and 3 do not depend on exchangeability of P, and can be repeated for a stationary P. In fact, up to a proper choice of the empirical measure (and thus up to Theorem 6), the results in this section hold for any $\mathbb{F}$-invariant p.m. P on $\mathcal{B}(X^\infty)$, with $\mathbb{F}$ countable. We state the stationary versions of Theorems 2 and 3 jointly, and we omit proofs.

Theorem 7

Let X, T *be Polish spaces,* P *a stationary* p.m. *on* $\mathcal{B}(X^\infty)$*, and* $\widetilde{t} : M_2^* \to T$ *a Borel function. Then,* (a_0) *is equivalent to* (b_0). *Moreover, if* $\widetilde{t}$ *is continuous on* M_2^* *and* $\mu(M_2^*) = 1$*, then conditions* (a_0), (b_0) *and* (c_0) *are equivalent.*

Acknowledgements: Research partially supported by: M.U.R.S.T. 40% "Processi Stocastici, Calcolo Stocastico ed Applicazioni".

References

Burkholder D.L. (1961) Sufficiency in the undominated case, *Annals of Mathematical Statistics*, 32, 1191-1200.

Fortini S., Ladelli L., Regazzini E. (2000) Exchangeability, predictive distributions
and parametric models, *Sankhya*, 62, Series A, 86-109.

Huber P.J. (1981) *Robust statistics*, J.Wiley, New York.

Maitra A. (1977) Integral representations of invariant measures, *Transactions of the American Mathematical Society*, 229, 209-225.

Olshen R. (1974) A note on exchangeable sequences, *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 28, 317-321.