Sankhyā: The Indian Journal of Statistics
2000, Volume 62, Series A, Pt. 1, pp. 86–109

EXCHANGEABILITY, PREDICTIVE DISTRIBUTIONS AND PARAMETRIC MODELS*

By SANDRA FORTINI
Università "L. Bocconi",
LUCIA LADELLI
Politecnico di Milano,
and
EUGENIO REGAZZINI
Università di Pavia, CNR-IAMI.

SUMMARY. In the general setting of predictive inference, when observations are exchangeable and take values in a Polish space, conditions are stated in order that parametric models turn out to be limiting forms of predictive distributions and parameters are limiting forms of suitable predictive sufficient statistics. The treatment is completed by a necessary and sufficient condition in order that a sequence of predictive distributions may be consistent with an exchangeable distribution. Moreover, main properties of predictive sufficiency are revisited in the general setting described above.

1. **Introduction**

De Finetti's characterization of exchangeable laws (see next Section 2) is a cornerstone of modern Bayesian theory. Generally speaking, implementation of statistical methods get simpler when de Finetti's measures are supported by parametric families of probabilities. Thus, Bayesian statisticians are concerned with the general question whether, and under what circumstances, it is coherent to reduce the class of all admissible distributions on a given sample space to some distinguished parametric family. In fact, "... it is not at all clear, at first sight, how we should interpret 'beliefs about parameters' ... or even whether such 'beliefs' have any intrinsic interest" (Bernardo and Smith, 1994, page 244).

These questions, which echo Chapter V of de Finetti (1937), have been answered (see next Section 5) by noting that the most common parametric models are limiting forms of predictive distributions depending on predictive sufficient statistics. [Apart from parametric forms, this is the kernel of de Finetti's representation which, in

Paper received. May 1998; revised October 1999.
AMS (1991) *subject classification.* Primary: 62A15. secondary: 60G09, 62B20.
Key words and phrases. de Finetti's measure, exchangeable random variables, exponential model, predictive distributions, predictive sufficient statistics.
* Research partially supported by Progetto Strategico "Decisioni statistiche: teoria e applicazioni" (1997).

point of fact, states that exchangeable laws of infinite sequences are mixtures of limits of empirical distributions; see (R3) of Section 2]. Acceptability of this answer is based on the following remarks: (a) Unlike prior and posterior distributions of parameters, predictive distributions [conditional distributions of future observable facts, given observed facts] can always be assessed; (b) Their dependence on a particular summary statistic follows from a direct, subjective judgement implying that possible differences between data do not actually bring about differences between previsions whenever data produce the same value for that statistic.

The present paper focuses on the question whether, and under what circumstances, statistical models can be interpreted as limiting forms of conditional previsions about observables. Although this problem is clear enough in modern Bayesian literature [see, e.g., Sections 4.5 and 4.6 of Bernardo and Smith (1994)], it has not been adequately addressed, to the best of our knowledge. In the final section [Section 7] of the present paper, conditions are given under which the support of de Finetti's measure reduces the issue to a family of laws depending on a parameter, the parameter being the limit of a sequence of predictive sufficient statistics, as the number of observed quantities goes to $+\infty$. It's worth noting that all those conditions are of a predictive nature, in the sense that they do not anticipate properties of the limiting models. Section 7 is the final step of a line of reasoning which involves exchangeability and significant connections between predictive sufficiency and other (conventional) forms of sufficiency. And some of the results achieved in the present paper seem indeed to be new. Section 3 states a necessary and sufficient condition in order that a sequence of predictive distributions be consistent with an exchangeable law. As an application of such condition it is shown how its use improves some well-known characterizations and, at the same time, simplifies their proofs; cf. Section 4. Section 5 is devoted to general aspects of predictive sufficiency, and tries to develop a mathematically adequate and more general than usual treatment of the subject. The proofs of some propositions are omitted. They can be found in Fortini, Ladelli and Regazzini (1998). Section 6 includes an application of the previous results to the characterization, in predictivistic terms, of exponential families. Technically speaking, this characterization is made possible by the result relative to Cauchy's functional equation obtained by Diaconis and Freedman within their general theory of exchangeability and sufficiency; see Diaconis and Freedman (1984) and (1990). By the way, it goes without saying that their approach is different from the predictivistic one followed in the present paper.

The final remarks of the present section focus on the adoption of the condition of exchangeability. In view of its mathematical simplicity, exchangeability leads to neat results by avoiding technical difficulties which could in point of fact hide the true nature of the statistical problems involved. On the other hand, the adoption of exchangeability does not represent a serious restriction, since: (a) Results deduced in its presence are generally extensible, almost directly, to partially exchangeable classes of observables [cf. de Finetti (1938)]; (b) There are forms of dependence – different from (partial) exchangeability but of interest to Bayesian inference – which can be reduced to (partial) exchangeability by means of suitable transformations of the sequence of observables. Recall, for example, that it is mixtures of recurrent

Markov chains which provide the general representation for the laws of partially exchangeable arrays of adjacent subsequent states; cf. de Finetti (1959), Diaconis and Freedman (1980), Fortini, Ladelli, Petris and Regazzini (1999).

2. Preliminaries and Notation

Consider a sequence of observable elements taking values in a set X, and let $\mathcal{X}$ be a σ-algebra of subsets of X. Write X^n for the n-fold Cartesian product and $\mathcal{X}^n$ for the usual product σ-algebra ($n = 1, 2, \ldots, \infty$). Define $\tilde{x}_1, \tilde{x}_2, \ldots$ to be the *coordinate random variables* (r.v.'s) of X^∞, i.e. $\tilde{x}_i(x) = x_i$ for every $x = (x_1, x_2, \ldots)$ in X^∞. We will assume that $\mathcal{X}^n = \mathcal{B}(X)^n$, whenever X is a topological space, where, for any topological space Y, $\mathcal{B}(Y)$ denotes the Borel σ-algebra on Y. Thus, if X is a *Polish* space [separable, complete, metric space], then $\mathcal{X}^n := \mathcal{B}(X)^n = \mathcal{B}(X^n)$ ($n = 1, 2, \ldots, \infty$).

For any finite subset $\{i_1, \ldots, i_n\}$ of $\mathbb{N}$ denote by $e_n(\tilde{x}_{i_1}, \ldots, \tilde{x}_{i_n})$ the *empirical measure* for $(\tilde{x}_{i_1}, \ldots, \tilde{x}_{i_n})$, i.e.

$$e_n(\tilde{x}_{i_1}, \ldots, \tilde{x}_{i_n}) = \frac{1}{n}\sum_{k=1}^{n} \delta_{\tilde{x}_{i_k}}$$

where δ_a denotes the unit mass at a. Here $e_n(\tilde{x}_{i_1}, \ldots, \tilde{x}_{i_n})$ is thought of as a function of the set X^∞ into the set M_1 of probability measures on $(X, \mathcal{X})$. M_1 is made into a measurable space by the σ-algebra $\mathcal{M}_1$ generated by all sets $\{p \in M_1 : p(A) \in B\}$, with A in $\mathcal{X}$ and B in $\mathcal{B}([0, 1])$. It is straightforward that $e_n(\tilde{x}_{i_1}, \ldots, \tilde{x}_{i_n})$ is an $(\mathcal{X}^\infty/\mathcal{M}_1)$-measurable function. If X is metric, and M_1 is made into a topological space by the topology of *weak convergence*, then $\mathcal{M}_1 = \mathcal{B}(M_1)$ [see, e.g. Kallenberg (1986), Lemma 1.3 and Lemma 1.4]. Recall that M_1 can be metrized as a separable metric space if X is a separable metric space, and that M_1 is topologically complete if X is so. Hence, when X is a Polish space, with $\mathcal{X} = \mathcal{B}(X)$, the same holds for M_1; cf., e.g., Theorems 6.2 and 6.5 in Chapter II of Parthasaraty (1967).

As far as the probabilistic aspects of $\tilde{x} = (\tilde{x}_n)_{n\geq 1}$ are concerned, the present paper focuses on probability measures P on $(X^\infty, \mathcal{X}^\infty)$, which make the $\tilde{x}_n$ *exchangeable*, i.e. the probability distribution (p.d.) of $(\tilde{x}_{\pi(n)})_{n\geq 1}$ is the same as the p.d. of $(\tilde{x}_n)_{n\geq 1}$ for any finite permutation π. In this case, P will be called exchangeable. Under suitable conditions about $(X, \mathcal{X})$, an exchangeable P admits the *de Finetti representation*.

If X is a Polish space, then the following statements are equivalent.

(**R1**) $\tilde{x}_1, \tilde{x}_2, \ldots$ *are exchangeable with respect to* P.

(**R2**) *There is a random probability measure* $\tilde{p}$ [= r.v. on $(X^\infty, \mathcal{B}(X^\infty))$ taking values in $(M_1, \mathcal{B}(M_1))$] *such that* $\tilde{p}^\infty(B)$ *is a version of the conditional probability* $P(\tilde{x} \in B \mid \tilde{p})$ *for every* B *in* $\mathcal{B}(X^\infty)$, *where* $\tilde{p}^\infty$ *is the power probability measure which makes the coordinates i.i.d. with p.d.* $\tilde{p}$.

(**R3**) *There exists a unique probability measure* ν *on* $(M_1, \mathcal{B}(M_1))$ *such that*

$$P(\tilde{x} \in B) = \int_{M_1} p^\infty(B)\nu(dp) \qquad\qquad (B \in \mathcal{B}(X^\infty)).$$

This measure is said to be the de Finetti measure of P [of $\tilde{x}$, equivalently] *and coincides with the p.d. of* $\tilde{p}$. *Moreover,*

$$e_n(\tilde{x}_1, \ldots, \tilde{x}_n) \Rightarrow \tilde{p}$$

a.s.-P [where $\Rightarrow$ denotes weak convergence].

In Bayesian terms, ν is the prior p.d. of $\tilde{p}$. For proofs see, e.g., Aldous (1985) or Schervish (1995). Throughout the present paper, conditional expectation is meant as an equivalence class of functions [versions] which differ from one another only on sets of zero probability. The same symbol will denote both conditional expectation and one of its versions.

Equivalence of (R1), (R2) implies that the $\tilde{x}_n$ are exchangeable if and only if they are conditionally i.i.d. given the σ-algebra $\sigma(\tilde{p})$ [given a r.v. χ on a measurable space, $\sigma(\chi)$ will denote the σ-algebra generated by χ]. It is a remarkable fact that if the $\tilde{x}_n$ are conditionally i.i.d. given a σ-algebra $\mathcal{G} \subset \mathcal{B}(X^\infty)$, then $\mathcal{G}^* = \sigma(\tilde{p})^*$ where, for any sub-σ-algebra $\mathcal{H}$ of $\mathcal{B}(X^\infty)$,

$$\mathcal{H}^* := \{H \triangle N : H \in \mathcal{H}, N \in \mathcal{B}(X^\infty), \text{ with } P(N) = 0\}.$$

This result is proved, e.g., in Chow and Teicher (1997), page 237.

Let $(T, \mathcal{T})$ be a measurable space. As usual, if T is a metric space, then $\mathcal{T}$ is meant as its Borel σ-algebra $\mathcal{B}(T)$. If $\tilde{t}_n$ is a $(\mathcal{X}^n/\mathcal{T})$-measurable function of X^n into T, then $\tilde{t}_n(\tilde{x}_{i_1}, \ldots, \tilde{x}_{i_n})$ is said to be a statistic of $(\tilde{x}_{i_1}, \ldots, \tilde{x}_{i_n})$. This paper deals with statistics defined by means of restrictions to the set of empirical measures, of functions of some subset of M_1 into T. More precisely, let M_1^* be a measurable subset of M_1 and $\tilde{t} : M_1^* \to T$ a $(M_1^* \cap \mathcal{M}/\mathcal{T})$-measurable function. Define D to be the union of the ranges of all empirical measures. If X is a Polish space, then $D \in \mathcal{B}(M_1)$. Indeed, the range D_n of $e_n(\tilde{x}_1, \ldots, \tilde{x}_n)$ is the same as the range of $e_n(\tilde{x}_1^*, \ldots, \tilde{x}_n^*)$, where $\tilde{x}_i^*$ is the i-th coordinate r.v. of X^n for $i = 1, \ldots, n$. The function $e_n(\tilde{x}_1^*, \ldots, \tilde{x}_n^*)$ is $(\mathcal{B}(X^n)/\mathcal{B}(M_1))$-measurable and, for any y in D_n, $[e_n(\tilde{x}_1^*, \ldots, \tilde{x}_n^*)]^{-1}(y)$ is a finite subset of X^n. Then, from Theorem III 21a in Dellacherie and Meyer (1975), D_n and, consequently, $D = \cup_{n \geq 1} D_n$ belong to $\mathcal{B}(M_1)$. Thus, if X is a Polish space and $M_1^* \supset D$, then $\tilde{t} \circ e_n$ is a statistic. Throughout the present paper, these conditions are supposed to be satisfied, and the term statistic designates the restriction $\tilde{t}_{|D}$ of $\tilde{t}$ to D.

3. Predictive Distributions and Exchangeability

Let $(X^\infty, \mathcal{X}^\infty, P)$ and $\tilde{x}$ be the same as in Section 2. Recall that a *regular conditional p.d.* for $\tilde{x}_n$ given $\tilde{x}_{(n-1)} := (\tilde{x}_1, \ldots, \tilde{x}_{n-1})$, with $n \in \{2, 3, \ldots\}$, is a function P_n on $X^{n-1} \times \mathcal{X}$ satisfying the following properties:

(**P1**) $P_n(x_{(n-1)}, \cdot)$ is a probability measure on $\mathcal{X}$ for each $x_{(n-1)} = (x_1, \dots, x_{n-1}) \in X^{n-1}$;

(**P2**) $P_n(\cdot, A)$ is $\mathcal{X}^{n-1}$-measurable for each A in $\mathcal{X}$;

(**P3**) For every A_1 and A_2 in $\mathcal{X}^{n-1}$ and $\mathcal{X}$, respectively,

$$P(\tilde{x}_{(n)} \in A_1 \times A_2) = \int_{A_1} P_n(y, A_2) P\tilde{x}_{(n-1)}^{-1}(dy)$$

where $P\tilde{x}_{(n-1)}^{-1}$ is the p.d. of $\tilde{x}_{(n-1)}$.

Any function P_n on $X^{n-1} \times \mathcal{X}$ which satisfies $(P1)$ and $(P2)$ is called *transition probability* with respect to (w.r.t.) $X^{n-1} \times \mathcal{X}$. Then, a regular conditional p.d. (r.c.p.d.) P_n on $X^{n-1} \times \mathcal{X}$ is a transition probability w.r.t. $X^{n-1} \times \mathcal{X}$ such that $P_n(\tilde{x}_{(n-1)}, A)$ is a version of $P(\tilde{x}_n \in A \mid \tilde{x}_{(n-1)})$ for every A in $\mathcal{X}$. In Bayesian terms, P_n is called *predictive p.d.* for $\tilde{x}_n$ given $\tilde{x}_{(n-1)}$ w.r.t. P. In view of this definition, given $(X^\infty, \mathcal{X}^\infty, P)$, a predictive p.d. for $\tilde{x}_n$ given $\tilde{x}_{(n-1)}$ w.r.t. P needn't exist [cf., e.g., (4) in Section 48 of Halmos (1950)]. It does exist, however, for every n, if X and $\mathcal{X}$ meet suitable conditions which include, for example, the case when X is a Polish space; cf. Section 10.29 in Hoffmann-Jørgensen (1994). On the other hand, it is possible to construct a probability measure P on $(X^\infty, \mathcal{X}^\infty)$, without requiring any special condition for the measurable space $(X, \mathcal{X})$, in such a way that a *given* sequence of transition probabilities with respect to $X^{n-1} \times \mathcal{X}$ $(n = 2, 3, \dots)$ represents a sequence of predictive p.d.'s w.r.t. P. This fact plays a leading role in the general setting of the operational point of view according to which predictive form is primary, i.e. predictive p.d.'s are thought of as primary constituents of any attempt at modelling and reporting uncertainty. Specifically, a straightforward translation into statistical terms of a basic theorem due to Ionescu Tulcea [cf., e.g., Section V-1 of Neveu (1980)] gives

PROPOSITION 3.1. *If P_1 is a probability measure on $(X, \mathcal{X})$ and P_n is a transition probability w.r.t. $X^{n-1} \times \mathcal{X}$ for every $n = 2, 3, \dots$, then there is a unique p.d. P for $\tilde{x}$ such that P_1 is the p.d. of $\tilde{x}_1$ and P_n is a predictive p.d. for $\tilde{x}_n$ given $\tilde{x}_{(n-1)}$, w.r.t. P, for every $n = 2, 3, \dots$.*

Under what circumstances is the p.d. P in Proposition 3.1 exchangeable? In order to answer this question, it is worth examining the inverse problem, i.e. investigating significant properties of predictive p.d's deduced from an exchangeable P.

PROPOSITION 3.2. *Let P be an exchangeable probability measure on $(X^\infty, \mathcal{X}^\infty)$ where $\mathcal{X}$ is assumed to be countably generated. Moreover, suppose that there is a predictive p.d. P_n for $\tilde{x}_n$ given $\tilde{x}_{(n-1)}$ for every $n = 2, 3, \dots$. Then there is a set N in $\mathcal{X}^\infty$ such that $P(N) = 0$ and, for each $x \in N^c$:*

(a) *$P_n(x_{(n-1)}, A) = P_n((x_{\pi(1)}, \dots, x_{\pi(n-1)}), A)$ holds true for every $n = 2, 3, \dots$, every A in $\mathcal{X}$ and every permutation π of $(1, \dots, n-1)$;*

(b) *$\int_B P_{n+1}(x_{(n)}, A) P_n(x_{(n-1)}, dx_n) = \int_A P_{n+1}(x_{(n)}, B) P_n(x_{(n-1)}, dx_n)$ is true for every A, B in $\mathcal{X}$ and for every n in $\mathbb{N}$, with $P_1(x_{(0)}, \cdot) := P\tilde{x}_1^{-1}(\cdot)$.*

[To say that $\mathcal{X}$ is *countably generated* means that there is a countable sub-family $\mathcal{F}$ of $\mathcal{X}$ such that the σ-algebra generated by $\mathcal{F}$, $\sigma(\mathcal{F})$, coincides with $\mathcal{X}$. Without real loss of generality, $\mathcal{F}$ can be thought of as an algebra. If X is a separable metric space, then $\mathcal{B}(X)$ is countably generated.]

PROOF OF PROPOSITION 3.2. (a) Note that exchangeability entails

$$\begin{aligned}\int_{A_1\times\ldots\times A_{n-1}} & P_n(x_{(n-1)},A)P\tilde{x}^{-1}_{(n-1)}(dx_{(n-1)}) \\ &= P(\tilde{x}_{(n)} \in A_1\times\ldots\times A_{n-1}\times A) \\ &= P(\tilde{x}_{(n)} \in A_{\pi(1)}\times\ldots\times A_{\pi(n-1)}\times A) \\ &= \int_{A_{\pi(1)}\times\ldots\times A_{\pi(n-1)}} P_n(x_{(n-1)},A)P\tilde{x}^{-1}_{(n-1)}(dx_{(n-1)}) \\ &= \int_{A_1\times\ldots\times A_{n-1}} P_n((x_{\pi(1)},\ldots,x_{\pi(n-1)}),A)P\tilde{x}^{-1}_{(n-1)}(dx_{(n-1)}).\end{aligned}$$

Thus, there is $N(n,\pi,A) \in \mathcal{X}^\infty$ with $P(N(n,\pi,A)) = 0$ such that

$$P_n(x_{(n-1)},A) = P_n((x_{\pi(1)},\ldots,x_{\pi(n-1)}),A) \qquad \ldots(3.1)$$

is true for every x in $(N(n,\pi,A))^c$. Hence, if $\mathcal{F}$ is a countable sub-algebra of $\mathcal{X}$ and n is a fixed integer ≥ 2, (3.1) is true for every n, every permutation π of $(1,\ldots,n-1)$, every A in $\mathcal{F}$ and every x in $N_1^c := (\cup_n \cup_\pi \cup_{A\in\mathcal{F}} N(n,\pi,A))^c$, where $P(N_1) = 0$.

(b) For every C in $\mathcal{X}^{n-1}$,

$$\begin{aligned}\int_C \left\{\int_B P_{n+1}(x_{(n)},A)P_n(x_{(n-1)},dx_n)\right\} & P\tilde{x}^{-1}_{(n-1)}(dx_{(n-1)}) \\ &= P(\tilde{x}_{(n+1)} \in C\times B\times A) \\ &= P(\tilde{x}_{(n+1)} \in C\times A\times B) \\ = \int_C \left\{\int_A P_{n+1}(x_{(n)},B)P_n(x_{(n-1)},dx_n)\right\} & P\tilde{x}^{-1}_{(n-1)}(dx_{(n-1)}).\end{aligned}$$

Hence there is $N(n,A,B) \in \mathcal{X}^\infty$ with $P(N(n,A,B)) = 0$ such that the desired property holds for every x in $N(n,A,B)$. Therefore, (b) holds true for every x in $N_2^c := (\cup_n \cup_{(A,B)\in\mathcal{F}^2} N(n,A,B))^c$, and for every $A\times B$ in the product algebra $\mathcal{F}^2$, where $P(N_2) = 0$.

To complete the proof, define N to be $N_1 \cup N_2$ and observe that, for any x in N^c, the previous statements extend from $\mathcal{F}$ and $\mathcal{F}^2$ to $\sigma(\mathcal{F})$ and $\sigma(\mathcal{F}^2)$, respectively, by the Carathéodory extension theorem. □

The following examples show that there is generally no relation between conditions (a) and (b) of Proposition 3.2.

EXAMPLE 3.1. Let $X = \{0,1\}$ and define P by

$$P(\tilde{x}_{(n)} = x_{(n)}) = p(x_1)p(x_1;x_2)\ldots p(x_{(n-1)};x_n),$$

where $x_{(n)} \in X^n$ for $n = 2, 3, \ldots$, with

$$\begin{aligned} p(0) = p(x_1; 0) &= \alpha \in (0,1) \\ \text{and } p(x_{(k)}; 1) &= \sum_{i=1}^{k} \frac{x_i}{k} \end{aligned}$$

for every $x_{(k)} \in X^k$ and $k = 2, 3, \ldots$.
P is, by definition, consistent with condition (a) of Proposition 3.2, but it does not satisfy condition (b). Indeed,

$$\begin{aligned} P(\tilde{x}_2 = 0, \tilde{x}_3 = 1 \mid \tilde{x}_1 = 0) &= 0, \\ P(\tilde{x}_2 = 1, \tilde{x}_3 = 0 \mid \tilde{x}_1 = 0) &= \tfrac{(1-\alpha)}{2}. \end{aligned}$$

Example 3.2. Let $X = \{\alpha, 1-\alpha\}$, with $\alpha \in (0,1)$. Define P by

$$\begin{aligned} P(\tilde{x}_1 = \alpha) &= \frac{1}{2} \\ P_{n+1}(x_{(n)}, \{\alpha\}) &= x_1. \end{aligned}$$

P is consistent with condition (b) of Proposition 3.2. In fact,

$$P(\tilde{x}_1 = \alpha, \tilde{x}_2 = 1-\alpha) = \tfrac{1}{2}(1-\alpha) = P(\tilde{x}_1 = 1-\alpha, \tilde{x}_2 = \alpha)$$

and

$$\begin{aligned} P(\tilde{x}_{n+1} = \alpha, \tilde{x}_{n+2} &= 1-\alpha \mid \tilde{x}_1, \ldots \tilde{x}_n) \\ &= \tilde{x}_1(1-\tilde{x}_1) \\ &= P(\tilde{x}_{n+1} = 1-\alpha, \tilde{x}_{n+2} = \alpha \mid \tilde{x}_1, \ldots \tilde{x}_n). \end{aligned}$$

On the other hand, condition (a) of Proposition 3.2 does not hold.

In the next proposition it is proved that (a) and (b) of Proposition 3.2 are sufficient conditions in order that a sequence of predictive p.d.'s can be consistent with an exchangeable P. Such a proposition generalizes an analogous statement proved by de Finetti for the particular case when $X = \{-1, 1\}$; cf. de Finetti (1952).

Theorem 3.1. *Let P be a probability measure on $(X^\infty, \mathcal{X}^\infty)$ w.r.t. which: the transition probability P_n is a predictive p.d. for $\tilde{x}_n$ given $\tilde{x}_{(n-1)}$ $(n = 2, 3, \ldots)$, and the probability measure P_1 on $(X, \mathcal{X})$ coincides with $P\tilde{x}_1^{-1}$. Moreover, assume that $(P_n)_{n\geq 1}$ satisfies conditions (a)-(b) of Proposition 3.2 for every x in N^c where $N \in \mathcal{X}^\infty$ and $P(N) = 0$. Then, P is exchangeable.*

Before proving the theorem, recall that $(\tilde{x}_1, \ldots \tilde{x}_n)$ and $(\tilde{x}_{\pi(1)}, \ldots, \tilde{x}_{\pi(n)})$ have identical p.d.'s, when π is a permutation of $(1, \ldots, n)$, if and only if

$$P(\tilde{x}_{(n)} \in A_{\pi^{-1}(1)} \times \ldots \times A_{\pi^{-1}(n)}) = P(\tilde{x}_{(n)} \in A_1 \times \ldots \times A_n) \qquad \ldots (3.2)$$

holds for every $A_1, \ldots, A_n$ in $\mathcal{X}$, and for the inverse π^{-1} of π. This fact can be proved by means of a monotone class argument.

PROOF OF Theorem 3.1. Let $\sigma(1, 2, \ldots) = (1, \ldots, j-1, j+1, j, j+2, \ldots)$.

$$\begin{aligned} &P(\tilde{x}_{(j+1)} \in A_1 \times \ldots \times A_{j-1} \times A_{j+1} \times A_j) \\ &= \int_{A_1} P\tilde{x}_1^{-1}(dx_1) \ldots \int_{A_{j-1}} P_{j-1}(x_{(j-2)}, dx_{j-1}) \int_{A_{j+1}} P_j(x_{(j-1)}, dx_j) P_{j+1}(x_{(j)}, A_j) \\ &= \int_{A_1} P\tilde{x}_1^{-1}(dx_1) \ldots \int_{A_{j-1}} P_{j-1}(x_{(j-2)}, dx_{j-1}) \int_{A_j} P_j(x_{(j-1)}, dx_j) P_{j+1}(x_{(j)}, A_{j+1}) \end{aligned}$$

from condition (b) of Proposition 3.2. This relation, together with (3.2), implies that $(\tilde{x}_{\pi_{(1)}}, \ldots, \tilde{x}_{\pi_{(j+1)}})$ and $(\tilde{x}_1, \ldots, \tilde{x}_{j+1})$ have the same p.d. when π is the inverse of σ.

Moreover, for every $n \geq j+2$, it follows from the first part of this proof that

$$\begin{aligned} &P(\tilde{x}_{(n)} \in A_{\sigma(1)} \times \ldots \times A_{\sigma(n)}) \\ &\quad= \int_{A_{\sigma(1)} \times \ldots \times A_{\sigma(j+1)}} \left[\int_{A_{j+2}} P_{j+2}(x_{(j+1)}, dx_{j+2}) \ldots \right. \\ &\quad\quad \left. \ldots \int_{A_n} P_n(x_{(n-1)}, dx_n) \right] P\tilde{x}_{(j+1)}^{-1}(dx_{(j+1)}) \\ &\quad= \int_{A_{\sigma(1)} \times \ldots \times A_{\sigma(j+1)}} \left[\int_{A_{j+2}} P_{j+2}(x_{(j+1)}, dx_{j+2}) \ldots \right. \\ &\quad\quad \left. \ldots \int_{A_n} P_n(x_{(n-1)}, dx_n) \right] P(\tilde{x}_{\pi(1)}, \ldots, \tilde{x}_{\pi_{(j+1)}})^{-1}(dx_1 \ldots dx_{j+1}) \\ &\quad= \int_{A_{\sigma(1)} \times \ldots \times A_{\sigma(j+1)}} \left[\int_{A_{j+2}} P_{j+2}(x_{(j+1)}, dx_{j+2}) \ldots \right. \\ &\quad\quad \left. \ldots \int_{A_n} P_n(x_{(n-1)}, dx_n) \right] P\tilde{x}_{(j+1)}^{-1}(dx_{\sigma_{(1)}} \ldots dx_{\sigma(j+1)}) \end{aligned}$$

which, by condition (a) of Proposition 3.2, can be written in the form

$$\begin{aligned} &\int_{A_{\sigma(1)} \times \ldots \times A_{\sigma(j+1)}} \left[\int_{A_{j+2}} P_{j+2}((x_{\sigma(1)}, \ldots, x_{\sigma(j+1)}), dx_{j+2}) \ldots \right. \\ &\left. \ldots \int_{A_n} P_n((x_{\sigma(1)}, \ldots x_{\sigma(n-1)}), dx_n) \right] P\tilde{x}_{(j+1)}^{-1}(dx_{\sigma_{(1)}}, \ldots, dx_{\sigma(j+1)}) \\ &\qquad\qquad = P(\tilde{x}_{(n)} \in A_1 \times \ldots \times A_n). \end{aligned}$$

This proves that $(\tilde{x}_1, \tilde{x}_2, \ldots)$ and $(\tilde{x}_{\pi(1)}, \tilde{x}_{\pi(2)}, \ldots)$ have identical p.d.. Since any finite permutation can be derived from $(1, 2, \ldots)$ by a finite number of exchanges of adjacent elements, the previous assertion extends to any finite permutation. Thus, P is exchangeable. □

REMARK. There is an equivalent formulation of condition (a) of Proposition 3.2 which highlights the role of the empirical measure $e_n(\tilde{x}_{(n)})$ in reducing the complexity of the observable element $\tilde{x}_{(n)}$, w.r.t. the predictive p.d. of $\tilde{x}_{n+1}$. In the same setting as in Theorem 3.1, let x and y be elements of N^c such that $(y_1, \dots y_{n-1})$ is a permutation of $(x_1, \dots x_{n-1})$. Then, by condition (a) of Proposition 3.2, $P_n(\tilde{x}_{(n-1)}(x), \cdot) = P_n(\tilde{x}_{(n-1)}(y), \cdot)$ and this is tantamount to saying that

$$P_n(\tilde{x}_{(n-1)}(x), \cdot) = f_n(e_{n-1}(\tilde{x}_{(n-1)}(x)), \cdot)$$

for x in N^c. Conversely, this equality entails condition (a) of Proposition 3.2.

Example 3.3. Theorem 3.1 can be exploited to prove that

$$\begin{aligned} P_n(x_{(n-1)}, A) &:= \int_A \alpha(y) \frac{\beta(h_{(n-1)}(x_1, \dots x_{n-1}), n-1)}{\beta(h_n(x_1, \dots x_{n-1}, y), n)} \mu(dy) \quad (A \in \mathcal{X},\ n = 2, 3, \dots) \\ P_1(A) &:= \int_A \alpha(y) \frac{1}{\beta(h_1(y), 1)} \mu(dy) \qquad\qquad (A \in \mathcal{X}) \end{aligned}$$

form a system of predictive p.d.'s of an exchangeable law provided h_n is a symmetric function for any n, and α, β, μ are chosen in such a way that P_1 and P_n turn out to be probability measures for every n and $x_{(n-1)}$. From the theory in Diaconis and Ylvisaker (1979), general exponential families together with their conjugate priors yield predictive p.d.'s of the above type. As an example of *non* exponential model with predictive p.d.'s of that very same type, consider the uniform distribution on $[0, \tilde{\beta}]$ where $\tilde{\beta}$ is a r.v. with Pareto law. See next Example (7.7).

Theorem 3.1 and Proposition 3.2 can be used to insist on the following point: adoption of improper priors can give rise to incompatibility with (σ-additive) exchangeability of observations.

Example 3.4. In many textbooks we read that

$$P_{n+1}(x_{(n)}, \{x\}) = \frac{1}{x!} \frac{n^{\frac{1}{2} + \sum_{i=1}^n x_i}}{\Gamma(\frac{1}{2} + \sum_{i=1}^n x_i)} \frac{\Gamma(\frac{1}{2} + \sum_{i=1}^n x_i + x)}{(n+1)^{\frac{1}{2} + \sum_{i=1}^n x_i + x}}$$

$(n = 1, 2, \dots; x, x_1, \dots \in \{0, 1, \dots\})$ is the predictive distribution of $\tilde{x}_{n+1}$ when the $\tilde{x}_k$ are i.i.d. Poisson observations conditional on the expectation $\tilde{\theta}$, and the p.d. of $\tilde{\theta}$ is "improper" with density $\propto (1/\sqrt{\theta}) \mathbf{1}_{(0,+\infty)}(\theta)$. To see that this statement is incorrect (within the usual measure-theoretic approach), observe that, from Theorem 3.1 and the expression of P_{n+1}, the following equality

$$P_1\{x_1\} \frac{1}{\Gamma(\frac{1}{2} + x_1) x_2!} = P_1\{x_2\} \frac{1}{\Gamma(\frac{1}{2} + x_2) x_1!}$$

has to hold for any (x_1, x_2) in $\{0, 1, \dots\}^2$, in order that the $\tilde{x}_n$ may be conditionally i.i.d., that is exchangeable. Whence,

$$P_1\{x\} = P_1\{0\} \frac{\Gamma(x + \frac{1}{2})}{\sqrt{\pi} \Gamma(x+1)} \qquad (x = 0, 1, \dots),$$

and $\sum_{x \geq 0} P_1\{x\} = 0$ or $+\infty$ according to $P_1\{0\} = 0$ or > 0. In both cases P_1 is not a probability measure. Thus, there is no probability p.d. for $\tilde{x}_1$ which, together with the above predictive p.d.'s, makes the $\tilde{x}_n$ exchangeable r.v.'s. The problem has at least one solution within the class of all finitely additive probabilities.

4. Application of the Previous Results

If (R2) holds and $A \in \mathcal{X}$, then $\tilde{p}(A)$ is a version of $P(\tilde{x}_n \in A \mid \tilde{x}_{(n-1)}, \tilde{p})$. Hence

$$P(\tilde{x}_n \in A \mid \tilde{x}_{(n-1)}) = E(\tilde{p}(A) \mid \tilde{x}_{(n-1)}). \qquad \ldots (4.1)$$

This relation comes in handy when expressing predictive p.d.'s.

EXAMPLE 4.1. When $\tilde{p}$ has the *Ferguson-Dirichlet* prior with parameter α, by a well-known closure property under sampling property of that prior [cf. Ferguson (1973)] and by (4.1), it is immediate to check that for A in $\mathcal{X}$, $n \geq 2$ and $c := \alpha(X) > 0$,

$$P_n^{(1)}(\tilde{x}_{(n-1)}, A) := \frac{\alpha(A) + \sum_{i=1}^{n-1} \delta_{\tilde{x}_i}}{n - 1 + c}$$

is a predictive p.d. for $\tilde{x}_n$ given $\tilde{x}_{(n-1)}$. Moreover,

$$P_1^{(1)} = \frac{\alpha}{c}$$

turns out to be the p.d. of $\tilde{x}_1$.

EXAMPLE 4.2. Let $\pi_1, \pi_2, \ldots$ be a binary tree of partitions of X,

$$\pi_1 = \{B_0, B_1\}, \pi_2 = \{B_{00}, B_{01}, B_{10}, B_{11}\}, \ldots$$

and $(\alpha_0, \alpha_1, \alpha_{00}, \alpha_{01}, \ldots)$ a sequence of nonnegative numbers. For all $\epsilon = (\epsilon_1, \ldots, \epsilon_m)$ in $\{0,1\}^m$, with m in $\mathbb{N}$, for all $n \geq 2$, define

$$P_n^{(2)}(\tilde{x}_{(n-1)}, B_\epsilon)$$
$$= \frac{\alpha_{\epsilon_1} + n_{\epsilon_1}}{\alpha_0 + \alpha_1 + n - 1} \frac{\alpha_{\epsilon_1\epsilon_2} + n_{\epsilon_1\epsilon_2}}{\alpha_{\epsilon_1 0} + \alpha_{\epsilon_1 1} + n_{\epsilon_1}} \cdots \frac{\alpha_{\epsilon_1\ldots\epsilon_m} + n_{\epsilon_1\ldots\epsilon_m}}{\alpha_{\epsilon_1\ldots\epsilon_{m-1}1} + \alpha_{\epsilon_1\ldots\epsilon_{m-1}0} + n_{\epsilon_1\ldots\epsilon_{m-1}}},$$

if all the denominators are strictly positive and 0 otherwise, where n_ϵ is the number of elements, from $\tilde{x}_{(n-1)}$, in B_ϵ. If $\tilde{p}$ has the *Pólya tree* prior with parameter $(\pi_1, \pi_2 \ldots, \alpha_0, \alpha_1, \ldots)$ [cf. Lavine (1992); cf. also Mauldin, Sudderth and Williams (1992)], then $P_n^{(2)}$ is the predictive p.d. for $\tilde{x}_n$ given $\tilde{x}_{(n-1)}$. Analogously, the p.d. of $\tilde{x}_1$ is

$$P_1^{(2)}(B_\epsilon) = \frac{\alpha_{\epsilon_1}}{\alpha_0 + \alpha_1} \cdots \frac{\alpha_{\epsilon_1\ldots\epsilon_m}}{\alpha_{\epsilon_1\ldots\epsilon_{m-1}1} + \alpha_{\epsilon_1\ldots\epsilon_{m-1}0}},$$

if all the denominators are strictly positive, and 0 otherwise.

In the general setting of the operational point of view mentioned in the previous section, the following problem is of interest: "If $P_n^{(i)}$ is the p.d. of $\tilde{x}_n$ given $\tilde{x}_{(n-1)}$ w.r.t. $P^{(i)}$ for $i = 1, 2$ and every $n = 2, 3, \ldots$, *need* this $P^{(i)}$ be exchangeable? If the answer is yes and (R2) holds, *need* the law of $\tilde{p}$ be Ferguson-Dirichlet, Pólya tree, according to i=1,2?"

Some authors have addressed only the latter part of the issue by means of arguments susceptible of some significant simplifications, as suggested by the content of Section 3; see, for example, Regazzini (1978), Lo (1991), Walker and Muliere (1997).

Firstly, as far as the former part of the question is concerned, there is a unique p.d. $P^{(i)}$ for $\tilde{x}$, w.r.t. which $P_1^{(i)}$ and $P_n^{(i)}$ are the p.d. of $\tilde{x}_1$ and the predictive p.d. of $\tilde{x}_n$ given $\tilde{x}_{(n-1)}$, respectively ($n \geq 2$); cf. Proposition 3.1. Moreover, it is almost immediate to check that $(P_n^{(i)})_{n\geq 1}$ satisfies conditions (a)-(b) of Proposition 3.2 for every x in X^∞, so that $P^{(i)}$ is exchangeable in view of Theorem 3.1, for $i = 1, 2$. Finally, if (R2) holds, the uniqueness of de Finetti's measure is conducive to addressing the latter part of the issue in a straightforward way.

A typical exploitation of Theorem 3.1 appears in

EXAMPLE 4.3. Assume that $(P_n)_{n\geq 1}$ is a sequence of Gaussian p.d.'s such that: $E(\tilde{x}_1) = 0$, $Var(\tilde{x}_1) = 1$; $E(\tilde{x}_{n+1} \mid \tilde{x}_{(n)}) = m_n(\tilde{s}_n)$, $Var(\tilde{x}_{n+1} \mid \tilde{x}_{(n)}) = \sigma_n^2$ for any $n \geq 1$, with $\tilde{s}_n := \sum_{k=1}^n \tilde{x}_k$ and σ_n^2 being a strictly positive number independent of $\tilde{x}_{(n)}$. The problem is that of determining m_n and σ_n^2 in such a way that the $\tilde{x}_n$ turn out to be exchangeable. Clearly the p.d.'s P_n satisfy condition (a) of Proposition 3.2. Hence, in view of condition (b) of that Proposition, m_n and σ_n^2 have to obey

$$\frac{1}{\sigma_n^2}[\{y - m_n(s_{n-1} + x)\}^2 - \{x - m_n(s_{n-1} + y)\}^2]$$

$$= \frac{1}{\sigma_{n-1}^2}[\{y - m_{n-1}(s_{n-1})\}^2 - \{x - m_{n-1}(s_{n-1})\}^2] \qquad \ldots (4.2)$$

for every x, y in $\mathbb{R}$, s_{n-1} in $\mathbb{R}$ with $s_0 := 0$, $m_0 := 0$, $\sigma_0^2 := 1$ and $n \geq 1$. Whence, for $n = 1$, (4.2) takes the form

$$[(y - m_1(x))^2 - (x - m_1(y))^2] = \sigma_1^2(y^2 - x^2) \qquad \ldots (4.3)$$

which, solving for $m_1(x)^2$ with $y = 0$, gives $m_1(x)^2 = (1 - \sigma_1^2)x^2 + \gamma^2 - 2\gamma x$, where $\gamma := m_1(0)$. By substituting this expression for $m_1(x)^2$ on the left-hand-side of (4.3), we obtain

$$\frac{m_1(y) - m_1(x)}{y - x} = \frac{m_1(x) - m_1(0)}{x} \qquad x, y \in \mathbb{R},\ y \neq x.$$

Thus, there is ρ in $\mathbb{R}$ for which $m_1(x) = \rho x + \gamma$ and, by substituting this expression on the left of (4.3), we have $(y + x)(1 - \rho^2 - \sigma_1^2) = 2\gamma(1 + \rho)$. Thus, since $\sigma_1^2 > 0$, we can write $\gamma = 0$ and $\sigma_1^2 = 1 - \rho^2$ with $\rho^2 < 1$.

Next, use induction on n to prove that $m_n(0) = 0$ for every $n \geq 1$. We have proved that this is true for $n = 1$. Assume it holds for every $1 \leq k < n$. Then (4.2), with $s_{n-1} = 0$, yields the same as (4.3) where σ_1^2 is replaced by $\sigma_n^2/\sigma_{n-1}^2$. Thus, $m_n(0) = 0$ and $m_n(x) = \rho_n x$ for some ρ_n with $1 - \rho_n^2 = \sigma_n^2/\sigma_{n-1}^2$.

Finally, from exchangeability, $0 \leq E(\tilde{x}_i\tilde{x}_j) = E(E(\tilde{x}_2 \mid \tilde{x}_1)\tilde{x}_1) = \rho$ for any $i \neq j$ and $n\rho = E(\tilde{x}_{n+1}\tilde{s}_n) = E(E(\tilde{x}_{n+1} \mid \tilde{s}_n)\tilde{s}_n) = \rho_n E(\tilde{s}_n^2) = n\rho_n\{1 + (n-1)\rho\}$, i.e.

$$m_n(x) = \frac{\rho}{1 + (n-1)\rho}x \qquad (x \in \mathbb{R},\ n \in \mathbb{N}).$$

Furthermore, $1 = E(\tilde{x}_{n+1}^2) = E(E(\tilde{x}_{n+1}^2 \mid \tilde{s}_n)) = \sigma_n^2 + \rho_n^2 n\{1 + (n-1)\rho\}$, i.e.

$$\sigma_n^2 = \frac{(1-\rho)(1+n\rho)}{1+(n-1)\rho} \qquad (n \in \mathbb{N}).$$

5. **Predictive Sufficiency**

In Section 3 qualitative conditions to obtain an exchangeable P have been indicated, but this is yet far from a concrete specification of P as it is needed in actual applications. Nevertheless, in Section 4 it has been shown that there are simple and natural definitions of $(P_n)_{n\geq 1}$ which lead to a complete characterization of what is needed to implement the Bayesian paradigm. The present section deals with what is implied when one assumes that data can be reduced, without actually altering previsions, by means of suitable statistics. Hence, these statistics will be called predictive sufficient. To make precise the notion of predictive sufficiency, assume that X and T are Polish spaces and keep the notation introduced in the final part of Section 2. Define a statistic $\tilde{t}_{|D}$ to be *predictive sufficient* w.r.t P if, for any n in $\mathbb{N}$, there is a r.c.p.d. Q_n for $(\tilde{x}_i)_{i\geq n+1}$ given $\tilde{t}(e_n(\tilde{x}_{(n)}))$ such that $Q_n(\tilde{t}(e_n(\tilde{x}_{(n)})), A)$ is a version of $P((\tilde{x}_i)_{i\geq n+1} \in A \mid \tilde{x}_{(n)})$ for every A in $\mathcal{B}(X^\infty)$. The r.c.p.d. Q_n exists since X^∞ and T are Polish spaces.

Predictive sufficiency and its properties have been investigated in a number of writings among which : Campanino and Spizzichino (1981), Cifarelli and Regazzini (1980, 1981, 1982), Dawid (1982), Secchi (1987), Muliere and Secchi (1992). More recently, Bernardo and Smith have devoted Section 4.5 of their book to this subject; cf. Bernardo and Smith (1994). Related notions of sufficiency have been studied by Lauritzen (1984, 1988), Diaconis and Freedman (1984). These works are reviewed in Section 2.4 of Schervish (1995). Here, some basic properties of predictive sufficient statistics are stated precisely and proved in a somewhat general setting, i.e. when X and T are Polish spaces. The missing proofs are given in Fortini, Ladelli and Regazzini (1998).

PROPOSITION 5.1. *Let $\tilde{x}$ be the sequence of coordinate r.v.'s of X^∞. Then $\tilde{t}_{|D}$ is a predictive sufficient statistic w.r.t. P if, and only if, $\tilde{x}_{(n)}$ and $(\tilde{x}_i)_{i\geq n+1}$ are conditionally independent given $\tilde{t}(e_n(\tilde{x}_{(n)}))$, for every n in $\mathbb{N}$.*

Here, conditional independence is understood in the sense that there is a r.c.p.d. of $\tilde{x}_{(n)}$ given $\tilde{t}(e_n(\tilde{x}_{(n)}))$, say Q'_n, such that $Q'_n(\tilde{t}(e_n(\tilde{x}_{(n)})), A)Q_n(\tilde{t}(e_n(\tilde{x}_{(n)})), B)$ is a version of $P(\tilde{x} \in A \times B \mid \tilde{t}(e_n(\tilde{x}_{(n)})))$ for every A in $\mathcal{B}(X^n)$ and B in $\mathcal{B}(X^\infty)$.

The previous proposition is useful to investigate the connections between predictive and Fisher's classical sufficiency.

LEMMA 5.1. *Let P be an exchangeable probability measure on $(X^\infty, \mathcal{B}(X^\infty))$ and let $\tilde{t}_{|D}$ be a predictive sufficient statistic w.r.t. P. Then $Q'_n(\tilde{t}(e_n(\tilde{x}_{(n)})), A)$ is a version of $P(\tilde{x}_{(n)} \in A \mid \tilde{p}, \tilde{t}(e_n(\tilde{x}_{(n)})))$ for every A in $\mathcal{B}(X^n)$ and n in $\mathbb{N}$, where $\tilde{p}$ has the same meaning as in (R2).*

THEOREM 5.1. *Under the same conditions as in Lemma (5.1), there is a set N in $\mathcal{B}(M_1)$ with $\nu(N) = 0$* [ν is the de Finetti measure as in (R3)] *such that, for every n in $\mathbb{N}$, $\tilde{t}_{|D} \circ e_n$ is classical sufficient for $\{p^n : p \in N^c\}$, where p^n is the n-th power probability.*

PROOF. Define $R_n : M_1 \times \mathcal{B}(T) \to [0,1]$ by

$$R_n(p, B) = p^\infty(\tilde{t}(e_n(\tilde{x}_{(n)})) \in B)$$

for every p in M_1 and B in $\mathcal{B}(T)$. Thus, for any C in $\mathcal{B}(M_1)$ and A in $\mathcal{B}(X^n)$,

$$\begin{aligned} \int_C \left[\int_B Q'_n(t, A) \quad R_n(p, dt) \right] \nu(dp) & \\ = P(\tilde{x}_{(n)} \in A, \tilde{t}(e_n(\tilde{x}_{(n)})) \in B, \tilde{p} \in C) & \qquad \text{[from Lemma 5.1]} \\ = \int_C p^\infty(\tilde{x}_{(n)} \in A, \tilde{t}(e_n(\tilde{x}_{(n)})) \in B) \nu(dp) & \qquad \text{[from } (R3)] \end{aligned}$$

Therefore,

$$\int_B Q'_n(t, A) R_n(p, dt) = p^\infty(\tilde{x}_{(n)} \in A, \tilde{t}(e_n(\tilde{x}_{(n)})) \in B) \qquad \ldots (5.1)$$

holds true for any p in $N(A, B, n)^c$ where $N(A, B, n) \in \mathcal{B}(M_1)$ and $\nu(N(A, B, n)) = 0$. By separability of X^n and T, the standard argument used, e.g., to prove Proposition 3.2, entails the existence of a set $N \in \mathcal{B}(M_1)$ with $\nu(N) = 0$, such that (5.1) remains valid for every p in N^c and for every A in $\mathcal{B}(X^n)$ and B in $\mathcal{B}(T)$. Hence, for any p in N^c, $Q'_n(\tilde{t}(e_n(\tilde{x}_{(n)})), A)$ is a version of the conditional probability of $\{\tilde{x}_{(n)} \in A\}$ given $\tilde{t}(e_n(\tilde{x}_{(n)}))$, when p^∞ is the p.d. of $\tilde{x}$, i.e. p^n is the p.d. of $\tilde{x}_{(n)}$. Since Q'_n is constant with respect to $p \in N^c$, then $\tilde{t}(e_n(\tilde{x}_{(n)}))$ is, by definition, a classical sufficient statistic for $\{p^n : p \in N^c\}$. □

Theorem 5.1 states that predictive sufficiency implies classical sufficiency - when P is exchangeable - for ν-almost all p, where ν is the de Finetti measure of P. As far as the converse is concerned, the following proposition holds true.

THEOREM 5.2 *Assume that the sequence $\tilde{x}$ of coordinate r.v.'s of X^∞ is exchangeable w.r.t. P. Moreover, let $\tilde{t}(e_n(\tilde{x}_{(n)}))$ be a classical sufficient statistic for $\{p^n : p \in N^c\}$, $n = 1, 2, \ldots$, where N is an element of $\mathcal{B}(M_1)$ with $\nu(N) = 0$. Then, $\tilde{t}_{|D}$ is predictive sufficient w.r.t. P.*

When P is exchangeable and $\tilde{p}$ is as in (R2), then $\tilde{t}_{|D}$ is called *Bayesian sufficient* w.r.t. P if, for each n in $\mathbb{N}$, there is a r.c.p.d., B_n, for $\tilde{p}$ given $\tilde{t}(e_n(\tilde{x}_{(n)}))$ such that $B_n(\tilde{t}(e_n(\tilde{x}_{(n)})), C)$ is a version of $P(\tilde{p} \in C \mid \tilde{x}_{(n)})$ for every C in $\mathcal{B}(M_1)$.

THEOREM 5.3. *Let the sequence $\tilde{x}$ of coordinate r.v.'s of X^∞ be exchangeable w.r.t. P. Then, w.r.t. P, $\tilde{t}_{|D}$ is Bayesian sufficient if, and only if, it is predictive sufficient.*

As far as relations between classical sufficiency and Bayesian sufficiency w.r.t. a *single* P are concerned, see Letta (1981). It is worth recalling that Bayesian sufficiency is usually understood as dependence of the posterior on a specific statistic *for every prior* on a fixed parameter space; see, e.g., Schervish (1995).

The last result of the present section is about a remarkable asymptotic property of the predictive distribution

$$Q''_n(\tilde{t}(e_n(\tilde{x}_{(n)})), A) := Q_n(\tilde{t}(e_n(\tilde{x}_{(n)})), A \times X^\infty) \qquad (A \in \mathcal{B}(X)).$$

PROPOSITION 5.2. *Let $\tilde{t}_{|D}$ be a predictive sufficient statistic w.r.t. P, where P is an exchangeable probability on $(X^\infty, \mathcal{B}(X^\infty))$. Then*

$$Q''_n(\tilde{t}(e_n(\tilde{x}_{(n)})), \cdot) \Rightarrow \tilde{p} \qquad \text{a.s.} - P.$$

The proposition can be proved starting from a result in Doob (1949).

6. Characterization of Exponential Families

In the classical setting of i.i.d. observations $\tilde{x}_1, \ldots, \tilde{x}_n$, there are precise statements of this fact : if $\sum_{i=1}^n \tilde{x}_i$ is a sufficient (in the classical sense) statistic for a parametric model, then the p.d. of each $\tilde{x}_i$ belongs to an exponential family. These characterizations differ from one another in the regularity conditions that are assumed, in addition to sufficiency, in order to formulate the problem in a reasonable mathematical form. The present section deals with a similar type of question, when classical sufficiency is replaced with predictive sufficiency and, consistently with this, regularity conditions are given in terms of predictive p.d.'s.

Here X is a Borel subset of $\mathbb{R}^d$ and P is supposed to be exchangeable and such that

(I1) $\tilde{t}_{|D} = \int_X x de_n$ $(e_n \in D)$ is predictive sufficient w.r.t. P, i.e. there is a r.c.p.d. Q_n for $(\tilde{x}_k)_{k \geq n+1}$ given $\sum_{i=1}^n \tilde{x}_i/n$ such that $Q_n(\sum_{i=1}^n \tilde{x}_i/n, A)$ is a version of $P((\tilde{x}_k)_{k \geq n+1} \in A \mid \tilde{x}_{(n)})$ for every A in $\mathcal{B}(X^\infty)$ and n in $\mathbb{N}$.

Then, by Proposition 5.2,

$$Q''_n(\sum_{i=1}^n \tilde{x}_i/n, \cdot) \Rightarrow \tilde{p}(\cdot) \qquad \ldots (6.1)$$

a.s.-P. Thus there is N in $\mathcal{B}(X^\infty)$, with $P(N) = 0$, such that $Q''_n(\sum_{i=1}^n x_i/n, \cdot) \Rightarrow p_x(\cdot)$, for every x in N^c and, by Theorem 5.1, $\sum_{i=1}^n x_i/n$ is classical sufficient for $\{p_x^n : x \in N^c\}$, where p_x denotes the value of $\tilde{p}$ at x.

Moreover, assume the following conditions:

(I2) For every $\epsilon > 0$ and x in N^c there is $\eta_x(\epsilon)$ for which

$$\liminf_n Q_n(\tilde{t}(e_n(\tilde{x}_{(n)})), G \times X^\infty) \leq \epsilon$$

whenever G is any open subset of X such that $\lambda(G) \leq \eta_x(\epsilon)$, λ being the restriction of the Lebesgue measure to $\mathcal{B}(X)$.

(I3) $\lambda(F) = 0$ for every closed subset F of X such that

$$\lim_n Q_n(\tilde{t}(e_n(\tilde{x}_{(n)})(x)), F \times X^\infty) = 0$$

for some x in N^c.

$(I2)-(I3)$ entail

$$p_x \text{ and } \lambda \text{ are equivalent measures on } (X, \mathcal{B}(X)) \text{ for every } x \text{ in } N^c. \qquad \ldots(6.2)$$

[To prove (6.2), choose x in N^c and to each $A_1 \in \mathcal{B}(X)$ with $\lambda(A_1) \leq \eta_x(\epsilon)/2$ associate an open set $G_{A_1} \supset A_1$ such that $\lambda(G_{A_1}) \leq \eta_x(\epsilon)$. Then, from (6.1) and Theorem 2.1 in Billingsley (1968),

$$p_x(A_1) \leq p_x(G_{A_1}) \leq \liminf_n Q_n(\tilde{t}(e_n(\tilde{x}_{(n)})), G_{A_1} \times X^\infty) \leq \epsilon \qquad \text{[from (I2)]},$$

i.e. p_x is absolutely continuous w.r.t. λ.

Conversely suppose that $p_x(A_1) = 0$ for some x in N^c. Then, for every closed subset F_{A_1} of A_1, by Theorem 2.1 in Billingsley (1968),

$$0 \leq \limsup_n Q_n(\tilde{t}(e_n(\tilde{x}_n(x))), F_{A_1} \times X^\infty) \leq p_x(F_{A_1}) = 0$$

which, by means of $(I3)$, yields $\lambda(A_1) = 0$.]

Plainly, (6.2) implies that p_x is absolutely continuous w.r.t. p_y, whenever x and y are in N^c. It is now possible to state the main result of this section, i.e. that the de Finetti measure ν of P is supported by an exponential family.

Theorem 6.1. *Let X be a Borel subset of $\mathbb{R}^d$ and P be any exchangeable law on $(X^\infty, \mathcal{B}(X^\infty))$, satisfying conditions $(I1)-(I3)$. Then, there are measurable functions: $\alpha : M_1 \to \mathbb{R}$, $\beta : M_1 \to \mathbb{R}^d$ and $\gamma : X \to (0, +\infty)$ such that for every $z \in X$*

$$l_x(z) := \gamma(z) \exp\{\alpha(p_x) + < \beta(p_x), z >\}$$

is a density function of p_x w.r.t. λ, for every x in N^c.

It follows from $\int_X l_x d\lambda = 1$ that $e^{\alpha(p_x)} = \rho(\beta(p_x))$. Hence, $\{p_x : x \in N^c\}$ reduces to a family of p.d.'s whose members are completely determined by the assignment of the d-dimensional parameter $\tilde{\beta} = \beta(\tilde{p})$. Therefore, in de Finetti's representation (R3) of P, the measure ν can be replaced with $\nu_{\tilde{\beta}} := \nu \circ \tilde{\beta}^{-1}$, defined on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, to obtain, for every $B \in \mathcal{B}(X^\infty)$,

$$P(B) = \int_{\mathbb{R}^d} (Ex_\beta)^\infty(B) \nu_{\tilde{\beta}}(d\beta) \qquad \text{where,} \quad Ex_\beta(B) := \rho(\beta) \int_B \gamma(z) e^{<\beta, z>} dz.$$

This fact is of practical primary importance when P has actually to be specified.

Proof of Theorem 6.1. Let l_x^* be a density of p_x w.r.t. p_{x_0}, where x_0 is a fixed element of N^c and x varies in the same set. Then, by Theorem 5.1, (6.2) and a well-known result due to Halmos, Savage and Bahadur [cf., e.g., Lemma 2.24 in Schervish (1995)],

$$\prod_{i=1}^n l_x^*(z_i) = g_{p_x, n}(\sum_{i=1}^n z_i) \qquad (n = 1, 2, \ldots) \qquad \ldots(6.3)$$

holds true for λ^n-almost all $(z_1, \dots z_n)$ in X^n, and for a suitable measurable function $g_{p_x,n} : X^{(n)} \to (0, +\infty)$, where $X^{(1)} := X$, $X^{(n+1)} := X^{(n)} + X$. Due to Theorem 2.1 in Diaconis and Freedman (1990), (6.3) yields

$$l_x^*(z) = \exp\{\alpha(p_x) + < \beta(p_x), z >\}$$

for $x \in N^c$ and for λ-almost all z in X. Since $l_x^* = l_x / l_{x_0}$ (a.e.-λ), where l_y denotes a density of p_y w.r.t. λ, the thesis follows by putting $\gamma = l_{x_0}$. □

The following examples illustrate how particular exponential families could be specified.

EXAMPLE 6.1. (Gaussian model with known precision). Suppose that $\tilde{x}$ is a (potentially) infinite sequence of real-valued measurements of a physical property of a given substance. Assume that the elements of $\tilde{x}$ are exchangeable and that their p.d. P satisfies $(I1) - (I3)$. Hence, from Theorem 6.1, the elements of $\tilde{x}$ are conditionally i.i.d. given a real-valued r.v. $\tilde{\beta}$ with p.d. $Ex_{\tilde{\beta}}$. In this case, many individuals might think of the expected value of $Ex_{\tilde{\beta}}$ as the unknown value of the physical property of interest. Thus we can assume that they judge $Ex_{\tilde{\beta}}$ to be completely determined when the expected value of the above physical property is assigned. This justifies the identification of $\tilde{\beta}$ with some one-to-one function of the expected value of $Ex_{\tilde{\beta}}$. Such expectation, as it is well-known, coincides with $\{-\rho'(\beta)/\rho(\beta)\}$. Cf. Chapter 12 in Hoffmann-Jørgensen (1994). We wonder whether these conditions are consistent with the additional assumption that the above one-to-one function is the identity mapping, that is

$$-\frac{\rho'(\beta)}{\rho(\beta)} = \beta \qquad (\beta \in \mathbb{R}) \qquad \dots (6.4)$$

Integration of (6.4) together with the definition of Ex_β, gives

$$ce^{\frac{\beta^2}{2}} = \frac{1}{\rho(\beta)} = \int_{\mathbb{R}} \gamma(x) e^{\beta x} dx \qquad (\beta \in \mathbb{R})$$

which entails $\gamma(x) = c_1 \exp(-x^2/2)$.

Thus, the Gaussian family with known variance ($= 1$) and unknown mean is the sole statistical model which is consistent with $(I1) - (I3)$ and (6.4).

A complete specification of P requires assessment of $\nu_{\tilde{\beta}}$. Here, two different "strategies" - both of a predictive nature - are illustrated. The former aims at searching for any prior $\nu_{\tilde{\beta}}$ such that $E(\tilde{x}_{n+1} \mid \tilde{x}_{(n)})$ takes the form of a convex combination of $\frac{1}{n}\sum_{k=1}^n \tilde{x}_k$ and of $E(\tilde{x}_1)$, for every n. It is well-known that such property holds true if and only if

$$\nu_{\tilde{\beta}}(d\beta) \propto \exp\{-\frac{1}{2\sigma_0^2}(\beta - E(\tilde{x}_1))^2\} d\beta \qquad (\beta \in \mathbb{R}).$$

Cf. Theorem 3 in Diaconis Ylvisaker (1979).

Assume now that, in the light of additional information, $E(\tilde{x}_{n+1} \mid \tilde{x}_{(n)})$ has to be positive. Consequently, one could search for any prior w.r.t. which $E(\tilde{x}_{n+1} \mid \tilde{x}_{(n)})$

is presentable as the expectation of a distribution truncated on $[0,+\infty)$, provided such distribution depends on $\tilde{x}_{(n)}$ only through $\frac{1}{n}\sum_{k=1}^{n}\tilde{x}_k$ [from $(I1)$] and becomes more and more concentrated around $\tilde{\beta}$ as $n \to +\infty$ [from Proposition 5.2]. As a typical illustration of this case consider

$$E(\tilde{x}_{n+1} \mid \tilde{x}_{(n)}) = \frac{\int_0^{+\infty} x \exp\{-\frac{n}{2}[x - \frac{1}{n}(\sum_{k=1}^{n}\tilde{x}_k + \gamma)]^2\}dx}{\int_0^{+\infty} \exp\{-\frac{n}{2}[x - \frac{1}{n}(\sum_{k=1}^{n}\tilde{x}_k + \gamma)]^2\}dx} \qquad \ldots (6.5)$$

Does there exist $\nu_{\tilde{\beta}}$ for which (6.5) holds true for every n and for some suitable γ? To answer this question, denote the right-hand-side of (6.5) by $f_{\gamma,n}(\sum_{k=1}^{n}\tilde{x}_k)$ and rewrite (6.5) with $n = 1$:

$$\frac{\int_{\mathbb{R}} \beta e^{-\frac{1}{2}(\tilde{x}_1-\beta)^2}\nu_{\tilde{\beta}}(d\beta)}{\int_{\mathbb{R}} e^{-\frac{1}{2}(\tilde{x}_1-\beta)^2}\nu_{\tilde{\beta}}(d\beta)} = f_{\gamma,1}(\tilde{x}_1)$$

i.e.

$$\{f_{\gamma,1}(\tilde{x}_1) - \tilde{x}_1\}\int_{\mathbb{R}} e^{-\frac{1}{2}(\tilde{x}_1-\beta)^2}\nu_{\tilde{\beta}}(d\beta) = \frac{d}{d\tilde{x}_1}\int_{\mathbb{R}} e^{-\frac{1}{2}(\tilde{x}_1-\beta)^2}\nu_{\tilde{\beta}}(d\beta).$$

Integration of this equation gives

$$\int_{\mathbb{R}} e^{-\frac{\beta^2}{2}+\beta\tilde{x}_1}\nu_{\tilde{\beta}}(d\beta) = ce^{F_{\gamma,1}(\tilde{x}_1)} \qquad \ldots (6.6)$$

with $F_{\gamma,1}(\tilde{x}_1) = \int_0^{\tilde{x}_1} f_{\gamma,1}(u)du$. Thus, the answer is yes provided $c\exp(F_{\gamma,1}(\tilde{x}_1))$ is a moment generating function and, in such case, (6.6) has a unique solution. In fact, write $\phi_\gamma(\tilde{x}_1)$ for the denominator of (6.5) with $n = 1$, and observe that it satisfies equation

$$f_{\gamma,1}(\tilde{x}_1)d\tilde{x}_1 = d\log\phi_\gamma(\tilde{x}_1) + (\tilde{x}_1 + \gamma)d\tilde{x}_1$$

i.e.

$$\begin{aligned} c\exp\{F_{\gamma,1}(\tilde{x}_1)\} &= k\phi_\gamma(\tilde{x}_1)\exp\{\gamma\tilde{x}_1 + \frac{\tilde{x}_1^2}{2}\} \\ &= k'\int_0^{+\infty} e^{-\frac{1}{2}\beta^2+\beta(\gamma+\tilde{x}_1)}d\beta. \end{aligned}$$

Combining this with (6.6), and recalling the uniqueness theorem for Laplace transforms, gives

$$\nu_{\tilde{\beta}}(d\beta) = -\gamma e^{\beta\gamma}\mathbf{1}_{(0,+\infty)}(\beta)d\beta$$

where $\gamma \in (-\infty, 0)$ in order that $\nu_{\tilde{\beta}}$ may be a probability measure. Now it is an easy task to verify that (6.5) is valid for every n if and only if $\nu_{\tilde{\beta}}$ has the previous form, that is an exponential distribution with parameter $-\gamma$.

Example 6.2. (Gamma model with unknown scale parameter). Let $\tilde{x}$ denote a sequence of exchangeable interarrival times, in a process of emission of particles from a radioactive source. Assume that the p.d. of $\tilde{x}$ obeys $(I1)-(I3)$. Then,

$X \subset [0, +\infty)$ and $Ex_{\tilde{\beta}}$ describes a conditional p.d. for any interarrival time given $\tilde{\beta}$. In this setting, arrival rate (=intensity) of the process is a characteristic of paramount importance and, thus, let us suppose that $Ex_{\tilde{\beta}}$ is completely determined by the specification of the value of that rate. Hence, think of $\tilde{\beta}$ as a one-to-one function of the (random) arrival rate. Afterwards define the expected interarrival time - under Ex_{β} - as the reciprocal of the intensity when β is the value taken by the above one-to-one mapping. Positive linear mappings cannot be used in the present case. In fact, for them we have

$$-\frac{\rho'(\beta)}{\rho(\beta)} = \frac{c}{\beta} \qquad (\beta > 0)$$

for some strictly positive c. This entails

$$k\beta^c = \int_0^{+\infty} \gamma(x) e^{\beta x} dx \qquad (\beta > 0)$$

which is absurd since $\lambda(\{x > 0 : \gamma(x) > 0\}) > 0$. Whence, we can try to obtain a consistent specification of $Ex_{\tilde{\beta}}$ by assuming $\tilde{\beta} = -\tilde{\alpha}$, where $\tilde{\alpha}$ denotes the arrival rate and, therefore,

$$\frac{\rho'(\beta)}{\rho(\beta)} = \frac{c}{\beta} \qquad (\beta < 0) \qquad \ldots (6.7)$$

i.e.

$$k\alpha^{-c} = \int_0^{+\infty} \gamma(x) e^{-\alpha x} dx \qquad (\alpha > 0)$$

which yields $\gamma(x) = x^{c-1}$. Thus, if P obeys $(I1) - (\mathrm{I}_3)$ and if Ex_{β} is supposed to be determined by the arrival rate and (6.7) holds, then Ex_{β} is the gamma distribution with parameters $c > 0$ and $\alpha = -\beta > 0$.

7. Predictive Sufficiency and Parametric Models

Suppose that $\tilde{t}_{|D}$ is a predictive sufficient statistic with M_1^* in $\mathcal{B}(M_1)$ and $\nu(M_1^*) = 1$. In view of the definition of predictive sufficiency [cf. Section 5], $Q_n''(\tilde{t}(e_n(\tilde{x}_{(n)})), A)$, with A in $\mathcal{B}(X)$, is a version of $P(\tilde{x}_{n+1} \in A \mid e_n(\tilde{x}_{(n)}))$. ¿From now on, Q_n'' will be thought of as a measurable function of T into M_1. This is justified by the factorization theorem [see, e.g., Hoffmann-Jørgensen (1994), Section 6.4]. Moreover ν^*, ν_n^*, ν_∞^* will denote the p.d.'s of $\tilde{t}(\tilde{p})$, $\tilde{t}(e_n(\tilde{x}_{(n)}))$ and $(\tilde{t}(e_n(\tilde{x}_{(n)})))_{n\geq 1}$, respectively.

The previous section showed, under somewhat restrictive conditions, how the existence of a predictive sufficient statistic can lead to represent an exchangeable P as a mixture of the p.d.'s of sequences of i.i.d. r.v.'s with common distribution from an exponential family. Reduction of the support of ν by low-dimensional parameters is of great moment for the actual assessment of P and the implementation of predictive Bayesian methods. Hence, when modelling uncertainty in terms of observables is thought of as primary, it is worth enquiring into the origins of parametric inference from predictive sources. This is the same as investigating under what *qualitative* assumptions on predictive p.d.'s, P disintegrates into distributions which make the observable coordinate r.v.'s i.d.d., and which depend on an informative low-dimensional parameter. Assume, for instance, that the actual judgements of an individual are consistent with the following "robustness" condition: small departures of $\tilde{t}_{|D}$ from t produce departures of $Q_n''(\tilde{t},\cdot)$ from $Q_n''(t,\cdot)$ which are uniformly small w.r.t. n. As an aid in the precise formulation of this idea, consider a *countable* π-class $\mathcal{G}\subset\mathcal{B}(X)$ such that $\sigma(\mathcal{G})=\mathcal{B}(X)$ and $P_1(\partial A)=0$ for each A in $\mathcal{G}$, where

$$P_1(\cdot):=\int_{X^\infty}\tilde{p}(x,\cdot)P(dx).$$

Now observe that by (R3) and by the continuity of $\tilde{t}$, we get the following immediate, fundamental fact: *the set S of sequences $(\tilde{t}(e_n(x_{(n)})))_{n\geq 1}$ that converge has probability* 1. Whence, the above mentioned robustness condition states that there exists $S'\subset S$ with $\nu_\infty^*(S')=1$ and with the following property.

(**S1**) *To each $(t_n)_{n\geq 1}$ in S', A in $\mathcal{G}$ and $\epsilon>0$ corresponds some $\delta=\delta((t_n)_{n\geq 1},A,\epsilon)$ such that $\sup_{n\geq n_2}|Q_n(t_n,A)-Q_n(t_n',A)|<\epsilon$ for some n_2 in $\mathbb{N}$, whenever $(t_n')_{n\geq 1}\in S'$ and $\sup_{n\geq n_1}d_T(t_n',t_n)<\delta$ for some n_1 in $\mathbb{N}$, d_T being a metric for T.*

The formulation of the next result involves the usual definition of *completion* which differs from the one recalled in Section 2. Given a probability space $(\Omega,\mathcal{F},P)$ and a sub-σ-algebra $\mathcal{H}$ of $\mathcal{F}$, the σ-algebra $\overline{\mathcal{H}}$ of all subsets of Ω defined by $G\triangle M$ with $G\in\mathcal{H}$, $M\subset N\in\mathcal{F}$ and $P(N)=0$, is called completion of $\mathcal{H}$. $\overline{P}$ defined by $\overline{P}(G\triangle M)=P(G)$ is thought of as the completion of the restriction of P to $\mathcal{H}$. We are now in a position to prove

THEOREM 7.1. *Let X, T be Polish spaces and P an exchangeable probability on $(X^\infty,\mathcal{B}(X^\infty))$. Assume that $\tilde{t}$ is continuous on M_1^* with $\nu(M_1^*)=1$, and that $\tilde{t}_{|D}$ is predictive sufficient. Then, if* (S1) *holds true, there is a function $g:T\to M_1$, measurable w.r.t. $\overline{\mathcal{B}(T)}/\mathcal{B}(M_1)$, such that $\tilde{p}=g(\tilde{t}(\tilde{p}))$ a.s.-P, and*

$$P(A)=\int_T g^\infty(\theta,A)\overline{\nu}^*(d\theta)\qquad \textit{for all } A \textit{ in } \mathcal{B}(X^\infty).$$

PROOF. First, since

$$0=P_1(\partial A)=\int_{X^\infty}\tilde{p}(x,\partial A)P(dx)\qquad (A\in\mathcal{G})$$

and $\mathcal{G}$ is countable, there is F_1 in $\mathcal{B}(X^\infty)$ with $P(F_1)=1$ and such that $\tilde{p}(x,\partial A)=0$ for all x in F_1 and A in $\mathcal{G}$.
We will prove that there is a $(\overline{\mathcal{B}(T)}/\mathcal{B}(M_1))$-measurable function g for which $\tilde{p}=g(\tilde{t}(\tilde{p}))$ holds true with probability 1. Define

$$F=F_1\cap\{e_n(\tilde{x}_{(n)})\Rightarrow\tilde{p},Q_n''(\tilde{t}(e_n(\tilde{x}_{(n)})))\Rightarrow\tilde{p},(\tilde{t}(e_n(\tilde{x}_{(n)})))_{n\geq 1}\in S'\}.$$

Then $P(F) = 1$, in view of (**3**), Proposition 5.2, (S1) and the definition of F_1. For a particular x in F, $(\tilde{t}(e_n(x_{(n)}))_{n\geq 1}$ belongs to S'. Hence, for any x' in F for which $\tilde{t}(\tilde{p}(x')) = \tilde{t}(\tilde{p}(x))$ and for any A in $\mathcal{G}$, we have $d_T(\tilde{t}(e_n(x_{(n)})), \tilde{t}(e_n(x'_{(n)}))) < \delta$, for large n. Thus, by (S1),

$$|Q''_n(\tilde{t}(e_n(x_{(n)})), A) - Q''_n(\tilde{t}(e_n(x'_{(n)})), A)| < \epsilon. \qquad \ldots (7.1)$$

holds true for large n. Moreover, by the definition of F and Theorem 2.1 in Billingsley (1968), $Q''_n(\tilde{t}(e_n(x_{(n)}), A)$ $[Q''_n(\tilde{t}(e_n(x'_{(n)})), A)$, respectively] converges to $\tilde{p}(x, A)$ $[\tilde{p}(x', A)$, respectively]. Combining this fact with (7.1) gives $|\tilde{p}(x, A) - \tilde{p}(x', A)| < \epsilon$, i.e., from the arbitrariness of ϵ,

$$\tilde{p}(x, A) = \tilde{p}(x', A).$$

Since $\mathcal{B}(X) = \sigma(\mathcal{G})$, the previous equality yields $\tilde{p}(x) = \tilde{p}(x')$ for all x, x' in F such that $\tilde{t}(\tilde{p}(x)) = \tilde{t}(\tilde{p}(x'))$. Thus $\tilde{p}(x) = Q''(\tilde{t}(\tilde{p}(x)))$ for all x in F and for some function $Q'' : T_0 \to M_1$ with $T_0 = \tilde{t}(\tilde{p}(F))$.

Fix p_0 in M_1 and set $g = Q''\mathbf{1}_{T_0} + p_0\mathbf{1}_{T_0^c}$. We have to prove that g is $(\overline{\mathcal{B}(T)}/\mathcal{B}(M_1))$-measurable. First we show that $T_0 \in \overline{\mathcal{B}(T)}$. Since $\tilde{t}(\tilde{p})$ is measurable, by the Lusin continuity theorem and the Prokhorov approximation theorem [cf., e.g., Doob (1994), Sections V.15 and IV.6], there is a compact $K_j \subset F$, such that $P(K_j) > 1 - 1/j$, with the property that $\tilde{t}(\tilde{p})$ is continuous on K_j. Then $\tilde{t}(\tilde{p}(K_j)) \subset T_0$ is compact and $\nu^*(\tilde{t}(\tilde{p}(K_j)) > 1 - 1/j$. Furthermore $B := \cup_j \tilde{t}(\tilde{p}(K_j)) \subset T_0$ has ν^*-probability 1 and, by the customary definition of completion, this implies that T_0 belongs to $\overline{\mathcal{B}(T)}$.

Next, we prove that $t \mapsto Q''(t, A)$ is continuous on T_0, for each A in $\mathcal{G}$. In fact, by (S1), if t, t' belong to T_0 and $d_T(t, t') < \delta/2$, there are x, x' in F such that $t = \tilde{t}(\tilde{p}(x))$, $t' = \tilde{t}(\tilde{p}(x'))$, and

$$|Q''_n(t_n, A) - Q''_n(t'_n, A)| < \epsilon$$

for large n, with $t_n = \tilde{t}(e_n(x_{(n)}))$, $t'_n = \tilde{t}(e_n(x'_{(n)}))$. Hence $|Q''(t, A) - Q''(t', A)| \leq \epsilon$, which shows that $t \mapsto Q''(t, A)$ is continuous on T_0, for each A in $\mathcal{G}$. Then, in view of the measurability of T_0 w.r.t. $\overline{\mathcal{B}(T)}$, $Q''(\cdot, A) : T_0 \to \mathbb{R}$ is measurable w.r.t. the same σ-algebra.

To prove that Q'' is $(\overline{\mathcal{B}(T)}/\mathcal{B}(M_1))$-measurable, consider the σ-algebra $\mathcal{D}$ *generated* by $\cup_{A\in\mathcal{G}}\sigma(\{p \in M_1 : p(A) \leq x\} : x \in \mathbb{R})$ and $\mathcal{B}' = \{B \in \mathcal{B}(X) : \{p \in M_1 : p(B) \leq x\} \in \mathcal{D}$ for all x in $\mathbb{R}\} \supset \mathcal{G}$. Then, $\{p : p(X) \leq x\} = \emptyset$ or M_1 according to $x < 1$ or $x \geq 1$, and, hence, $\{p : p(X) \leq x\} \in \mathcal{D}$, i.e. $X \in \mathcal{B}'$. Moreover, if $B_1, B_2 \in \mathcal{B}'$ and $B_1 \subset B_2$, then $\{p : p(B_2 \setminus B_1) > x\} = \{p : p(B_2) - p(B_1) > x\} = \cup_{r\in\mathbb{Q}}\{p(B_2) > r\} \cap \{p(B_1) < r - x\} \in \mathcal{D}$, which implies that $B_2 \setminus B_1 \in \mathcal{B}'$. Finally, if $B_n \in \mathcal{B}'$ $(n \geq 1)$ and $B_n \uparrow B$, then $\{p(B) \leq x\} = \{\lim p(B_n) \leq x\} = \cap_{n\geq 1}\{p(B_n) \leq x\} \in \mathcal{D}$ so that $B \in \mathcal{B}'$. This shows that $\mathcal{B}'$ is a λ-class and, by Theorem 2 on page 7 of Chow and Teicher (1997), $\mathcal{B}(X) \supset \mathcal{B}' \supset \sigma(\mathcal{G}) = \mathcal{B}(X)$. Hence, $\mathcal{D} = \sigma(\cup_{A\in\mathcal{B}(X)}\sigma(\{p(A) \leq x\} : x \in \mathbb{R})) = \mathcal{B}(M_1)$, i.e., $\cup_{A\in\mathcal{G}}\sigma(\{p(A) \leq x\} : x \in \mathbb{R})$ generates $\mathcal{B}(M_1)$ and, therefore, Q'' is

$(\overline{\mathcal{B}(T)}/\mathcal{B}(M_1))$-measurable if $(Q'')^{-1}(\{p(A) \leq x\}) \in \overline{\mathcal{B}(T)}$ for all A in $\mathcal{G}$ and x in $\mathbb{R}$. In fact, for any A in $\mathcal{G}$ and x in $\mathbb{R}$:

$$(Q'')^{-1}(\{p(A) \leq x\}) = \{t : Q''(t) \in \{p(A) \leq x\}\} = \{t : Q''(t, A) \leq x\} \in \overline{\mathcal{B}(T)}$$

in view of the previous part of this proof.
Combination of this statement with measurability of T_0 shows that g is $(\overline{\mathcal{B}(T)}/\mathcal{B}(M_1))$-measurable. To complete the proof, observe that $\tilde{p} = g(\tilde{t}(\tilde{p}))$ a.s.-P, and extend de Finetti's representation (R3) to $(X^\infty, \overline{\mathcal{B}(X^\infty)}, \overline{P})$. Finally, use the change-of-variable formula. □

Theorem 7.1 states that the r.v.'s $\tilde{x}_n$ are conditionally i.i.d. [w.r.t. the completion $\overline{P}$ of P] given $\tilde{\theta} = \tilde{t}(\tilde{p})$. In view of the meaning of $\tilde{\theta}$, ν^* can be assessed by thinking of it as an approximating p.d. of the p.d. of $\tilde{t}(e_n(\tilde{x}_{(n)}))$ when n is sufficiently large. The attention to be paid to the evaluation of ν^* is a logical consequence of the special status of the predictive sufficient statistic $\tilde{t}_{|D}$ in modelling uncertainty in terms of observables.

As far as assumptions in Theorem 7.1 are concerned, observe that they are trivially satisfied when $\tilde{t}$ is the identity function, i.e. when no actual reduction of data is assumed in the process of assessing predictive beliefs. Hence, it seems that (S1) may be necessary in order that $\tilde{t}(\tilde{p})$ plays the same role as $\tilde{p}$ in de Finetti's representation of P. The proof of Theorem 7.1 shows that $g(\cdot, A)$ is continuous on T_0 with $\overline{\nu}^*(T_0) = 1$. Thus the necessity of (S1) will be investigated under the additional hypothesis that $g(\cdot, A)$ is continuous almost certainly.

Theorem 7.2. *Let X, T be Polish spaces and P an exchangeable probability on $(X^\infty, \mathcal{B}(X^\infty))$. Assume that $\tilde{t}$ is continuous on M_1^* with $\nu(M_1^*) = 1$, and that $\tilde{t}_{|D}$ is predictive sufficient. If there is a function $g : T \to M_1$, measurable w.r.t. $\overline{\mathcal{B}(T)}/\mathcal{B}(M_1)$, such that, for every A in $\mathcal{G}$, $g(\cdot, A)$ is continuous $\overline{\nu}^*$-a.s. and*

$$P(A) = \int_T g^\infty(\theta, A)\overline{\nu}^*(d\theta) \qquad \text{for all } A \text{ in } \mathcal{B}(X^\infty), \qquad \ldots(7.2)$$

then (S1) *holds true.*

Proof. From (7.2) it follows that there exists F_0 in $\mathcal{B}(X^\infty)$ with $P(F_0) = 1$ and such that $\tilde{p}(x) = g(\tilde{t}(\tilde{p}(x)))$ for every x in F_0. Define $G_1 = \cap_{A \in \mathcal{G}}\{t \in T : g(\cdot, A) \text{ is continuous in } t\}$, $F_1 = (\tilde{t}(\tilde{p}))^{-1}(G_1)$. Then the set

$$F = F_0 \cap F_1 \cap \{e_n(\tilde{x}_{(n)}) \Rightarrow \tilde{p}, Q''_n(\tilde{t}(e_n(\tilde{x}_{(n)}))) \Rightarrow \tilde{p}\}$$

has probability 1. Moreover $S' := \{(\tilde{t}(e_n(x_{(n)})))_{n\geq 1} : x \in F\}$ belongs to $\overline{\mathcal{B}(T^\infty)}$ and $\overline{\nu}^*_\infty(S') = 1$, by the same argument used, in the proof of Theorem 7.1, to show that T_0 belongs to $\overline{\mathcal{B}(T)}$ and $\overline{\nu}^*(T_0) = 1$.

Let $(t_n)_{n\geq 1}$ be any fixed point in S'. Then $t_n = \tilde{t}(e_n(x_{(n)}))$ for some x in F. Since $t := \tilde{t}(\tilde{p}(x))$ belongs to G_1, to every A in $\mathcal{G}$ and $\epsilon > 0$ corresponds $\delta = \delta(t, A, \epsilon) > 0$ such that $|g(t, A) - g(t', A)| < \epsilon/2$, whenever t' belongs to G_1 and $d_T(t, t') < \delta$.

Now, if $(t'_n)_{n\geq 1}$ belongs to S' and n_1 is such that $\sup_{n\geq n_1} d_T(t_n, t'_n) < \delta/2$, then $t'_n = \tilde{t}(e_n(x'_{(n)}))$ for some x' in F. Moreover, since $\tilde{t}(e_n(x_{(n)})) \to \tilde{t}(\tilde{p}(x))$,

$\tilde{t}(e_n(x'_{(n)})) \to \tilde{t}(\tilde{p}(x'))$, we have $d_T(\tilde{t}(\tilde{p}(x)), \tilde{t}(\tilde{p}(x'))) < \delta$. This, together with the fact that $\tilde{t}(\tilde{p}(x))$ and $\tilde{t}(\tilde{p}(x'))$ belong to G_1, yields $|g(\tilde{t}(\tilde{p}(x)), A) - g(\tilde{t}(\tilde{p}(x')), A)| < \epsilon/2$, i.e.

$$|\tilde{p}(x, A) - \tilde{p}(x', A)| < \frac{\epsilon}{2}. \qquad \ldots (7.3)$$

On the other hand, $Q''_n(\tilde{t}(e_n(x_{(n)})), A) \to \tilde{p}(x, A)$ and $Q''_n(\tilde{t}(e_n(x'_{(n)})), A) \to \tilde{p}(x', A)$ (as $n \to +\infty$), since x and x' belong to F. Finally, combining this fact with (7.3), gives $\sup_{n \geq n_2} |Q''_n(\tilde{t}(e_n(x_{(n)})), A) - Q''_n(\tilde{t}(e_n(x'_{(n)})), A)| < \epsilon$. □

In connection with applicability of Theorem 7.1, it is worth pointing out that specification of ν^*_∞ and $(Q''_n)_{n\geq 1}$ is not needed. In point of fact, assumptions in Theorem 7.1 concern qualitative aspects of the above mentioned distributions. Then, one has simply to decide whether or not one's judgements on those qualitative aspects agree with what is stated by assumption. If the answer is yes, de Finetti's representation of P can be given in a parametric form justified by large-sample arguments on observable random entities. Further propositions, which give conditions in order that de Finetti's representation holds in a parametric form, are proved in Fortini, Ladelli and Regazzini (1998).

When $(Q''_n)_{n\geq 1}$ is specified, Theorem 7.1 can be exploited to deduce the precise form of the ingredients of a parametric representation.

Example 7.1. Let $X = [0, +\infty)$ and $(\tilde{x}_n)_{n\geq 1}$ be a sequence of exchangeable r.v.'s such that the p.d. of $\tilde{x}_{(n)}$ has a density (w.r.t. the Lebesgue measure) defined, for any $(x_1, \ldots x_n)$ in $(0, +\infty)^n$ and n in $\mathbb{N}$, by

$$\alpha x_0^\alpha (\alpha + n)^{-1} (x_0 \vee x_1 \vee \ldots \vee x_n)^{-\alpha - n} \qquad \ldots (7.4)$$

where α and x_0 are strictly positive constants. Hence,

$$\frac{\alpha + n}{\alpha + n + 1} \frac{(x_0 \vee x_1 \vee \ldots \vee x_n)^{\alpha + n}}{(x_0 \vee x_1 \vee \ldots \vee x_n \vee x)^{\alpha + n + 1}} \qquad \ldots (7.5)$$

represents a predictive density, at $x > 0$, for $\tilde{x}_{n+1}$ given $\tilde{x}_{(n)} = (x_1, \ldots x_n)$. Define $\tilde{t}(e_n) = \sup(\text{Support}(e_n))$. Clearly, $\tilde{t}(e_n) = M_n(x_{(n)}) := x_1 \vee \ldots \vee x_n$ is a predictive sufficient statistic. Hence, de Finetti's representation can be established in a parametric form with $\tilde{t}(\tilde{p}) = \sup(\text{Support}(\tilde{p}))$ if we show that:

- $M_n(\tilde{x}_{(n)})$ converges a.s.;
- $\tilde{t}(\tilde{p})$ obeys the continuity condition assumed in Theorem 7.1.

Apropos of these points note that the p.d. function of $M_n(\tilde{x}_{(n)})$ at $x > 0$ can be deduced from (7.4) in the following form

$$\alpha x_0^\alpha (\alpha + n)^{-1} \int_{[0,x]^n} (x_0 \vee x_1 \vee \ldots \vee x_n)^{-\alpha - n} dx_1 \ldots dx_n =$$

$$\alpha x_0^\alpha (\alpha + n)^{-1} n! \int_0^x dx_n \int_0^{x_n} dx_{n-1} \ldots \int_0^{x_2} (x_0 \vee x_1 \vee \ldots \vee x_n)^{-\alpha - n} dx_1$$

by using the symmetry of the integrand. Hence, a density at x for M_n is

$$\frac{n\alpha x_0^\alpha}{\alpha+n}\frac{x^{n-1}}{(x_0 \vee x)^{\alpha+n}}.$$

Thus, the p.d. of M_n converges weakly to the Pareto distribution with parameters α, x_0. Hence, since the monotone sequence $(M_n(\tilde{x}_{(n)}))_{n\geq 1}$ converges in probability to a real r.v. [and hence a.s.], then the limiting r.v. $\tilde{\theta}$ is greater than x_0 with probability 1. At this stage it is easy to show that (S1) holds for every sequence $(M_n)_{n\geq 1}$, with $\mathcal{G} = \{[0,x] : x \in \mathbb{Q}^+\}$, and that (7.5) converges to the uniform density

$$\frac{1}{\tilde{\theta}}\mathbf{1}_{[0,\tilde{\theta}]}(x)$$

with probability 1. These remarks lead us to *conjecture* that the p.d. of $\tilde{x}$ is a mixture of distributions of i.i.d. r.v.'s, having the uniform distribution on $[0,\tilde{\theta}]$, $\tilde{\theta}$ being a Pareto p.d.. In fact, equality

$$\alpha x_0^\alpha(\alpha+n)^{-1}(x_0 \vee x_1 \vee \ldots \vee x_n)^{-\alpha-n} = \int_{x_0}^{+\infty}\{\prod_{k=1}^{n}\frac{1}{\theta}\mathbf{1}_{[0,\theta]}(x_k)\}\frac{\alpha x_0^\alpha}{\theta^{\alpha+1}}d\theta$$

is valid for any $x_1,\ldots x_n$ in X and n in $\mathbb{N}$. Therefore, since sup(Support($\cdot$)) is continuous on the set $\mathcal{U}_{x_0}$ of all uniform p.d.'s supported by $[0,\theta]$ for any $\theta > x_0$, then, by recalling Proposition 5.2, sup(Support($\cdot$)) is continuous on $M_1^* := D \cup \mathcal{U}_{x_0}$. Thus, by Theorem 7.1, if $\tilde{x}_n$ are exchangeable r.v.'s according to (7.4), then they are conditionally independent and identically distributed, given a Pareto r.v. $\tilde{\theta}$ with parameters α, x_0, and $P\{\tilde{x}_1 \leq x \mid \tilde{\theta}\} = (x/\tilde{\theta})\mathbf{1}_{[0,\tilde{\theta}]}(x) + \mathbf{1}_{(\tilde{\theta},+\infty)}$.

Acknowledgments. The authors would like to acknowledge the helpful suggestions of an anonymous referee.

REFERENCES

ALDOUS, D.J. (1985). Exchangeability and related topics. *École d'Été de Probabilités de Saint-Flour XIII - Lecture Notes in Math. 1117.* Springer-Verlag, Berlin.

BERNARDO, J.M. and SMITH A.F.M. (1994). *Bayesian Theory.* Wiley, Chichester.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

CAMPANINO, M and SPIZZICHINO, F. (1981). Prediction, sufficiency and representation of infinite sequences of exchangeable random variables. *Technical report.* Ist. di Matematica "G. Castelnuovo" Univ. Roma I.

CHOW, Y.S. and TEICHER, H. (1997). *Probability Theory* (3rd Edition). Springer-Verlag, New-York.

CIFARELLI, D.M. and REGAZZINI, E. (1980). Sul ruolo dei riassunti esaustivi ai fini della previsione in contesto bayesiano (Parte 1^a). *Rivista di Mat. per le Scienze Econ. e Soc.* **3** 109-125.

− − −− (1981). Sul ruolo dei riassunti esaustivi ai fini della previsione in contesto bayesiano (Parte 2^a). *Rivista di Mat. per le Scienze Econ. e Soc.* **3** 3-11.

− − −− (1982). Some considerations about mathematical statistics teaching methodology suggested by the concept of exchangeability. *Exchangeability in Probability and Statistics* (G. Koch and F. Spizzichino, eds.) 185-205. North-Holland, Amsterdam.

DAWID, A.P. (1982). Intersubjective statistical models. *Exchangeability in Probability and Statistics* (G. Koch and F. Spizzichino, eds.) 217-232. North-Holland, Amsterdam.

Dellacherie, C. and Meyer, P.A. (1975). *Probabilités et Potentiel.* (Chapitres I à IV). Hermann, Paris.
Diaconis, P. and Freedman D. (1980). de Finetti theorem for Markov chains. *Ann. Probab.* **8** 115-130.
−−−− (1984). Partial exchangeability and sufficiency. *Statistics: Applications and New Directions* (J.K. Ghosh and J. Roy, eds.) 205-274.
−−−− (1990). Cauchy's equation and de Finetti's theorem. *Scandinavian J. Statist.* **17** 235-274.

Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families *Ann. Statist.* **7** 265-281.
Doob, J.L., (1949). Application of the theory of martingales. *Colloque international Centre Nat. Rech. Sci.*, Paris, 22-28.
−−−− (1994). *Measure Theory.* Springer, Berlin.
Ferguson, T.S. (1973). A Bayesian analysis of some non parametric problems *Ann. Statist.* **1** 209-230.
de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré* **7** 1-68. [English translation in Studies in Subjective Probability (1980) (H.E. Kyburg and H.E. Smokler, eds) 53-118. Krieger, Malabar, FL.]
−−−− (1938). Sur la condition d'équivalence partielle. *Actualités Scientifique et Industrielles* N^o 739. Hermann, Paris.
−−−− (1952). Gli eventi equivalenti e il caso degenere. *Giorn. Istit. Ital. Attuari.* **15** 40-64.
−−−− (1959). La probabilità e la statistica nei rapporti con l'induzione secondo i diversi punti di vista. *Atti Corso CIME su Induzione e Statistica* 1-115. Cremonese, Roma. [English traslation in B. de Finetti, *Probability , Induction and Statistics* (1972) 147-227. Wiley, New York.]
Fortini, S., Ladelli, L. and Regazzini, E. (1998). Exchangeability, predictive distributions and parametric models. *Technical Report* C.N.R.-IAMI 98.5.
Fortini, S., Ladelli, L., Petris, G. and Regazzini, E. (1999). On mixtures of distributions of Markov chains. Technical Report C.N.R.-IAMI 99.9.
Halmos, P.R. (1950). *Measure theory.* Van Nostrand-Reinhold, Princeton, New Jersey.
Hoffmann-jørgensen, J. (1994). *Probability with a view toward Statistics.* Chapman and Hall, New York.
Kallenberg, O. (1986). *Random Measures.* Akademie-Verlag, Berlin.
Lauritzen, S.L. (1984). Extreme point models in statistics (with discussion) *Scandinavian J. Statist.* **11** 65-91.
−−−− (1988). Extremal families and systems of sufficient statistics. *Lecture Notes in Statist.* **49**. Springer, New York.
Lavine, M. (1992). Some aspects of Pólya tree distribution for statistical modelling. *Ann. Statist.* **20** 1222-1235.
Letta, G. (1981). Sulla nozione di riassunto esaustivo. *Rivista di Mat. per le Scienze Econ. e Soc.* **3** 109-125.
Lo, A.Y. (1991). A characterization of the Dirichlet process. *Statist. Prob. Lett.* **12** 163-168.
Mauldin, R.D., Sudderth, W.D., and Williams, S.C. (1992). Pólya trees and random distributions. *Ann. Statist.* **20** 1203-1221.
Muliere, P. and Secchi, P. (1992). Exchangeability, predictive sufficiency and Bayesian bootstrap. *J. Ital. Statist. Soc.* **3** 377-404.
Neveu, J. (1980). *Bases Mathématiques du Calcul des Probabilités.* Masson, Paris.
Parthasaraty, K.L. (1967). *Probability Measures on Metric Spaces* Academic Press, New York.
Regazzini, E. (1978). Intorno ad alcune questioni relative alla definizione del premio secondo la teoria della credibilità. *Giorn. Istit. Ital Attuari* **41** 77-89.
Schervish, M.J. (1995). *Theory of Statistics.* Springer-Verlag, New York.
Secchi P. (1987). Processi aleatori scambiabili e costruzione di modelli statistici. Degree thesis, Dip. Mat. "F. Enriques", Univ. degli Studi di Milano.
Walker, S. and Muliere P. (1997) A characterization of Pólya tree distributions. *Statist. Prob. Lett.* **31** 163-168.

Sandra Fortini, Università "L. Bocconi", Milano, Italy.
e-mail: sandra.fortini@uni-bocconi.it
Lucia Ladelli, Politecnico di Milano, Italy.
e-mail: lucia.ladelli@mate.polimi.it
Eugenio Regazzini, Università di Pavia and CNR-IAMI, Italy.
e-mail: eugenio@iami.mi.cnr.it