Exchangeable sequences of random variables

Donato Michele Cifarelli e Pietro Muliere
Istituto di Metodi Quantitativi
Università "Luigi Bocconi" - Milano

Piercesare Secchi
Dipartimento di Matematica "Francesco Brioschi"
Politecnico di Milano

May 3, 2002

Abstract

We present an introduction to the theory for infinite exchangeable sequences of events and of random variables with values in a Polish space.

This work is part of a book on 'Bayesian theory in a completely predictive perspective' that the authors are writing under contract with John Wiley & Sons, Ltd.

1 Introduction

The notion of a family of exchangeable events has been introduced by the dutch mathematician Jules Haag in 1924. However it was Bruno de Finetti who, after independently conceiving the same notion, announced in 1928, and proved in three papers published in 1930, the first fundamental results about sequences of exchangeable events. They include the law of large numbers and the celebrated Representation Theorem. These results were extended in 1933 by de Finetti to cover the case of exchangeable sequences of random variables (see also, de Finetti (1937).) Since then, many other authors contributed to the understanding of the properties of exchangeable families of random elements with values in abstract spaces: we only cite the fundamental work of Hewitt and Savage (1955).

The following pages are an introduction to the probabilistic theory for infinite sequences of exchangeable events and random variables with values in a Polish space.

2 Exchangeable events

Let $(\Omega, \mathcal{F}, P)$ be a probability space; we associate to a given family $(A_1, A_2, ..., A_N)$ of events in $\mathcal{F}$, the class $\mathcal{C}$ of the 2^N constituents generated by the family. That is

$$\mathcal{C} = \{\bigcap_{i=1}^{N} A_i^{s_i} : s_1, ..., s_N \in \{-1, 1\}\}$$

where $A_i^{s_i} = A_i$ if $s_i = 1$ and $A_i^{s_i} = A_i^c$ if $s_i = -1$. Every element of $\mathcal{C}$ corresponds to a trajectory on the lattice of points

$$\mathcal{L} = \{(i, j) : i, j \in \{0, 1, ..., N\}, -i \leq j \leq i\}.$$

Indeed, the constituent $\bigcap_{i=1}^{n} A_i^{s_i}$ may be represented by the trajectory that starts from the origin (0,0) and visits the points

$$(1, s_1), (2, s_1 + s_2), ..., (N, \sum_{i=1}^{N} s_i). \tag{2.1}$$

Viceversa, for $s_1, s_2, ..., s_N \in \{-1, 1\}$, the trajectory that starts from (0,0) and visits the points listed in (2.1) corresponds uniquely to the constituent $\bigcap_{i=1}^{N} A_i^{s_i}$.

For instance, the constituent

$$A_1^c \bigcap A_2^c \bigcap A_3 \bigcap A_4 \bigcap A_5 \bigcap A_6 \bigcap A_7 \bigcap A_8$$

generated by the family $(A_1, ..., A_8)$ is represented by the trajectory appearing in Figure 1.

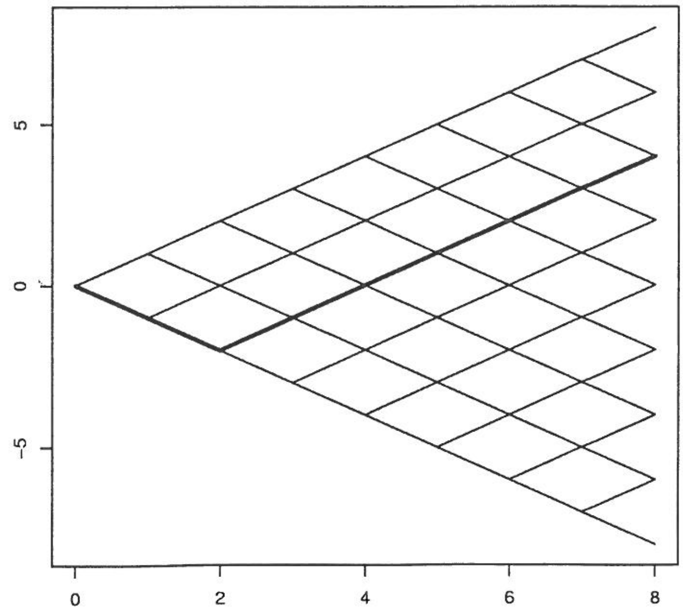

Figure 1: The trajectory representing $A_1^c \bigcap A_2^c \bigcap A_3 \bigcap A_4 \bigcap A_5 \bigcap A_6 \bigcap A_7 \bigcap A_8$.

Given $(n, v) \in \mathcal{L}$, let us focus on the set of trajectories that start from (0,0) and reach (n, v) after visiting the points

$$(1, s_1), (2, s_1 + s_2), ..., (n, \sum_{i=1}^{n} s_i), \tag{2.2}$$

where $s_1, ..., s_n \in \{-1, 1\}$ and are subject only to the constrain $\sum_{i=1}^{n} s_i = v$. Note that any such trajectory has $k = (n + v)/2$ ascending tracts and $n - k = (n - v)/2$ descending tracts; clearly $k \in \{0, 1, ..., n\}$. We want to assign the same probability to each of these trajectories; how can we do that?

For $n \leq N$ and $k \in \{0, 1, ..., n\}$, let us indicate with $A(n, k)$ the event that is true if among the first n events of the family $(A_1, ..., A_N)$, k are true and $n - k$ are false and write $P(A(n, k))$ for its probability; if $k = (n+v)/2$, this probability coincides with that of the set of trajectories described in (2.2). Write $\pi(n, k)$ for the probability of each trajectory of this set, that is the probability of the event that is true if k events with *assigned index* among the first n of the family $(A_1, ..., A_N)$, say $A_{i_1}, ..., A_{i_k}$, are true and the remaining $n - k$ are false. Then our previous request is equivalent to the following condition:

$$\pi(n, k) = \frac{P(A(n, k))}{\binom{n}{k}}. \tag{2.3}$$

2.4 **Definition.** *The family of events $(A_1, ..., A_N)$ is said to be exchangeable if (2.3) holds for all $n \in \{1, 2, ..., N\}$ and $k \in \{0, ..., n\}$.*

Here is a momentous characterization of the condition expressed in the previous definition.

2.5 **Proposition.** *The family of events $(A_1, ..., A_N)$ is exchangeable if and only if, for all $1 \leq n \leq N$ and $1 \leq i_1 < i_2 < \cdots < i_n \leq N$, $P(A_{i_1} \bigcap A_{i_2} \bigcap \cdots \bigcap A_{i_n})$ depends on n but not on the particular choice of indexes $i_1, i_2, ..., i_n$.*

Proof. Suppose that the family $(A_1, ..., A_N)$ is exchangeable, let $1 \leq n \leq N$ and $1 \leq i_1 < i_2 < \cdots < i_n \leq N$, and consider the event $A_{i_1} \bigcap A_{i_2} \bigcap \cdots \bigcap A_{i_n}$. This event is true if the events $A_{i_1}, A_{i_2}, ..., A_{i_n}$ are true and, among the remaining $N - n$ events of the family, k are true, with $k \in \{0, 1, ..., n - N\}$. Then, from (2.3), it follows that

$$\begin{aligned} P(A_{i_1} \bigcap A_{i_2} \bigcap \cdots \bigcap A_{i_n}) &= \sum_{k=0}^{N-n} \binom{N-n}{k} \pi(N, n+k) \\ &= \sum_{q=n}^{N} \binom{N-n}{q-n} \pi(N, q) \qquad (2.6) \\ &= \sum_{q=n}^{N} \frac{\binom{N-n}{q-n}}{\binom{N}{q}} P(A(N, q)) \qquad (2.7) \end{aligned}$$

which does not depend on the choice of $i_1, i_2, ..., i_n$.

Conversely, suppose that for all $1 \leq n \leq N$ and $1 \leq i_1 < i_2 < \cdots < i_n \leq N$, the probability $P(A_{i_1} \bigcap A_{i_2} \bigcap \cdots \bigcap A_{i_n})$ does not depend on the particular choice of the indexes $i_1, i_2, ..., i_n$, but only on their number n. Fix $n \in \{1, ..., N\}$, $k \in \{0, 1, ..., n\}$, $i_1, i_2, ..., i_n$ and compute

$$\begin{aligned} & P(A_{i_1} \bigcap A_{i_2} \bigcap \cdots \bigcap A_{i_k} \bigcap A^c_{i_{k+1}} \bigcap \cdots \bigcap A^c_{i_n}) = \\ & = P(A_{i_1} \bigcap A_{i_2} \bigcap \cdots \bigcap A_{i_k}) - \sum_{k+1 \leq j \leq n} P(A_{i_1} \bigcap A_{i_2} \bigcap \cdots \bigcap A_{i_k} \bigcap A_{i_j}) + \cdots \\ & \cdots + (-1)^{n-k} P(A_{i_1} \bigcap A_{i_2} \bigcap \cdots \bigcap A_{i_n}); \qquad (2.8) \end{aligned}$$

note that all the probabilities on the right member do not depend on the indices appearing but only on their number. Therefore, the probability on the left member depends only on n and k, but not on the choice of indexes $i_1, i_2, ..., i_n$. Define

$$\pi(n, k) = P(A_{i_1} \bigcap A_{i_2} \bigcap \cdots \bigcap A_{i_k} \bigcap A^c_{i_{k+1}} \bigcap \cdots \bigcap A^c_{i_n});$$

Then

$$P(A(n, k)) = \binom{n}{k} \pi(n, k)$$

and this shows that (2.3) holds. ◇

When $(A_1, ..., A_N)$ is exchangeable, condition (2.3) implies that for any choice of $1 \leq i_1 < i_2 < \cdots < i_n \leq N$,

$$P(A_{i_1} \bigcap A_{i_2} \bigcap \cdots \bigcap A_{i_n}) = P(A_1 \bigcap A_2 \bigcap \cdots \bigcap A_n) = \sum_{q=n}^{N} \frac{\binom{N-n}{q-n}}{\binom{N}{q}} P(A(N,q)).$$

From this in turn it follows that

$$\begin{aligned}
&P(A_{i_1} \bigcap A_{i_2} \bigcap \cdots \bigcap A_{i_k} \bigcap A_{i_k}^c \bigcap \cdots \bigcap A_{i_n}^c) = \\
&= \sum_{j=0}^{n-k} (-1)^j \binom{n-k}{j} P(A_{i_1} \bigcap A_{i_2} \bigcap \cdots \bigcap A_{i_{k+j}}) \\
&= \sum_{j=0}^{n-k} (-1)^j \binom{n-k}{j} \sum_{q=k+j}^{N} \frac{\binom{N-k-j}{q-k-j}}{\binom{N}{q}} P(A(N,q)) \\
&= \sum_{q=k}^{N} \frac{P(A(N,q))}{\binom{N}{q}} \sum_{j=0}^{\min(n-k,q-k)} (-1)^j \binom{n-k}{j} \binom{N-k-j}{q-k-j} \\
&= \sum_{q=k}^{N-n+k} \frac{\binom{N-n}{q-k}}{\binom{N}{q}} P(A(N,q)).
\end{aligned} \tag{2.9}$$

Formula (2.9) is very important since it shows that, under the assumption of exchangeability for the family $(A_1, ..., A_N)$, the probability of any sequence of events of the family can be computed once we assign the probabilities $P(A(N,q))$, for $q = 0, 1, ..., N$.

The notion of exchangeability can be extended to infinite sequences of events.

2.10 **Definition**. *An infinite sequence of events (A_n) is said to be exchangeable if, for all $N \geq 1$, the family $(A_1, ..., A_N)$ is exchangeable.*

For $x \in \Re$ and $N \geq 1$, let

$$F_N(x) = \sum_{k=0}^{N} P(A(N,k)) I[\frac{k}{N} \leq x]$$

where, for $A \in \mathcal{F}$, $I[A]$ is the indicator function of the event A. (We assume, by convention, that the sum is 0 for $x < 0$.) Note that F_N is a distribution function that assigns probability $P(A(N,k))$ to the value k/N, for $k = 0, 1, ..., N$. In fact, F_N is the distribution function of the random variable Y_N that is equal to k/N if among the events $(A_1, ..., A_N)$, k are true and $N-k$ are false; that is, Y_N is the frequency of success among the first N events of the sequence (A_n). The following Representation Theorem, due to de Finetti (1933c), is our first fundamental result for infinite exchangeable sequences of events.

2.11 **Theorem**. *An infinite sequence of events (A_n) is exchangeable if and only if there exists a distribution F on $[0, 1]$ such that*

$$\pi(n,k) = \int_0^1 \theta^k (1-\theta)^{n-k} dF(\theta) \tag{2.12}$$

for $n = 1, 2, \ldots$ *and* $k = 0, 1, \ldots, n$.

Moreover, F *is the weak limit of the sequence of distribution functions* (F_N).

Proof. It is almost trivial to check that the sequence (A_n) is exchangeable when (2.12) holds. To prove the converse, let us assume that (A_n) is exchangeable.

Because of (2.9), for $1 \leq n \leq N$ and $k = 0, \ldots, n$,

$$\begin{aligned}
\pi(n,k) &= \\
&= \sum_{q=k}^{N-n+k} \frac{\binom{N-n}{q-k}}{\binom{N}{q}} P(A(N,q)) \\
&= \sum_{q=k}^{N-n+k} \frac{q(q-1)\cdots(q-k+1)}{N(N-1)\cdots(N-n+1)}(N-q)(N-q-1)\cdots(N-q-(n-k-1))P(A(N,q)) \\
&= \int_0^1 I[\frac{k}{N} \leq \theta \leq \frac{N-n+k}{N}] \cdot \\
&\quad \cdot \frac{N\theta(N\theta-1)\cdots(N\theta-k+1)}{N(N-1)\cdots(N-n+1)} N(1-\theta)(N(1-\theta)-1)\cdots(N(1-\theta)-(n-k-1))dF_N(\theta).
\end{aligned}$$

Now observe that the integrand above converges uniformly in θ to

$$I[0 \leq \theta \leq 1]\theta^k(1-\theta)^{n-k}$$

when N grows to infinity. Therefore, for any $\epsilon > 0$, there is $N = N(\epsilon)$ such that,

$$\int_0^1 \theta^k(1-\theta)^{n-k} dF_N(\theta) - \epsilon \leq \pi(n,k) \leq \int_0^1 \theta^k(1-\theta)^{n-k} dF_N(\theta) + \epsilon. \tag{2.13}$$

It follows from Helly's Selection Theorem (see, for instance, Ash (1972), Theorem 8.2.1), that there exists a non decreasing, right continuous function F and a subsequence (F_{N_j}) of (F_N) such that

$$\lim_{N_j \to \infty} F_{N_j}(x) = F(x)$$

at every continuity point x of F. Moreover, F is a distribution function since the support of F_{N_j} is contained in $[0,1]$ for every N_j and

$$\lim_{N_j \to \infty} \int_0^1 \theta^k(1-\theta)^{n-k} dF_{N_j}(\theta) = \int_0^1 \theta^k(1-\theta)^{n-k} dF(\theta).$$

This fact, together with (2.13), implies that

$$\pi(n,k) = \int_0^1 \theta^k(1-\theta)^{n-k} dF(\theta).$$

This concludes the proof of the first part of the theorem.

For the second part, let $x \in [0,1]$ be a continuity point for F and note that for $N \geq 1$,

$$F_N(x) = \sum_{k=0}^{N} P(A(N,k)) I[\frac{k}{N} \leq x] = \int_0^1 \sum_{k=0}^{[Nx]} \binom{N}{k} \theta^k (1-\theta)^{N-k} dF(\theta).$$

However, the law of large numbers for independent and identically distributed Bernoulli random variables, guarantees that for $\theta \in [0,1] - \{x\}$

$$\lim_{N\to\infty} \sum_{k=0}^{[Nx]} \binom{N}{k} \theta^k (1-\theta)^{N-k} = I[\theta < x].$$

Hence, by Dominated Convergence Theorem,

$$\lim_{N\to\infty} F_N(x) = \int_0^1 I[\theta < x] dF(\theta) = F(x).$$

This shows that (F_N) weakly converges to F and concludes the proof of the theorem. ◇

Given an exchangeable infinite sequence of events (A_n), let us indicate with X_i the indicator function of the event A_i, for $i = 1, 2, ...$ Then, an equivalent way to express the representation (2.12) is the following:

$$P(X_1 = x_1, ..., X_n = x_n) = \int_0^1 \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n - \sum_{i=1}^n x_i} dF(\theta) \tag{2.14}$$

for $n \geq 1$ and $x_1, ..., x_n \in \{0,1\}$. Moreover, being $Y_N = N^{-1} \sum_{i=1}^N X_i$, for $N = 1, 2, ...$, it is

$$F(x) = \lim_{N\to\infty} P(\frac{1}{N} \sum_{i=1}^{N} X_i \leq x) \tag{2.15}$$

at any continuity point x of F.

2.16 **Example.** Let (X_n) be the sequence of indicators of the events of an infinite exchangeable sequence (A_n) of elements of $\mathcal{F}$. For all $N \geq 1$ and $q = 0, ..., N$, set

$$P(X_1 + X_2 + \cdots + X_N = q) = P(A(N,q)) = \frac{1}{N+1}.$$

We want to compute the probability distribution of the random vector $(X_1, X_2, ..., X_n)$, for $n = 1, 2, ...$

For $N \geq 1$, the characteristic function of $S_N = \sum_{i=1}^N X_i$ is

$$\phi_{S_N}(t) = \begin{cases} \frac{1}{N+1} \sum_{k=0}^N \exp(itk) = \frac{1}{N+1} \frac{\exp(it(N+1))-1}{\exp(it)-1} & \text{for } t \neq 0, \\ 1 & \text{for } t = 0. \end{cases}$$

Therefore the characteristic function of $N^{-1} S_N$ is

$$\phi_{\frac{1}{N}S_N}(t) = \begin{cases} \frac{1}{N+1} \frac{\exp(it\frac{N+1}{N})-1}{\exp(i\frac{t}{N})-1} & \text{for } t \neq 0, \\ 1 & \text{for } t = 0. \end{cases}$$

Hence

$$\lim_{N\to\infty} \phi_{\frac{1}{N}S_N}(t) = \begin{cases} \frac{\exp(it)-1}{it} & \text{for } t \neq 0, \\ 1 & \text{for } t = 0. \end{cases}$$

The right member of the previous equality determines the characteristic function of the uniform distribution; therefore Levy's Continuity Theorem and Theorem 2.11 imply that

$$\lim_{N\to\infty} P(\frac{1}{N}S_N \leq \theta) = \theta = F(\theta)$$

at every $\theta \in [0,1]$, and thus

$$P(X_1 = x_1, ..., X_n = x_n) = \int_0^1 \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} d\theta$$

for $n \geq 1$ and $x_1, ..., x_n \in \{0,1\}$. $\diamond$

Equation (2.15) shows that the sample mean of the indicators of an infinite exchangeable sequence of events converges in distribution. In fact, a stronger convergence holds.

2.17 **Theorem.** *Let (X_n) be the sequence of indicators of the events of an infinite exchangeable sequence. Then*

$$Y_N = \frac{1}{N}\sum_{i=1}^N X_i$$

converges almost surely to a random variable Θ as N grows to infinity.

Proof. We will show that

$$\lim_{N\to\infty} P(\sup_{k\geq 0}|Y_{N+k} - Y_N| \geq \epsilon) = 0;$$

indeed, this condition implies the existence of a random variable Θ such that $\lim_{N\to\infty} Y_N = \Theta$ on a set of probability one.

Let us consider the subsequence (Y_{N^2}). Given $\epsilon > 0$ and an integer q, whenever

$$|Y_{(N+k)^2} - Y_{(N+q)^2}| < \frac{\epsilon}{2}$$

for $k = 0, 1, ..., q$, then

$$|Y_{(N+k)^2} - Y_{N^2}| \leq |Y_{(N+k)^2} - Y_{(N+q)^2}| + |Y_{(N+k)^2} - Y_{N^2}| < \epsilon$$

for $k = 0, 1, ..., q$. Therefore

$$\begin{aligned} P(\sup_{0\leq k\leq q}|Y_{(N+k)^2} - Y_{N^2}| \geq \epsilon) &\leq P(\sup_{0\leq k\leq q}|Y_{(N+k)^2} - Y_{(N+q)^2}| \geq \frac{\epsilon}{2}) \\ &= P(\bigcup_{k=0}^q |Y_{(N+k)^2} - Y_{(N+q)^2}| \geq \frac{\epsilon}{2}) \\ &\leq \sum_{k=0}^q P(|Y_{(N+k)^2} - Y_{(N+q)^2}| \geq \frac{\epsilon}{2}) \\ &\leq \frac{4}{\epsilon^2}\sum_{k=0}^q E(|Y_{(N+k)^2} - Y_{(N+q)^2}|^2) \qquad (2.18) \end{aligned}$$

where the last inequality follows from that due to Markov. Now, for $m > n$, compute

$$
\begin{aligned}
&E(|Y_m - Y_n|^2) = \\
&= E[(\frac{1}{m}\sum_{i=1}^{m} X_i - \frac{1}{n}\sum_{i=1}^{n} X_i)^2] \\
&= E[(\frac{n\sum_{i=n+1}^{m} X_i - (m-n)\sum_{i=1}^{m} X_i}{mn})^2] \\
&= \frac{1}{m^2n^2}E[n^2(\sum_{i=n+1}^{m} X_i)^2 - 2n(m-n)\sum_{i=1}^{n} X_i \sum_{j=n+1}^{m} X_j + (m-n)^2(\sum_{i=1}^{n} X_i)^2] \\
&= \frac{1}{m^2n^2}E[n^2\sum_{i=n+1}^{m} X_i^2 + n^2 \sum_{n+1\le i\ne j\le m} X_iX_j - 2n(m-n)\sum_{i=1}^{n} X_i \sum_{j=n+1}^{m} X_j + \\
&\qquad +(m-n)^2\sum_{i=1}^{n} X_i^2 + (m-n)^2 \sum_{1\le i\ne j\le n} X_iX_j] \\
&= \frac{1}{m^2n^2}[n^2(m-n)P(X_1 = 1) + n^2(m-n)(m-n-1)P(X_1 = 1, X_2 = 1) \\
&-2n^2(m-n)^2P(X_1 = 1, X_2 = 1) + (m-n)^2nP(X_1 = 1) + (m-n)^2n(n-1)P(X_1 = 1, X_2 = 1)] \\
&= \frac{m-n}{mn}[P(X_1 = 1) - P(X_1 = 1, X_2 = 1)] \\
&\le \frac{1}{n}[P(X_1 = 1) - P(X_1 = 1, X_2 = 1)].
\end{aligned}
$$

The next to the last equality holds because the variables X_i are indicators of exchangeable events. Therefore (2.18) becomes

$$
\begin{aligned}
P(\sup_{0\le k\le q} |Y_{(N+k)^2} - Y_{N^2}| \ge \epsilon) &\le \frac{4}{\epsilon^2}\sum_{k=0}^{q}\frac{1}{(N+k)^2}[P(X_1 = 1) - P(X_1 = 1, X_2 = 1)] \\
&\le \frac{4}{\epsilon^2}[P(X_1 = 1) - P(X_1 = 1, X_2 = 1)]\sum_{j=N}^{\infty}\frac{1}{j^2}.
\end{aligned}
$$

Since the previous inequality is true for any q, it implies that

$$
\lim_{N\to\infty} P(\sup_{k} |Y_{(N+k)^2} - Y_{N^2}| \ge \epsilon) \le \lim_{N\to\infty}\frac{4}{\epsilon^2}[P(X_1 = 1) - P(X_1 = 1, X_2 = 1)]\sum_{j=N}^{\infty}\frac{1}{j^2} = 0.
$$

Therefore there is a random variable Θ such that

$$
P(\lim_{N\to\infty} Y_{N^2} = \Theta) = 1. \tag{2.19}
$$

For all $N \ge 1$, let $k = k(N)$ be integer and such that $k^2 < N \le (k+1)^2$. Then

$$
\begin{aligned}
|Y_N - Y_{K^2}| &= |\frac{1}{N}\sum_{i=1}^{N} X_i - \frac{1}{k^2}\sum_{i=1}^{k^2} X_i| \\
&= \frac{k^2\sum_{i=1}^{N} X_i - N\sum_{i=1}^{k^2} X_i}{Nk^2} \\
&= \frac{(N-k^2)\sum_{i=1}^{k^2} X_i - k^2\sum_{i=k^2+1}^{N} X_i}{Nk^2}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{(N-k^2)\sum_{i=1}^{k^2} X_i + k^2 \sum_{i=k^2+1}^{N} X_i}{Nk^2} \\
&\leq \frac{2k^2(N-k^2)}{Nk^2} \leq \frac{2}{N}(2k+1).
\end{aligned}
$$

Therefore

$$
|Y_N - \Theta| \leq |Y_N - Y_{k^2}| + |Y_{k^2} - \Theta| \leq |Y_{k^2} - \Theta| + \frac{2}{N}(2k+1);
$$

from this and (2.19) it follows that

$$
P(\lim_{N\to\infty} Y_N = \Theta) = 1
$$

◇

We are finally ready to prove a different version of the Representation Theorem.

2.20 **Theorem.** *Let (X_n) be the sequence of indicators of the events of an infinite exchangeable sequence. Then there is a random variable $\Theta \in [0,1]$ such that, conditionally on Θ, the random variables $X_1, X_2, ..., X_n, ...$ are independent and identically distributed Bernoulli(Θ). Moreover Θ is the almost sure limit of the sequence $(n^{-1}\sum_{i=1}^{n} X_i)$: hence its distribution is uniquely determined by the sequence (X_n).*

Proof. Let Θ be the almost sure limit of the sequence (Y_N) and F be its distribution function. It is enough to show that, for all $n \geq 1$, $x_1, ..., x_n \in \{0,1\}$ and $t \in [0,1]$,

$$
P(X_1 = x_1, ..., X_n = x_n, \Theta \leq t) = \int_0^t \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} dF(\theta).
$$

For $N > n$, let us compute

$$
\begin{aligned}
&P(X_1 = x_1, ..., X_n = x_n, Y_N \leq t) \\
&= P(X_1 = x_1, ..., X_n = x_n, \sum_{i=n+1}^{N} X_i \leq Nt - \sum_{i=1}^{n} x_i) \\
&= \sum_{q=0}^{N-n} \binom{N-n}{q} \int_0^1 \theta^{\sum_{i=1}^n x_i + q}(1-\theta)^{N-\sum_{i=1}^n x_i - q} I[q \leq Nt - \sum_{i=1}^n x_i] dF(\theta) \\
&= \int_0^1 \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} \left\{ \sum_{q=0}^{N-n} \binom{N-n}{q} \theta^q (1-\theta)^{N-n-q} I[q \leq Nt - \sum_{i=1}^n x_i] \right\} dF(\theta)
\end{aligned}
$$

The second equality is true because of exchangeability and (2.14). However,

$$
\lim_{N\to\infty} \sum_{q=0}^{N-n} \binom{N-n}{q} \theta^q (1-\theta)^{N-n-q} I[q \leq Nt - \sum_{i=1}^n x_i] = I[\theta \leq t]
$$

because of the strong law of large numbers for sequences of independent and identically distributed Bernoulli random variables. Therefore,

$$
\begin{aligned}
P(X_1 = x_1, ..., X_n = x_n, \Theta \leq t) &= \lim_{N\to\infty} P(X_1 = x_1, ..., X_n = x_n, Y_N \leq t) \\
&= \int_0^t \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} dF(\theta);
\end{aligned}
$$

note that the first equality holds because Θ is the limit of the sequence (Y_N) almost surely. $\diamond$

2.21 **Remark.** A different approach for proving the previous theorem considers the characteristic function of the vector $(X_1, ..., X_n, \Theta)$, for $n \geq 1$; let it be $\phi_{X_1,...,X_n,\Theta}$. Indeed, by computing $\phi_{X_1,...,X_n,\Theta}$ as the limit of the sequence of characteristic functions $(\phi_{X_1,...,X_n,N^{-1}\sum_{i=1}^N X_i})$ for N going to infinity, one obtains

$$\phi_{X_1,...,X_n,\Theta}(t_1, ..., t_n, \theta) = \int_0^1 \exp(i\theta v) \prod_{i=1}^n [1 - v + v\exp(it_j)]dF(v)$$

for all $(t_1, ..., t_n, \theta) \in \Re^{n+1}$; this shows that, conditionally on Θ, the random variables of the sequence (X_n) are independent and identically distributed Benoulli(Θ).

This approach is often fruitful for characterizing the distribution F; we will follow it in Example 4.15. $\diamond$

2.22 **Example. Pólya urn.** Consider an urn initially containing $b_0 > 0$ black balls and $w_0 > 0$ white balls. At time $n = 1, 2, 3, ...$, a ball is sampled from the urn and replaced into it along with $m > 0$ balls of the same color. This generates an infinite sequence (X_n) of Bernoulli random variables, where X_n is 1 or 0 according to the color black or white respectively of the ball extracted from the urn at time n.

The sequence (X_n) is exchangeable. In fact, let $n \geq 1$, $0 \leq k \leq n$ and $i_1, ..., i_n$ such that $i_j \in \{0, 1\}$ and $\sum_{j=1}^n i_j = k$. Then

$$\begin{aligned} P(X_1 = i_1, ..., X_n = i_n) &= \frac{\prod_{j=0}^{k-1}(b_0 + jm) \cdot \prod_{j=0}^{n-k-1}(w_0 + jm)}{\prod_{j=0}^{n-1}(b_0 + w_0 + jm)} \\ &= \frac{\Gamma(\frac{b_0}{m} + \frac{w_0}{m})}{\Gamma(\frac{b_0}{m})\Gamma(\frac{w_0}{m})} \frac{\Gamma(\frac{b_0+k}{m})\Gamma(\frac{w_0+n-k}{m})}{\Gamma(\frac{b_0}{m} + \frac{w_0}{m} + n)} \\ &= \int_0^1 \theta^k(1-\theta)^{n-k}[\frac{\Gamma(\frac{b_0}{m} + \frac{w_0}{m})}{\Gamma(\frac{b_0}{m})\Gamma(\frac{w_0}{m})}\theta^{\frac{b_0}{m}-1}(1-\theta)^{\frac{w_0}{m}-1}]d\theta. \end{aligned} \tag{2.23}$$

This representation and Theorem 2.11 show that the sequence is exchangeable. Moreover, together with Theorem 2.20, (2.23) shows that, if $\Theta \in [0, 1]$ is the almost sure limit of the sequence $(n^{-1}\sum_{i=1}^n X_i)$, then Θ has distribution Beta with parameters $\frac{b_0}{m}$ and $\frac{w_0}{m}$.

3 Mixtures of i.i.d. random variables and exchangeability

There are many different approaches one can follow in order to introduce the notions of mixture of sequences of i.i.d random variables and that of exchangeability.

A very general approach begins by considering a metric space $\mathbf{X}$, separable and complete, endowed with its Borel σ-field $\mathcal{X}$. We write $\mathbf{P}$ for the class of all probability measures defined on $(\mathbf{X}, \mathcal{X})$ and we endow $\mathbf{P}$ with the σ-field $\mathcal{P}$ generated by the topology of weak convergence, that is the smallest σ-field according to which are measurable all mappings $\wp_A : \mathbf{P} \to [0,1]$, defined for all $A \in \mathcal{X}$ and $p \in \mathbf{P}$ by

$$\wp_A(p) = p(A).$$

Next, we introduce the product space $(\mathbf{X}^\infty, \mathcal{X}^\infty)$ where $\mathbf{X}^\infty$ indicates the set of all infinite sequences $x = (x_1, x_2, \ldots)$ of elements of $\mathbf{X}$ whereas $\mathcal{X}^\infty$ is the σ-field generated by the subsets of $\mathbf{X}^\infty$ of the type

$$A_1 \times \cdots \times A_n \times \mathbf{X}^\infty = \{x \in \mathbf{X}^\infty : x_1 \in A_1, \ldots, x_n \in A_n\}$$

with $n \geq 1$ and $A_1, \ldots, A_n \in \mathcal{X}$. It is a standard result of measure theory (Ash (1972), Corollary 2.7.3) that, for all $p \in \mathbf{P}$, there exists a unique probability p^∞ on $(\mathbf{X}^\infty, \mathcal{X}^\infty)$, called the product probability measure, such that

$$p^\infty(A_1 \times \cdots \times A_n \times \mathbf{X}^\infty) = \prod_{i=1}^{n} p(A_i)$$

for all $n \geq 1$ and $A_1, \ldots, A_n \in \mathcal{X}$.

Finally, we consider the product space $(\mathbf{X}^\infty \times \mathbf{P}, \mathcal{X}^\infty \times \mathcal{P})$ where

$$\mathbf{X}^\infty \times \mathbf{P} = \{(x,p) : x \in \mathbf{X}^\infty, p \in \mathbf{P}\}$$

while $\mathcal{X}^\infty \times \mathcal{P}$ is the smallest σ-field containing the family of measurable rectangles $\{A \times B : A \in \mathcal{X}^\infty, B \in \mathcal{P}\}$.

Any probability distribution ν defined on $(\mathbf{P}, \mathcal{P})$, induces a special probability measure π on $(\mathbf{X}^\infty \times \mathbf{P}, \mathcal{X}^\infty \times \mathcal{P})$ through the relation

$$\pi(C) = \int_{\mathbf{P}} p^\infty(\{x \in \mathbf{X}^\infty : (x,p) \in C\})\nu(dp) \tag{3.1}$$

that is assumed to hold for all $C \in \mathcal{X}^\infty \times \mathcal{P}$. Note that, when $C = (A_1 \times \cdots \times A_n \times \mathbf{X}^\infty) \times B$, with $n \geq 1, A_i \in \mathcal{X}$ for $i = 1, \ldots, n$ and $B \in \mathcal{P}$, then

$$\pi(C) = \int_B \prod_{i=1}^{n} p(A_i)\nu(dp).$$

3.2 **Lemma.** *π is a probability measure on $(\mathbf{X}^\infty \times \mathbf{P}, \mathcal{X}^\infty \times \mathcal{P})$.*

Proof. First observe that

$$\pi(\mathbf{X}^\infty \times \mathbf{P}) = \int_{\mathbf{P}} p^\infty(\mathbf{X}^\infty)\nu(dp) = 1.$$

Now let C_1, C_2 be two disjoint sets in $\mathcal{X}^\infty \times \mathcal{P}$; then

$$\begin{aligned}
&\pi(C_1 \bigcup C_2) \\
&= \int_{\mathbf{P}} p^\infty(\{x \in \mathbf{X}^\infty : (x,p) \in C_1 \bigcup C_2\})\nu(dp) \\
&= \int_{\mathbf{P}} p^\infty(\{x \in \mathbf{X}^\infty : (x,p) \in C_1\} \bigcup \{x \in \mathbf{X}^\infty : (x,p) \in C_2\})\nu(dp) \\
&= \int_{\mathbf{P}} p^\infty(\{x \in \mathbf{X}^\infty : (x,p) \in C_1\})\nu(dp) + \int_{\mathbf{P}} p^\infty(\{x \in \mathbf{X}^\infty : (x,p) \in C_2\})\nu(dp) \\
&= \pi(C_1) + \pi(C_2)
\end{aligned}$$

where the third equality follows from the fact that the sets

$$\{x \in \mathbf{X}^\infty : (x,p) \in C_1\} \text{ and } \{x \in \mathbf{X}^\infty : (x,p) \in C_2\}$$

are disjoint. This proves that π is finitely additive.

In order to prove that π is countably additive, let $C_1, C_2, \ldots, C_n, \ldots \in \mathcal{X}^\infty \times \mathcal{P}$ be an infinite sequence of sets decreasing to the empty set; note that, given any $p \in \mathbf{P}$, the sets

$$\{x \in \mathbf{X}^\infty : (x,p) \in C_n\} \in \mathcal{X}^\infty, \quad n = 1, 2, \ldots,$$

decrease to the empty set when n grows to infinity. Then

$$\begin{aligned}
&\lim_{n\to\infty} \pi(C_n) \\
&\quad = \lim_{n\to\infty} \int_{\mathbf{P}} p^\infty(\{x \in \mathbf{X}^\infty : (x,p) \in C_n\})\nu(dp) \\
&\quad = \int_{\mathbf{P}} \lim_{n\to\infty} p^\infty(\{x \in \mathbf{X}^\infty : (x,p) \in C_n\})\nu(dp) = 0
\end{aligned}$$

where the next to the last equality holds because of Dominated Convergence Theorem.

Hence π is a probability on $(\mathbf{X}^\infty \times \mathbf{P}, \mathcal{X}^\infty \times \mathcal{P})$. ◇

Since π is a probability measure on the product space $(\mathbf{X}^\infty \times \mathbf{P}, \mathcal{X} \times \mathcal{P})$, it induces two probability measures on the spaces $(\mathbf{X}^\infty, \mathcal{X}^\infty)$ and $(\mathbf{P}, \mathcal{P})$ respectively. In fact, for all $B \in \mathcal{P}$,

$$\pi(\mathbf{X}^\infty \times B) = \nu(B).$$

Analogously, for all $A \in \mathcal{X}^\infty$, we set

$$\tau(A) = \pi(A \times \mathbf{P}) = \int_{\mathbf{P}} p^\infty(A)\nu(dp); \tag{3.3}$$

τ is a probability measure on the space $(\mathbf{X}^\infty, \mathcal{X}^\infty)$.

Consider now a probability space $(\Omega, \mathcal{F}, P)$ and a sequence (X_n) of random variables defined on Ω and with values in $\mathbf{X}$. The sequence (X_n) induces a probability distribution τ' on $(\mathbf{X}^\infty, \mathcal{X}^\infty)$; in fact, τ' is uniquely determined by

$$\tau'(A_1 \times A_2 \times \cdots \times A_n \times \mathbf{X}^\infty) = P[X_1 \in A_1, \cdots, X_n \in A_n]$$

for $n \geq 1$ and $A_1, ..., A_n \in \mathcal{X}$. The probability distribution τ' is called the law of the sequence (X_n). At this point it is important to notice that, in our treatment of exchangeability the space $(\Omega, \mathcal{F})$ will not play any role; indee once the law τ' of (X_n) is defined on the space $(\mathbf{X}^\infty, \mathcal{X}^\infty)$, we can assume that, for $n \geq 1$, the random variable X_n is defined as

$$X_n : (\mathbf{X}^\infty, \mathcal{X}^\infty) \to (\mathbf{X}, \mathcal{X})$$

with $X_n(x) = x_n$ for all $x = (x_1, x_2, ...) \in \mathbf{X}^\infty$. That is, without loss of generality, we may assume that $(\Omega, \mathcal{F})$ coincides with $(\mathbf{X}^\infty, \mathcal{X}^\infty)$ and P with τ'.

When the law of (X_n) is the τ represented in (3.3), we will say that it is a mixture of laws of sequences of independent random variables with identical probability distribution. Note that if ν concentrates all its mass on the probability distribution $p_0 \in \mathbf{P}$, then (X_n) is an infinite sequence of i.i.d. random variables with distribution equal to p_0.

For $\sigma = (\sigma_1, \sigma_2, ...)$ a finite permutation of the integers, set $(\sigma(X_n)) = (X_{\sigma_1}, X_{\sigma_2}, ...)$.

3.4 **Proposition.** *Assume that the law of the sequence (X_n) is the τ represented in (3.3). Then, for any finite permutation σ of the integers, the laws of (X_n) and of $(\sigma(X_n))$ are the same.*

Proof. For all $A \in \mathcal{X}^\infty$, let

$$\sigma^{-1}(A) = \{(x_1, x_2, ...) \in \mathbf{X}^\infty : (x_{\sigma_1}, x_{\sigma_2}, ...) \in A\}$$

and

$$\tau_\sigma(A) = \tau(\sigma^{-1}(A)).$$

Clearly τ_σ is the law of $(\sigma(X_n))$.

For all $p \in \mathbf{P}$, $p^\infty(A) = p^\infty(\sigma^{-1}(A))$ since p^∞ is the law of an i.i.d. sequence of random variables with values in $\mathbf{X}$ and common distribution p. Therefore

$$\tau_\sigma(A) = \int_{\mathbf{P}} p^\infty(\sigma^{-1}(A))\nu(dp) = \int_{\mathbf{P}} p^\infty(A)\nu(dp) = \tau(A).$$

◇

3.5 **Definition.** *A sequence of random variables (X_n) is exchangeable if, for any finite permutation σ of the integers, it has the same law as that of the sequence $(\sigma(X_n))$.*

Proposition 3.4 shows that when the law of (X_n) is a τ with representation (3.3), the sequence (X_n) is exchangeable. Before we proceed any further, let us observe that two probability distributions are connected with any such sequence (X_n); τ and the law p^∞ appearing in the representation (3.3). While τ shows some sort of dependence among the random variables of the sequence (X_n), p^∞ assumes that they are independent and identically distributed with common probability distribution p. This assumption is in actual fact not observable, but for situations where one could imagine that p would be known if the entire population were observed (for instance, through a census); hence the difficulty of eliciting a probability distribution ν through which $p \in \mathbf{P}$ is selected. On the contrary, the assumption of exchangeability concerning the law τ attains to observable facts for which it is always possible, at least theoretically, to postulate different forms of dependence.

In the next section we will obtain the most important result concerning infinite exchangeable sequences of random variables, namely de Finetti's representation theorem; for the moment, let us notice a few elementary properties that follow directly from the assumption of exchangeability.

It is immediate to check that the definition of exchangeability implies that all the variables of the sequence X_n have the same marginal distributions; that is, the probability distributions of X_i and X_j are the same for all $i, j \geq 1$. Analogously, for $n \geq 1$, all finite vectors $(X_{i_1}, ..., X_{i_n})$, $i_1, ..., i_n$ being different integers, have the same probability distributions.

For $n \geq 1$, let $\phi_{X_1,...,X_n}$ be the characteristic function of the random vector $(X_1, ..., X_n)$.

3.6 **Proposition.** *The infinite sequence (X_n) is exchangeable if and only if, for all $n \geq 1$ and every permutation $(\sigma_1, ..., \sigma_n)$ of the integers $(1, ..., n)$,*

$$\phi_{X_1,...,X_n}(t_1, ..., t_n) = \phi_{X_1,...,X_n}(t_{\sigma_1}, ..., t_{\sigma_n}), \tag{3.7}$$

for $t_1, ..., t_n \in \Re$.

Proof. Assume first that (X_n) is exchangeable and let $n \geq 1$ and $(\sigma_1, ..., \sigma_n)$ be a permutation of $(1, ..., n)$; indicate with $(\sigma_1^{-1}, ..., \sigma_n^{-1})$ the inverse of $(\sigma_1, ..., \sigma_n)$ that is another permutation of $(1, ..., n)$. Then, for all $t_1, ..., t_n \in \Re$,

$$\begin{aligned}
\phi_{X_1,...,X_n}(t_1, ..., t_n) &= \phi_{X_{\sigma_1^{-1}},...,X_{\sigma_n^{-1}}}(t_1, ..., t_n) \\
&= E[\exp(i \sum_{k=1}^{n} t_k X_{\sigma_k^{-1}})] \\
&= E[\exp(i \sum_{k=1}^{n} t_{\sigma_k} X_k)] \\
&= \phi_{X_1,...,X_n}(t_{\sigma_1}, ..., t_{\sigma_n});
\end{aligned}$$

the first equality holds since $(X_1, ..., X_n)$ and $(X_{\sigma_1^{-1}}, ..., X_{\sigma_n^{-1}})$ have the same distribution because (X_n) is exchangeable.

Along analogous computations one shows that if (3.7) holds, then for all n and all permutations $(\sigma_1, ..., \sigma_n)$ of $(1, ..., n)$, the characteristic function of $(X_1, ..., X_n)$ and $(X_{\sigma_1}, ..., X_{\sigma_n})$ are the same. This, easily implies that the sequence (X_n) is exchangeable. $\diamond$

The following result shows that the assumption of exchangeability for the sequence (X_n) implies that the correlation between couples of random variables extracted from the sequence is always nonnegative. Let us indicate with $\rho(X_i, X_j)$ the correlation between the random variables X_i and X_j.

3.8 **Proposition.** *If the infinite sequence (X_n) is exchangeable,*

$$\rho = \rho(X_i, X_j) \geq 0$$

for all integers i, j.

Proof. For all i, j, $\rho(X_i, X_j)$ is constant since all vectors (X_i, X_j) have the same distribution. Let us indicate with $\sigma^2 = \text{Var}(X_1)$ the variance of the random variable X_1; this is also the variance of the variable X_i, for $i = 1, 2, \dots$, since (X_n) is exchangeable. For $n \geq 1$, compute

$$\text{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \text{Var}(X_i) + \sum_{1 \leq i \neq j \leq n} \text{Cov}(X_i, X_j) = n\sigma^2 + n(n-1)\rho\sigma^2.$$

Therefore, for all $n \geq 1$,

$$\rho \geq -\frac{1}{n-1}$$

and thus $\rho \geq 0$. ◇

3.9 **Example.** Let $\rho \in [0, 1]$, and (Y_n) be a sequence of independent and identically distributed random variables. For $n \geq 1$, set

$$X_n = Y_0 + aY_n$$

with $a = \sqrt{\rho/(1-\rho)}$. Then (X_n) is exchangeable and

$$\rho(X_i, X_j) = \frac{a^2}{a^2 + 1} = \rho.$$

4 de Finetti's Representation Theorem

A general formulation of the famous de Finetti's representation theorem states that:

> The law of an infinite exchangeable sequence of random variables is a mixture of laws of i.i.d. sequences.

Many different approaches are known in the literature for proving the theorem. We will follow one based on the law of large numbers for exchangeable sequences.

Our treatment of de Finetti's theorem will stress the fact that the result regards the law of an infinite sequence of random variables, and not the abstract space $(\Omega, \mathcal{F})$ where the variables are defined. Indeed, without loss of generality we endow the the product space $(\mathbf{X}^\infty, \mathcal{X}^\infty)$ with a probability distribution τ and we consider the sequence of random variables

$$X_n : (\mathbf{X}^\infty, \mathcal{X}^\infty) \to (\mathbf{X}, \mathcal{X})$$

defined, for $n \geq 1$ and $x = (x_1, x_2, ...) \in \mathbf{X}^\infty$ by $X_n(x) = x_n$. Hence the law of the inifinite sequence (X_n) is τ.

4.1 **Definition.** *A Borel-measurable, real function f defined on $(\mathbf{X}^\infty, \mathcal{X}^\infty)$ is said to be n-symmetric if*

$$f(x_1, x_2, ..., x_n, x_{n+1} ...) = f(x_{\sigma_1}, x_{\sigma_2}, ..., x_{\sigma_n}, x_{n+1}, ...)$$

for all $(x_1, x_2, ...) \in \mathbf{X}^\infty$ and all permutations $(\sigma_1, ..., \sigma_n)$ of $(1, ..., n)$.

For instance, if $\mathbf{X} = \Re$ and, for all $(x_1, x_2, ...) \in \Re^\infty$,

$$f(x_1, x_2, ...) = (x_1 + x_2 + x_3) \exp(-x_1 x_2 x_3 + x_4 + x_5),$$

then f is 3-symmetric but not 4-symmetric.

Trivially, if f is n-symmetric then it is m-symmetric for all $m \leq n$. Therefore if Σ_n is the sigma-field generated by all real valued, n-symmetric functions defined on $(\mathbf{X}^\infty, \mathcal{X}^\infty)$, then

$$\mathcal{X}^\infty = \Sigma_1 \supset \Sigma_2 \supset \cdots \supset \Sigma_n \supset \Sigma_{n+1} \supset \cdots.$$

Set $\Sigma_\infty = \bigcap_{n=1}^\infty \Sigma_n$.

4.2 **Theorem.** *Let (X_n) be exchangeable. Then, for any Borel-measurable, real valued function ϕ defined on $(\mathbf{X}, \mathcal{X})$ and such that $E[|\phi(X_1)|] < \infty$, there exists a real valued random variable Y defined on $(\mathbf{X}^\infty, \mathcal{X}^\infty)$ such that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \phi(X_i) = Y$$

with probability one.

Proof. Let $n \geq 1$ and f be an n-symmetric function defined on $(\mathbf{X}^\infty, \mathcal{X}^\infty)$. For $i \in \{1, ..., n\}$,

$$\begin{aligned} E[\phi(X_1) f(X_1, X_2, ..., X_n, ...)] &= E[\phi(X_1) f(X_i, X_2, ..., X_{i-1}, X_1, X_{i+1}, ..., X_n, ...)] \\ &= E[\phi(X_i) f(X_1, X_2, ..., X_{i-1}, X_i, X_{i+1}, ..., X_n, ...)]; \end{aligned}$$

the first equality is true because f is n-symmetric, the second because the sequence (X_n) is exchangeable. Therefore

$$E[\frac{1}{n}\sum_{i=1}^{n}\phi(X_i)f(X_1,X_2,...,X_n,...)]$$
$$=\frac{1}{n}\sum_{i=1}^{n}E[\phi(X_i)f(X_1,X_2,...,X_n,...)]=E[\phi(X_1)f(X_1,X_2,...,X_n,...)] \qquad (4.3)$$

for all n-symmetric f defined on $(\mathbf{X}^\infty, \mathcal{X}^\infty)$. Since $\frac{1}{n}\sum_{i=1}^{n}\phi(X_i)$ is n-symmetric, (4.3) shows that

$$E[\phi(X_1)|\Sigma_n]=\frac{1}{n}\sum_{i=1}^{n}\phi(X_i).$$

Now consider the sequence of random variables

$$Z_n = E[\phi(X_1)|\Sigma_n],$$

for $n = 1, 2, ...$; (Z_n) is a reverse martingale with repect to the filtration (Σ_n). Therefore, on a set of probability one,

$$\lim_{n\to\infty} Z_n = E[\phi(X_1)|\Sigma_\infty]$$

(cfr. Ash (1972), Corollary 7.4.4). Thus,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\phi(X_i)=Y$$

with probability one if Y is a version of $E[\phi(X_1)|\Sigma_\infty]$. ◇

The result reached in the previous theorem could be stated in a different way. In fact, let $\mathcal{T}$ be the tail sigma-field of the sequence (X_n). That is

$$\mathcal{T}=\bigcap_{n=1}^{\infty}\sigma(X_n, X_{n+1},\cdots).$$

Then it is not difficult to show that

$$\mathcal{T}\subset\bigcap_{n=1}^{\infty}\Sigma_n=\Sigma_\infty.$$

Therefore, for every real valued, Borel measurable function ϕ defined on $\mathbf{X}$,

$$E[\phi(X_1)|\mathcal{T}]=E[E[\phi(X_1)|\Sigma_\infty]|\mathcal{T}]=E[\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\phi(X_i)|\mathcal{T}]=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\phi(X_i),$$

on a set of probability one, where the last equality holds because $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\phi(X_i)$ is $\mathcal{T}$-measurable. Hence

$$E[\phi(X_1)|\Sigma_\infty]=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\phi(X_i)=E[\phi(X_1)|\mathcal{T}]$$

with probability one. Note that for $A \in \mathcal{X}$ and ϕ the indicator function of A,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \phi(X_i) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}(A) = E[I[X_1 \in A]|\mathcal{T}] = \tau[X_1 \in A|\mathcal{T}]$$

where, for $y \in \mathbf{X}$, δ_y is the point mass at y and, given $X_1, ..., X_n$, $\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$ is the empirical probability distribution on $(\mathbf{X}, \mathcal{X})$ that assigns mass $1/n$ to each of the observed X_i's.

Since $\mathbf{X}$ is metric, separable and complete, there exists $p_1 : \mathbf{X}^\infty \times \mathcal{X} \to [0,1]$ a regular version of the conditional probability distribution of X_1 given $\mathcal{T}$. That is, for $A \in \mathcal{X}$ and $x = (x_1, x_2, \ldots) \in \mathbf{X}^\infty$,

$$p_1(x, A) = \tau[X_1 \in A|\mathcal{T}](x);$$

moreover, for all $A \in \mathcal{X}$, $p_1(\cdot, A)$ is $\mathcal{T}$-measurable whereas $p_1(x, \cdot)$ is a probability measure on $\mathcal{X}$, for all $x \in \mathbf{X}^\infty$.

4.4 **Theorem.** *Let (X_n) be exchangeable. Then, for $x = (x_1, x_2, ...)$ belonging to a set of probability one, the sequence of probability distributions $(\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(x)})$ weakly converges to $p_1(x, \cdot)$.*

Proof. Since $\mathbf{X}$ is metric, separable and complete, there exists a sequence (g_k) of real valued, uniformly continuous and bounded functions defined on $\mathbf{X}$ such that a sequence (μ_n) of probability measures on $(\mathbf{X}, \mathcal{X})$ weakly converges to a probability measure μ if and only if

$$\lim_{n\to\infty} \int_{\mathbf{X}} g_k d\mu_n = \int_{\mathbf{X}} g_k d\mu$$

for $k = 1, 2, ...$ (For a proof of this fact, see Parthasarathy (1967).)

For $n = 1, 2, ...$, $x = (x_1, x_2, ...) \in \mathbf{X}^\infty$, and $A \in \mathcal{X}$, let

$$\mu_n(x, A) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(x)}(A).$$

Fix k; then it follows from the previous theorem that there exists a $G_k \in \mathcal{X}^\infty$ such that $\tau(G_k) = 1$ and

$$\lim_{n\to\infty} \int_{\mathbf{X}} g_k(y)\mu_n(x, dy) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} g_k(X_i(x)) = E[g_k(X_1)|\mathcal{T}](x) = \int_{\mathbf{X}} g_k(y)p_1(x, dy)$$

for $x = (x_1, x_2, ...) \in G_k$.

Let $G = \bigcap_{k=1}^{\infty} G_k$. Then $\tau(G) = 1$ and

$$\lim_{n\to\infty} \int_{\mathbf{X}} g_k(y)\mu_n(x, dy) = \int_{\mathbf{X}} g_k(y)p_1(x, dy)$$

for $x \in G$ and $k = 1, 2, ...$ ◇

We now come to the representation for the finite marginal distributions of an exchangeable infinite sequence (X_n); that is, a representation of

$$\tau[X_1 \in A_1, ..., X_k \in A_k]$$

for $k \geq 1$ and $A_1, ..., A_k \in \mathcal{X}$. We begin by considering the case where $A_1, ..., A_k$ are elements of a finite measurable partition of $\mathbf{X}$.

Let thus $\{B_1, ..., B_{n+1}\}$ be a finite measurable partition of $\mathbf{X}$. That is:

(i) $B_1, ..., B_{n+1} \in \mathcal{X}$;

(ii) $\bigcup_{i=1}^{n+1} B_i = \mathbf{X}$;

(iii) $B_i \bigcap B_j = \emptyset$ for $i, j = 1, ..., n+1, i \neq j$.

For $N \geq 1$ and $N_1, ..., N_n \geq 0$, let $A_N(N_1, ..., N_n) \in \mathcal{X}^\infty$ be the event that is true if, among the first N random variables $X_1, ..., X_N$, N_1 belong to B_1,...,N_n belong to B_n and $N_{n+1} = N - \sum_{i=1}^n N_i$ belong to B_{n+1}. Let $B_N(N_1, ..., N_n) \in \mathcal{X}^\infty$ be the event that is true if N_1 assigned random variables among $X_1, ..., X_N$ belong to B_1, N_2 assigned random variables belong to B_2, ..., N_{n+1} assigned random variables belong to B_{n+1}. Then, as in the case for exchangeable sequences of events,

$$\tau(A_N(N_1, ..., N_n)) = \binom{N}{N_1\ N_2\ \cdots\ N_n} \tau(B_N(N_1, ..., N_n)).$$

Moreover, for $1 \leq r \leq N$ and $r_1 \geq 0, ..., r_n \geq 0$,

$$\tau[B_r(r_1, ..., r_n) | A_N(q_1, ..., q_n)] = \frac{\binom{N-r}{q_1 - r_1\ \cdots\ q_n - r_n}}{\binom{N}{q_1\ \cdots\ q_n}}$$

for $q_1 \geq r_1, ..., q_n \geq r_n$ and $\sum_{i=1}^n (q_i - r_i) \leq N - r$. Therefore

$$\tau(B_r(r_1, ..., r_n)) = \sum_{q_1 \geq r_1, ..., q_n \geq r_n, \sum_{i=1}^n (q_i - r_i) \leq N - r} \frac{\binom{N-r}{q_1 - r_1\ \cdots\ q_n - r_n}}{\binom{N}{q_1\ \cdots\ q_n}} \tau(A_N(q_1, ..., q_n)). \tag{4.5}$$

Let now F_N be the probability distribution function, with support on the simplex

$$S_n = \{(\theta_1, ..., \theta_n) : \theta_1 \geq 0, ..., \theta_n \geq 0, \sum_{i=1}^n \theta_i \leq 1\},$$

that assigns mass $\tau(A_N(q_1, ..., q_n))$ to the point $(q_1/N, ..., q_n/N)$, for $q_1, ..., q_n$ non negative integers and such that $\sum_{i=1}^n q_i \leq N$. In fact, F_N is the probability distribution of the random vector

$$(\frac{1}{N} \sum_{i=1}^N \delta_{X_i}(B_1), ..., \frac{1}{N} \sum_{i=1}^N \delta_{X_i}(B_n)).$$

We can then rewrite (4.5) as

$$\tau(B_r(r_1,...,r_n)) = \int_{S_n} I[(\theta_1,...,\theta_n) \in G_N] f_N(\theta_1,...,\theta_n) F_N(d\theta_1,...,d\theta_n)$$

where

$$G_N = \{(\theta_1,...,\theta_n) \in S_n : r_1 \le N\theta_1, ..., r_n \le N\theta_n, \sum_{i=1}^{n} \theta_i \le 1 - \frac{r - \sum_{i=1}^{n} r_i}{N}\}$$

and

$$f_N(\theta_1,...,\theta_n) = \frac{N^r[\prod_{i=1}^n \theta_i(\theta_i - \frac{1}{N})\cdots(\theta_i - \frac{r_i-1}{N})][(1-\sum_{i=1}^n \theta_i)\cdots(1-\sum_{i=1}^n \theta_i - \frac{r-\sum_{i=1}^n r_i}{N})]}{N(N-1)\cdots(N-r+1)},$$

for $(\theta_1,...,\theta_n) \in S_n$.

Notice that

$$\lim_{N\to\infty} I[(\theta_1,...,\theta_n) \in G_N] = I[(\theta_1,...,\theta_n) \in S_n]$$

and

$$\lim_{N\to\infty} f_N(\theta_1,...,\theta_n) = \theta_1^{r_1}\theta_2^{r_2}\cdots\theta_n^{r_n}(1-\sum_{i=1}^n \theta_i)^{r-\sum_{i=1}^n r_i}$$

uniformly in $(\theta_1,...,\theta_n) \in S_n$. Moreover, because of Theorem 4.4, the probability distribution F_N weakly converges to the distribution F of the random vector

$$(p_1(\cdot, B_1), ..., p_1(\cdot, B_n))$$

on a set of probability one. Then, arguing as in the proof of Theorem 2.11,

$$\begin{aligned}
&\tau(B_r(r_1,...,r_n)) = \\
&= \lim_{N\to\infty} \int_{S_n} I[(\theta_1,...,\theta_n) \in G_N] f_N(\theta_1,...,\theta_n) F_N(d\theta_1,...,d\theta_n) \\
&= \int_{S_n} \theta_1^{r_1}\theta_2^{r_2}\cdots\theta_n^{r_n}(1-\sum_{i=1}^n \theta_i)^{r-\sum_{i=1}^n r_i} F(d\theta_1,...,d\theta_n) \\
&= E[p_1^{r_1}(\cdot, B_1)p_1^{r_2}(\cdot, B_2)\cdots p_1^{r_n}(\cdot, B_n)[1-\sum_{j=1}^n p_1(\cdot, B_j)]^{r-\sum_{i=1}^n r_i}] \\
&= \int_{\mathbf{X}} p_1^{r_1}(x, B_1)p_1^{r_2}(x, B_2)\cdots p_1^{r_n}(x, B_n)[1-\sum_{j=1}^n p_1(x, B_j)]^{r-\sum_{i=1}^n r_i}\tau(dx) \\
&= \int_{\mathbf{P}} p^{r_1}(B_1)p^{r_2}(B_2)\cdots p^{r_n}(B_n)[1-\sum_{j=1}^n p(B_j)]^{r-\sum_{i=1}^n r_i}\nu(dp).
\end{aligned} \tag{4.6}$$

where ν is the probability distribution on $(\mathbf{P}, \mathcal{P})$ of the random probability $\tilde{p}_1 : (\mathbf{X}^\infty, \mathcal{X}^\infty) \to (\mathbf{P}, \mathcal{P})$ such that $\tilde{p}_1(x)(A) = p_1(x, A)$ for $x \in \mathbf{X}^\infty$ and $A \in \mathcal{X}$.

With the aid of (4.6), we may now move to the general representation of

$$\tau[X_1 \in A_1, ..., X_k \in A_k]$$

for $k \geq 1$ and $A_1, ..., A_k \in \mathcal{X}$. Our first step is to consider the class $\mathcal{C}$ of constituents generated by $A_1, ..., A_k$. Clearly $\mathcal{C}$ is a finite measurable partition of $\mathbf{X}$; assume that $\mathcal{C} = \{B_1, ..., B_{n+1}\}$. Then

$$\tau[X_1 \in A_1, ..., X_k \in A_k] = \sum \tau[X_1 \in B_{i_1}, ..., X_k \in B_{i_k}]$$

where the sum is over all the possible k-tuples $(B_{i_1}, ..., B_{i_k})$ such that $B_{i_j} \subseteq A_j$ for $j = 1, ..., k$. Each term appearing in the sum can be represented by means of (4.6). By summing these representations we obtain:

$$\tau[X_1 \in A_1, ..., X_k \in A_k] = \int_{\mathbf{P}} p(A_1)p(A_2) \cdots p(A_k)\nu(dp). \tag{4.7}$$

An example clarifies the argument. Suppose $k = 2$ and $A_1, A_2 \in \mathcal{X}$, $A_1 \neq A_2$. Then

$$B_1 = A_1 \bigcap A_2,\ B_2 = A_1 \bigcap A_2^c,\ B_3 = A_1^c \bigcap A_2,\ B_4 = A_1^c \bigcap A_2^c.$$

and

$$A_1 = B_1 \bigcup B_2,\ A_2 = B_1 \bigcup B_3.$$

Moreover

$$\begin{aligned}
&\tau[X_1 \in A_1, X_2 \in A_2] \\
&\quad = \tau[X_1 \in B_1, X_2 \in B_1] + \tau[X_1 \in B_1, X_2 \in B_3] + \tau[X_1 \in B_2, X_2 \in B_1] + \tau[X_1 \in B_2, X_2 \in B_3] \\
&\quad = \int_{\mathbf{P}} p^2(B_1)\nu(dp) + \int_{\mathbf{P}} p(B_1)p(B_3)\nu(dp) + \int_{\mathbf{P}} p(B_2)p(B_1)\nu(dp) + \int_{\mathbf{P}} p(B_2)p(B_3)\nu(dp) \\
&\quad = \int [p(B_1) + p(B_2)][p(B_1) + p(B_3)]\nu(dp) \\
&\quad = \int_{\mathbf{P}} p(A_1)p(A_2)\nu(dp).
\end{aligned}$$

Note that for the second equality we used the representation (4.6).

We are finally ready for a first version of de Finetti's representation theorem.

4.8 **Theorem**. *The sequence (X_n) is exchangeable if and only if there exists a probability ν on $(\mathbf{P}, \mathcal{P})$ such that*

$$\tau[(X_1, X_2, ...) \in A] = \int_{\mathbf{P}} p^\infty(A)\nu(dp) \tag{4.9}$$

for all $A \in \mathcal{X}^\infty$. Moreover, ν is the probability distribution of the weak limit of the sequence $(\frac{1}{n}\sum_{i=1}^n \delta_{X_i})$; hence is unique.

Proof. If (4.9) holds for the law of the sequence (X_n), Proposition 3.4 shows that (X_n) is exchangeable.

Viceversa, assume that (X_n) is exchangeable; then (4.7) holds. Therefore

$$\begin{aligned}
&\tau[X_1 \in A_1, ..., X_k \in A_k] = \\
&\quad = \int_{\mathbf{P}} p(A_1)p(A_2) \cdots p(A_k)\nu(dp) = \int_{\mathbf{P}} p^\infty(A_1 \times A_2 \times \cdots A_k \times \mathbf{X}^\infty)\nu(dp)
\end{aligned}$$

for $k \geq 1$ and $A_1, ..., A_k \in \mathcal{X}$. This is enough to prove that (4.9) holds.

The fact that ν is the probability distribution of the weak limit of the sequence $(\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i})$ follows from Theorem 4.4. $\diamond$

The probability distribution ν appearing in (4.9) is called *de Finetti measure* of the exchangeable sequence (X_n) (or *prior* probability distribution in Bayesian inference.) Another version of de Finetti's representation theorem, more commonly encountered in the literature, is the following for which we follow Kingman (1974).

4.10 **Theorem.** *The sequence (X_n) is exchangeable if and only if there exists a random probability measure $\tilde{p}_1 : (\mathbf{X}^\infty, \mathcal{X}^\infty) \to (\mathbf{P}, \mathcal{P})$ such that, given $\tilde{p}_1$, the random variables of the sequence (X_n) are conditionally i.i.d. with distribution $\tilde{p}_1$.*

Proof. Suppose that there exists a $\tilde{p}_1 : (\mathbf{X}^\infty, \mathcal{X}^\infty) \to (\mathbf{P}, \mathcal{P})$ such that, given $\tilde{p}_1$, the random variables of the sequence (X_n) are conditionally i.i.d. with distribution $\tilde{p}_1$. Then the law of (X_n) admits the representation (3.3) and it is thus exchangeable.

Viceversa, assume that (X_n) is exchangeable. Let $1 \leq k \leq n$, $i_1, ..., i_k \in \{1, ..., n\}$ distinct and $\phi_1, ..., \phi_k$ real valued, measurable and bounded function defined on $(\mathbf{X}^\infty, \mathcal{X}^\infty)$. For $n \geq 1$ and any n-symmetric f

$$\begin{aligned} E[\phi_1(X_{i_1}) \cdots \phi_k(X_{i_k}) f(X_1, X_2, ..., X_n, X_{n+1}, ...)] = \\ = E[\phi_1(X_1) \cdots \phi_k(X_k) f(X_{i_1}, X_{i_2}, ..., X_{i_k}, ..., X_n, X_{n+1}, ...)] \\ = E[\phi_1(X_1) \cdots \phi_k(X_k) f(X_1, X_2, ..., X_n, X_{n+1}, ...)] \end{aligned}$$

where the first equality is true because of exchangeability and the second holds since f is n-symmetric. By the same argument as in the proof of Theorem 4.2, one then shows that

$$\lim_{n\to\infty} \frac{1}{n(n-1)\cdots(n-k+1)} \sum_{1\leq i_1 \neq i_2 \neq \cdots \neq i_k \leq n} \phi_1(X_{i_1}) \cdots \phi_k(X_{i_k}) = E[\phi_1(X_1)\cdots\phi_k(X_k)|\Sigma_\infty] \tag{4.11}$$

with probability one.

Let $A_1, ..., A_k \in \mathcal{X}^\infty$. Then (4.11) implies that

$$\lim_{n\to\infty} \frac{1}{n(n-1)\cdots(n-k+1)} \sum_{1\leq i_1 \neq i_2 \neq \cdots \neq i_k \leq n} I[X_{i_1} \in A_1] \cdots I[X_{i_k} \in A_k] = \tau[X_1 \in A_1, ..., X_k \in A_k|\Sigma_\infty] \tag{4.12}$$

with probability one. Notice that

$$\sum_{1\leq i_1 \neq i_2 \neq \cdots \neq i_k \leq n} I[X_{i_1} \in A_1] \cdots I[X_{i_k} \in A_k]$$

$$= \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_k=1}^{n} I[X_{i_1} \in A_1]\cdots I[X_{i_k} \in A_k] - \sum_{1\le i_1=i_2\neq\cdots\neq i_k\le n} I[X_{i_1} \in A_1]\cdots I[X_{i_k} \in A_k] - \cdots$$
$$\cdots - \sum_{1\le i_1\neq i_2\neq\cdots\neq i_{k-1}=i_k\le n} I[X_{i_1} \in A_1]\cdots I[X_{k-1} \in A_{k-1}]I[X_{i_k} \in A_k] - \cdots$$
$$\cdots - \sum_{i_1=1}^{n} I[X_{i_1} \in A_1]\cdots I[X_{i_1} \in A_k].$$

Except for the first term, all the other terms in the sum on the right side of the previous equality are of order less than or equal to n^{k-1}. Therefore

$$\lim_{n\to\infty} \frac{1}{n(n-1)\cdots(n-k+1)} \sum_{1\le i_1\neq i_2\neq\cdots\neq i_k\le n} I[X_{i_1} \in A_1]\cdots I[X_{i_k} \in A_k]$$
$$= \lim_{n\to\infty} \frac{1}{n^k}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_k=1}^{n} I[X_{i_1} \in A_1]\cdots I[X_{i_k} \in A_k]$$
$$= \lim_{n\to\infty} \prod_{j=1}^{n}[\frac{1}{n}\sum_{i=1}^{n} I[X_i \in A_j]] \tag{4.13}$$

with probability one. However, because of Theorem 4.2 and the argument following its proof, for $j = 1, ..., k$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} I[X_i \in A_j] = \tau[X_1 \in A_j|\mathcal{T}] \tag{4.14}$$

with probability one. Equations (4.12), (4.13) and (4.14) imply that

$$\prod_{j=1}^{n} \tau[X_1 \in A_j|\mathcal{T}] = \tau[X_1 \in A_1, ..., X_k \in A_k|\Sigma_\infty]$$

with probability one. Since $\mathcal{T} \subseteq \Sigma_\infty$,

$$\tau[X_1 \in A_1, ..., X_k \in A_k|\mathcal{T}] = E[\tau[X_1 \in A_1, ..., X_k \in A_k|\Sigma_\infty]|\mathcal{T}] = \prod_{j=1}^{n} \tau[X_1 \in A_j|\mathcal{T}]$$

on a set of probability one. Let $\tilde{p}_1 : (\mathbf{X}^\infty, \mathcal{X}^\infty) \to (\mathbf{P}, \mathcal{P})$ be a random probability defined for $x \in \mathbf{X}^\infty$ and $A \in \mathcal{X}$ by $\tilde{p}_1(x)(A) = p_1(x, A)$. Then $\tilde{p}_1$ is measurable with respect to $\mathcal{T}$. Therefore

$$\tau[X_1 \in A_1, ..., X_k \in A_k|\tilde{p}_1]$$
$$= E[\tau[X_1 \in A_1, ..., X_k \in A_k|\mathcal{T}]|\tilde{p}_1] = E[\prod_{j=1}^{n} \tau[X_1 \in A_j|\mathcal{T}]|\tilde{p}_1] = \prod_{j=1}^{n} \tilde{p}_1(A_j)$$

with probability one. This proves that, given $\tilde{p}_1$, the random variables of the sequence (X_n) are i.i.d. with probability distribution $\tilde{p}_1$. ◇

4.15 **Example.** Let (X_n) be a sequence of identically distributed gaussian random variables with values in $\Re$ and suppose the correlation coefficient ρ between any two different variables of the sequence is constant. Without loss of generality we assume that, for every n, X_n has mean 0 and variance 1.

If $\rho = 0$, the random variables of the sequence (X_n) are i.i.d. N$(0, 1)$, and the sequence is trivially exchangeable. So, let us focus on the case where $\rho \neq 0$; notice that it must be $\rho > 0$ since, for all $n \geq 1$,

$$0 \leq \text{Var}(\frac{\sum_{i=1}^{n} X_i}{n}) = \frac{1}{n} + \frac{n(n-1)}{n^2}\rho.$$

The sequence (X_n) is exchangeable. In fact, for $n \geq 1$, compute the characteristic function of $(X_1, ..., X_n)$,

$$\phi_{X_1,...,X_n}(t_1, ..., t_n) = \exp\Big[-\frac{1}{2}[(1-\rho)\sum_{i=1}^{n} t_i^2 + \rho(\sum_{i=1}^{n} t_i)^2]\Big] = \phi_{X_1,...,X_n}(t_{\sigma_1}, ..., t_{\sigma_n})$$

for every $t_1, ..., t_n \in \Re$ and every permutation $(\sigma_1, ..., \sigma_n)$ of the integers $(1, ..., n)$. Exchangeability now follows from Proposition 3.6. Note that

Theorem 4.2 implies the existence of a real random variable μ such that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu$$

with probability one. Indeed, μ has distribution N$(0, \rho)$ since, for $n \geq 1$, $n^{-1}\sum_{i=1}^{n} X_i$ has distribution N$(0, n^{-1} + n^{-1}(n-1)\rho)$.

For $n \geq 1$, we now compute the characteristic function of $(X_1, ..., X_n, \mu)$. Let $t_1, ..., t_n, \theta \in \Re$; then

$$\phi_{X_1,...,X_n,\mu}(t_1, ..., t_n, \theta) = \lim_{N\to\infty} \phi_{X_1,...,X_n,\frac{1}{N}\sum_{i=1}^{n} X_i}(t_1, ..., t_n, \theta).$$

However, for $N > n$,

$$\begin{aligned}
&\phi_{X_1,...,X_n,\frac{1}{N}\sum_{i=1}^{n} X_i}(t_1, ..., t_n, \theta) \\
&\quad = E\Big[\exp[i\sum_{i=1}^{n}(t_i + \frac{\theta}{N})X_i + i\frac{\theta}{N}\sum_{i=n+1}^{N} X_i]\Big] \\
&\quad = \exp\Big[-\frac{1}{2}[(1-\rho)\sum_{i=1}^{n}(t_i + \frac{\theta}{N})^2 + (1-\rho)\frac{\theta^2}{N^2}(N-n) + \rho(\sum_{i=1}^{n} t_i + \frac{\theta}{N}(N-n))^2]\Big].
\end{aligned}$$

Hence

$$\begin{aligned}
\phi_{X_1,...,X_n,\mu}(t_1, ..., t_n, \theta) &= \exp\Big[-\frac{1}{2}[(1-\rho)\sum_{i=1}^{n} t_i^2 + \rho(\sum_{i=1}^{n} t_i + \theta)^2]\Big] \\
&= \exp\Big[-\frac{1}{2}[(1-\rho)\sum_{i=1}^{n} t_i^2]\Big]E\Big[\exp[i\mu(\sum_{i=1}^{n} t_i + \theta)]\Big] \\
&= E\Big[\exp[i\mu\theta]\prod_{i=1}^{n}\exp[i\mu t_i - \frac{1}{2}(1-\rho)t_i^2]\Big]. \qquad (4.16)
\end{aligned}$$

The next to the last equality holds because we know already that μ has $N(0, \rho)$ distribution. The previous representation for the characteristic function of $(X_1, ..., X_n, \mu)$ proves that, conditionally on μ, the random variables of the sequence (X_n) are independent and identically distributed with distribution $N(\mu, 1-\rho)$.

Let us indicate with Φ_{m,σ^2} the Normal probability distribution with mean m and variance σ^2 and with φ_{m,σ^2} its density. If τ is the law of the process (X_n), equation (4.16) implies that

$$\tau[(X_1, X_2, ...) \in A] = \int_{-\infty}^{\infty} \Phi_{m,1-\rho}^{\infty}(A)\varphi_{0,\rho}(m)dm, \tag{4.17}$$

for all $A \in \mathcal{B}^{\infty}$. In view of Theorem 4.8, how shall we interpret this representation for τ?

Let $(\mathbf{P}, \mathcal{P})$ be the space whose elements are probabilities on $(\Re, \mathcal{B})$ endowed, as before, with the sigma-field generated by the topology of weak convergence. Observe that, for all Borel sets $B \in \mathcal{B}$,

$$\{\Phi_{m,1-\rho} : m \in B\} \in \mathcal{P}.$$

In fact, it is easy to show that $\{\Phi_{m,1-\rho}\} \in \mathcal{P}$ for every $m \in \Re$. Hence, for $B \in \mathcal{B}$,

$$\{\Phi_{m,1-\rho} : m \in B\} = \overline{\{\Phi_{m,1-\rho} : m \in Q \bigcap B\}} \in \mathcal{P}$$

where Q indicates the set of rational numbers and $\overline{A}$ is the closure of A. Notice that the first equality holds since if, $m_n \in Q \bigcap B$ for $n \geq 1$ and the sequence (m_n) converges to m, then the sequence $(\Phi_{m_n,1-\rho})$ weakly converges to $\Phi_{m,1-\rho}$. Now set ν to be a probability on $(\mathbf{P}, \mathcal{P})$ such that

$$\nu(\{\Phi_{m,1-\rho} : m \in B\}) = \Phi_{0,\rho}(B)$$

for all $B \in \mathcal{B}$. Then (4.17) is equivalent to

$$\tau[(X_1, X_2, ...) \in A] = \int_{\mathbf{P}} p^{\infty}(A)\nu(dp)$$

and, because of Theorem 4.8, ν is the unique probability distribution of the weak limit of the sequence $(\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i})$.

In many applications in statistics, when facing an exchangeable sequence (X_n) one assumes that the probability ν appearing in the representation (2.14) is such that there exist a random vector $\Theta \in \Re^k$ with probability distribution G, and a family $\{F_\theta \in \mathbf{P} : \theta \in \Re^k\}$ such that, for $B \in \mathcal{B}^k$, $\{F_\theta : \theta \in B\} \in \mathcal{P}$ and

$$\nu(\{F_\theta : \theta \in B\}) = G(B).$$

Then, conditionally on Θ, the random variables of the sequence (X_n) are independent and identically distributed with distribution F_Θ and representation (2.14) becomes equivalent to

$$\tau[(X_1, X_2, ...) \in A] = \int_{\Re^k} F_\theta^{\infty}(A)G(d\theta);$$

in this case we speak of a *parametric model* for the exchangeable sequence (X_n). This will be the topic of a future paper.

References

[1] ASH, R.B. (1972). *Real Analysis and Probability*, Academic Press, New York.

[2] DE FINETTI, B. (1928). Funzione caratteristica di un fenomeno aleatorio. *Atti del Convegno Internazionale dei Matematici*, Zanichelli, Bologna, 1932, Vol. 6, 179–190.

[3] DE FINETTI, B. (1930). Funzione caratteristica di un fenomeno aleatorio. *Memorie della R. Accademia dei Lincei*, IV, fasc. V, 86-133.

[4] DE FINETTI, B. (1933a). Classi di numeri aletori equivalenti. *Rendiconti della R. Accademia Nazionale dei Lincei*, 18, 107–110.

[5] DE FINETTI, B. (1933b). La legge dei grandi numeri nel caso dei numeri aleatori equivalenti. *Rendiconti della R. Accademia Nazionale dei Lincei*, 18, 203–207.

[6] DE FINETTI, B. (1933c). Sulla legge di distribuzione dei valori in una successione di numeri aleatori equivalenti. *Rendiconti della R. Accademia Nazionale dei Lincei*, 18, 279–294.

[7] DE FINETTI, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institute Henry Poincaré*, tome VII, fasc. I, 1–68.

[8] HEWITT, E. and SAVAGE, L.J. (1955). Symmetric measures on cartesian products. *Trans. Amer. Math. Soc.*, 80, 470–501.

[9] PARTHASARATHY, K.R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York.