ASYMPTOTICS FOR RELATIVE
FREQUENCY WHEN POPULATION IS
DRIVEN BY ARBITRARY EVOLUTION

Silvano Fiorin

Dipartimento di Scienze Statistiche, via C. Battisti, 241-35121
Padova, Italy - fiorin@stat.unipd.it
**ABSTRACT** Strongly consistent estimates are shown, via relative frequency, for the probability of "white balls" inside a dichotomous urn when such a probability is an arbitrary continuous time dependent function over a bounded time interval. The asymptotic behaviour of relative frequency is studied in a nonstationary context using a Riemann-Dini type theorem for SLLN of random variables with arbitrarily different expectations; furthermore, the theoretical results concerning the SLLN can be applied for estimating the mean function of unknown form of a general nonstationary process.

1 **INTRODUCTION**

Several different areas of statistics deal with an urn model including "white" and "black" balls with probability $p$ and $1 - p$ respectively. In this very classical context a time dependent component is introduced: $p$ is replaced with $p_0(t)$ which denotes a time varying quantity $0 \leq p_0(t) \leq 1$ in such a way that at any instant $t \in [0, T]$ only one observation is taken from the corresponding urn with probability $p_0(t)$ and the random variable $Y(t)$ is obtained such that $P(Y(t) = 1) = p_0(t), P(Y(t) = 0) = 1 - p_0(t), E(Y(t)) = p_0(t) \forall t \in [0, T]$, defining the nonstationary process

$$\{Y(t) : t \in [0, T]\}$$

with mean function $E(Y(t)) = p_0(t)$. The description of the above model is specified introducing some reasonable assumptions:

**A 1** the continuity is assumed for the usually unknown mean function $p_0 : [0, T] \mapsto [0, 1]$;

**A 2** for any fixed pair of instants $t_1, t_2 \in [0, T]$ the independence is assumed for the random variables $Y(t_1)$ and $Y(t_2)$.

This assumption is introduced in order to apply the Rajchman theorem (see next section). Namely: only pairwise uncorrelation is requested for $Y(t_1)$ and $Y(t_2)$ but, it can be easily checked in this case, the uncorrelation implies independence; furthermore independence is here a very mild condition: in fact we may suppose that the total number of white and black balls in the urn is big enough that the knowledge of $Y(t_1) = 1$ or $Y(t_1) = 0$ does not produce a meaningful modification of the probability distribution for $Y(t_2)$.

The main purpose is estimating the unknown function $p_0$, i.e. the mean function $p_0(t) = E(Y(t))$ of the nonstationary process $\{Y(t) : t \in [0, T]\}$. which is an arbitrary
continuous map form \([0, T]\) into \([0, 1]\).

i) An approach to estimation for the mean function \(m(.)\) of a nonstationary process was given by M.B. Priestley (see [5] at page 587 and [6] at page 140) when the form of \(m\) is known and the case is suggested of a polynomial function in \(t\). Vice versa: "with no information on the form of \(m\) we obviously cannot construct a consistent estimate of it". The approach here adopted is quite different from classical methods of time series analysis; the only information available for \(m\) is the continuity property over \([0, T]\) and no approximation of \(m\) is introduced by continuous functions of a known form. The estimation technique involves the process \(\Pi\) which is a specified case of nonstationarity but the theoretical results given in the last section hold true for a general nonstationary process. The case \(\Pi\) is only a concrete example of a process having no regularity properties; nevertheless the continuity for the mean function \(m\) is a reasonable and not restrictive assumption which denotes compatibility with a context of an arbitrary but not brutal evolution for the composition of the urn.

ii) The urn evolution has effects concerning sampling; for instance if the observations number \(n\) is big enough a not slight time interval will be needed in order to receive the \(n\) observations which surely are not values taken by the same random variable. Then, for sake of simplification, we assume that any r.v. \(Y(t)\) may be observed at most only one time. The point of view we adopt is then characterized by a strong nonstationarity and the consistent estimation for the mean \(m(t_0)\) at a fixed time \(t_0\) may appear as a very hard objective.

iii) The answer to above arguments is the relative frequency

\[
\frac{1}{n}\sum_{j=1}^{n} Y(t_j)
\]  

(2)

where \(\{t_j : j = 1, ..., n\}\) are the first \(n\) observation times of a sequence \(\{t_j : j \geq 1\} \subset [0, T]\) and the main purpose is that of getting consistent estimations of \(m(t) = p_0(t)\) via almost sure convergence for the sequence (2). The SLLN is then the theoretical tool needed in the below analysis, but the classical approach based on the zero-mean r.v.'s \((Y(t_j) - p_0(t_j))\), i.e.

\[
\frac{1}{n}\sum_{j=1}^{n} (Y(t_j) - p_0(t_j)) \to 0 \text{ a.s.}
\]  

(3)

is not enough; in fact we need convergence for (2) with the not zero mean r.v.'s \(Y(t_j)\). This argument, investigated by Fiorin [4] is now improved with the help of new results given in section (5).

iv) The convergence of (2) is studied via the sequence \(\{E(Y(t_j)) = p_0(t_j) :\)
\( j \geq 1 \) and permutations (i.e. bijections) \( \pi : N \to N \) in fact, if a permutation \( \pi \) is introduced, the possible almost sure limit of

\[
\frac{1}{n} \sum_{j=1}^{n} Y(t_{\pi(j)})
\]  

(4)

is depending on \( \pi \). If \( \{P_{\pi_n}^0\} \) is a sequence of probability measures, where each \( P_{\pi_n}^0 \) assigns mass \( \frac{1}{n} \) to each point \( \{p_0(t_{\pi(j)}) : j = 1, \ldots, n\} \), then the ”weak” or ”vague” convergence for the sequence \( \{P_{\pi_n}^0\} \) to a probability measure \( P^0 \) implies almost sure convergence of (4) to the limit \( \int_0^1 I(v) dP^0(v) \) where \( I(v) \) is the identity map over \([0, 1]\) and \( P^0 \) depends on the sequence \( \{Y(t_j) : j \geq 1\} \) and on permutation \( \pi \). All the below analysis is based on the possibility of finding a permutation \( \pi \) in such a way that the convergence of (4) is driven to a limit \( \int_0^1 I(v) dP^0(v) \) where \( P^0 \) is a previously chosen probability measure over \([0, 1]\) under a theoretical point of view this is a result for SLLN (4) which is the analogous of the well known Riemann-Dini theorem for real simply convergent (but not absolutely convergent) series. Under the operative point of view the strongly consistent estimates, i.e. the a.s. limits \( \int_0^1 I(v) dP^0(v) \), are the result of an experimental design based on choosing:

I) the sequence of observation times \( \{t_j : j \geq 1\} \subset [0, T] \);
II) the permutation \( \{t_{\pi(j)} : j \geq 1\} \).

2 CONVERGENCE ELEMENTS

If the observation times \( \{t_j : j \geq 1\} \) are given jointly with the observable r.v.’s \( \{Y(t_j) : j \geq 1\} \), an intuitive approach for studying the almost sure convergence for (2) is suggested by the classical Rajchman theorem

**Theorem 1** If the \( Y(t_j) \)'s are pairwise uncorrelated and their second moments have a common bound then

\[
\frac{1}{n} \sum_{j=1}^{n} (Y(t_j) - p_0(t_j))
\]

is convergent to 0 almost surely.

Because of assumption A2) and the inequality \( |Y(t_j)| \leq 1 \) the \( Y(t_j) \)'s satisfy theorem (1) and then

\[
\frac{1}{n} \sum_{j=1}^{n} (Y(t_j) - p_0(t_j)) \to 0 \ a.s.
\]  

(5)
Now an intuitive and simple condition which implies (together with (5)) the almost sure convergence for (2) is the possible limit for the deterministic sequence
\[
\frac{1}{n} \sum_{j=1}^{n} p_0(t_j) = \frac{1}{n} \sum_{j=1}^{n} E(Y(t_j)).
\] (6)
In fact, if such a limit exists, i.e.
\[
L = \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} p_0(t_j),
\]
we have
\[
\frac{1}{n} \sum_{j=1}^{n} Y(t_j) \to L \text{ a.s.}
\]

**Definition 1** Let us define as a "pseudoempirical measure" (P.E.M. hereafter) any probability measure giving the weight \( \frac{1}{n} \) to each of the assigned points \( \{x_j : j = 1, \ldots, n\} \), where the "pseudo" means that the \( x_j \)'s are arbitrarily fixed deterministic values and not a sequence of i.i.d. observations.

The notion of "Vague Convergence" (V.C. hereafter) is introduced mainly for application to sequences of P.E.M.'s; such a concept, which implies existence of limit \( L \) for the sequence (6), is the main technical tool for studying the asymptotic behaviour of relative frequency (2). Only the really necessary elements for below analysis are here given; for an exhaustive exposition see Chung [3].

**Definition 2** A sequence \( \{\mu_n : n \geq 1\} \) of probability measures (P.M. hereafter) defined over the Borel \( \sigma \)-field \( B^1 \) of \( R^1 \) is said to converge vaguely to the P.M. \( \mu \) iff there exists a dense subset \( D \) of \( R^1 \) such that
\[
\mu_n(a, b] \to \mu(a, b], \forall a \in D, b \in D, a < b.
\]

**Theorem 2** (see Theorem 4.3.1, page 85 Chung [3]) The sequence of P.M.'s \( \mu_n \) is vaguely convergent to the P.M. \( \mu \) if and only if
\[
\lim_{n \to \infty} \mu_n(a, b] = \mu(a, b]
\]
for every continuity interval \( (a, b] \) of \( \mu \), i.e. for every interval whose endpoints satisfy \( \mu(a) = \mu(b) = 0 \).

By theorem (2) the equivalence is stated between vague and weak convergence for P.M.'s \( \mu_n \) to \( \mu \). A further classical result needed in the below proofs is the following characterization of V.C.:
**Theorem 3** (see Theorem 4.4.2., page 93 Chung [3]) \( \mu_n \) is vaguely convergent to \( \mu \) if and only if the convergence is stated

\[
\lim_{n \to \infty} \int_R f \, d\mu_n = \int_R f \, d\mu
\]

for each bounded, continuous and real \( f \).

Even if vague and weak convergence of P.M.’s are equivalent, in the main proofs the V.C. is preferable because the convergence has to be proved \( \mu_n(a, b) \to \mu(a, b) \) for countably many \( a, b \) in a dense subset of \( R \).

The above theorem (3) can be directly applied for convergence of sequence (6) via the equality

\[
\frac{1}{n} \sum_{j=1}^{n} p_0(t_j) = \int_0^1 p \, dP^0_n
\]

where \( P^0_n \) is the P.E.M. giving weight \( \frac{1}{n} \) to each point \( \{p_0(t_j) : j = 1, ..., n\} \). Thus a condition which implies the convergence of (6) is the vague convergence for the sequence of P.E.M.’s \( P^0_n \) to a P.M. \( P^0 \). In fact if \( P^0_n \) is V.C. to \( P^0 \), having \( p_0(t_j) \in [0, 1] \forall j \), and taking the function

\[
\begin{align*}
  f(p) &= p \quad \forall p \in [0, 1] \\
  f(p) &= 1 \quad \forall p \in [1, +\infty) \\
  f(p) &= 0 \quad \forall p \in (-\infty, 0],
\end{align*}
\]

by theorem (3), the convergence is stated

\[
\int_0^1 p \, dP^0_n \to \int_0^1 p \, dP^0
\]

i.e.

\[
\frac{1}{n} \sum_{j=1}^{n} p_0(t_j) \to \int_0^1 p \, dP^0
\]

which jointly with theorem (1) implies

\[
\frac{1}{n} \sum_{j=1}^{n} Y(t_j) \to \int_0^1 p \, dP^0 \text{ a.s.}
\]

**Remark 1** For the almost sure convergence (9) an alternative proof is given by theorem (6) below: working with the sequence \( \frac{1}{n} \sum_{j=1}^{n} Y(t_j) \) its direct approximation to the integral \( \int_0^1 p \, dP^0 \) is proved.
The central argument concerning convergence (9) is the assumption of vague convergence for \( P_n^0 \) to \( P^0 \). Several questions may arise: for instance it is evident that such a condition is not so easy to reach. In fact the restrictivity of this assumption will be evident via Definition (2), for an assigned sequence of expectations \( \{ E(Y(t_j)) = p_0(t_j) : j \geq 1 \} \) and a fixed interval \( (a, b] \subset [0, 1] \) the convergence \( P_n^0(a, b] \to P^0(a, b] \) holds true where

\[
P_n^0(a, b) = \frac{n(a, b)}{n}
\]

and \( n(a, b] \) is the total number of points \( \{ p_0(t_j) : j = 1, ..., n \} \) belonging to \( (a, b] \); this means that inside the first \( n \) elements of the sequence \( \{ p_0(t_j) : j \geq 1 \} \) the proportion of points falling into \( (a, b] \) is "so regular" to approach a limit \( P^0(a, b] \), when \( n \to \infty \). And this for an arbitrary deterministic sequence \( \{ p_0(t_j) : j \geq 1 \} \subset [0, 1] \).

Our purpose, in the sequel, will consist of a strategy to obtain a vaguely convergent sequence of P.E.M.'s \( P_n^0 \); recalling I) and II) at the end of introduction, we may choose an experimental design which consists of two steps; we may decide when to observe the continuous time process \( \{ Y(t) : t \in [0, T] \} \) and then we choose the observation times consisting of a sequence \( \{ t_j : j \geq 1 \} \subset [0, T] \). Not only: we may decide also, for each \( n \) fixed, the \( n \) observable r.v.'s to choose inside \( \{ Y(t_j) : j \geq 1 \} \), i.e. we do not consider necessarily the first \( n \) r.v.'s \( \{ Y(t_j) : j = 1, ..., n \} \) but we select \( \{ Y(t_{j}) : j = 1, ..., n \} \) with the respective expectations \( \{ E(Y(t_{j}) : j = 1, ..., n) \} \) where \( \{ \pi(j) : j = 1, ..., n \} \) are the first \( n \) values taken by a permutation (a bijection) \( \pi : N \to N \), in such a way that, if \( P_{n} \) denotes the P.E.M. giving mass \( \frac{1}{n} \) to each point \( \{ t_{j} : j = 1, ..., n \} \), the sequence \( P_{n} \) is vaguely or weakly convergent to some P.M. \( P_{\pi} \).

Then, using the relevant property that the induced measures \( p_0(P_{n}) \)'s and \( P_{\pi} \)'s keep the weak convergence, we reach the V.C. \( p_0(P_{n}) \to p_0(P_{\pi}) \), where \( p_0(P_{n}) \) assigns mass \( \frac{1}{n} \) to each point \( \{ p_0(t_{j}) : j = 1, ..., n \} \). But, for a complete description of the above strategy, we need to introduce the relevant tool of permutations.

### 3 PERMUTATIONS

Given the family of r.v.'s \( \{ Y(t_{j}) : j \geq 1 \} \) with expectations \( \{ E(Y(t_{j})) = p_0(t_{j}) : j \geq 1 \} \), for any assigned bijection \( \pi : N \to N \) the respective process may be defined

\[
\{ Y(t_{\pi(j)}) : j \geq 1 \}
\]

with expectations \( \{ E(Y(t_{\pi(j)}) = p_0(t_{\pi(j)}) : j \geq 1 \} \) and the P.E.M.'s \( P_{n}^{0} \pi \), which gives mass \( \frac{1}{n} \) to each point \( \{ p_0(t_{\pi(j)}) : j = 1, ..., n \} \). A direct comparison shows
that $P^n_0$ and $P^n_0\pi_n$, in the general case, define different probability measures over the Borel $\sigma$-field $B[0, 1]$. Consequently the possible vague limits $P^n_0$ and $P^n_0\pi_n$, if they exist, are different P.M.’s and, applying Theorem (3) and Theorem (6), we have:

$$\frac{1}{n} \sum_{j=1}^{n} p_0(t_j) \rightarrow \int_0^1 p dP^n_0 \quad \text{and} \quad \frac{1}{n} \sum_{j=1}^{n} Y(t_j) \rightarrow \int_0^1 p dP^n_0 \text{a.s.}$$

and using permutation $\pi$

$$\frac{1}{n} \sum_{j=1}^{n} p_0(t_{\pi(j)}) \rightarrow \int_0^1 p dP^n_\pi \quad \text{and} \quad \frac{1}{n} \sum_{j=1}^{n} Y(t_{\pi(j)}) \rightarrow \int_0^1 p dP^n_0 \pi \text{a.s..}$$

Permutations are an important argument in below analysis with several implications concerning estimation; then this topic needs further attention: the vague convergence for a sequence of P.M.’s $P^n_0\pi_n$ was introduced above only as an hypothesis. Now, in order to obtain an estimation procedure, the following three steps have to be examined:

1) the vague convergence for an assigned sequence of P.E.M.’s $P^n_0\pi_n$ has really to be proved.

2) Given the sequence of points $\{t_j : j \geq 1\} \subset [0, T]$, the class $\mathcal{M}$ has to be found of P.M.’s $P$ over $B[0, T]$ for which a permutation $\{t_{\pi(j)} : j \geq 1\}$ can be computed such that the P.E.M.’s $P^n_\pi$ (which assigns weight $\frac{1}{n}$ to each point $\{t_{\pi(j)} : j = 1, ..., n\}$) are vaguely convergent to $P$ and then the induced measures $p_0(P^n_\pi)$ over $B[0, 1]$ are vaguely convergent to $p_0(P)$ (because of continuity of $p_0$), where

$$p_0(P^n_\pi)(B) = P^n_\pi(p_0^{-1}(B)) \quad \text{and} \quad p_0(P)(B) = P(p_0^{-1}(B)) \forall B \in B[0, 1].$$

3) The possibility of choosing a measure $P \in \mathcal{M}$, and then of computing a permutation $\{t_{\pi(j)} : j \geq 1\}$ such that, applying theorem (6),

$$\frac{1}{n} \sum_{j=1}^{n} Y(t_{\pi(j)}) \rightarrow \int_0^1 p dP_0(P) \text{a.s.,}$$

is a good chance for consistent estimation: through the choice of the vague limit measure $P$ and of $\pi$ the convergence for the SLLN may be driven to different limit values.

A rigorous characterization of class $\mathcal{M}$ is given by definition (5) which needs more technical details given later; nevertheless it may be useful to anticipate the content of assumption under which $\mathcal{M}$ contains infinitely many measures: if the set of points $\{t_j : j \geq 1\} \subset [0, T]$ has at least two different limit values, i.e. if there are at least two values $L_1 \neq L_2$ such that there exist two subsequences

$$\lim_{k \to \infty} t_{j_1(k)} = L_1 \quad \text{and} \quad \lim_{k \to \infty} t_{j_2(k)} = L_2,$$
then $\mathcal{M}$ contains infinitely many probability measures. Furthermore, for an assigned measure $P \in \mathcal{M}$, the procedure of finding a permutation $\{t_{\pi(j)} : j \geq 1\}$ such that the respective P.E.M.'s $P_{\pi n}$ are vaguely convergent to the assigned $P$ is available in the proof of theorem (7). Our aim consists now in applying the above results for estimation.

4 ESTIMATING $p_0$

As examples of estimation problems two different procedures are shown below where suitable choices of the sequences of observation times $\{t_j : j \geq 1\}$ and of permutations $\{t_{\pi(j)} : j \geq 1\}$ imply almost sure convergence for SLLN

$$\frac{1}{n} \sum_{j=1}^{n} Y(t_{\pi(j)})$$

10 to different estimations.

4.1 PROBLEM 1

Let us suppose to choose a sequence of observation times $\{t_j : j \geq 1\}$ which is dense into $[0, T]$; then by Corollary (1) the class $\mathcal{M}$ contain the uniform probability measure $P_U$ over $B[0, T]$ which is characterized by the respective density function $f_U(t) = \frac{1}{T} \forall t \in [0, T]$ and, applying the proof of theorem (7), a permutation $\{t_{\pi(j)} : j \geq 1\}$ is computed such that the P.E.M.'s $P_{\pi n}$ are vaguely convergent to $P_U$. Now, for a fixed interval $(a, b) \subset [0, T]$ and for any assigned natural $n$, the following set is introduced:

$$A(\pi, n, (a, b)) = \{t_{\pi(j)} \in (a, b) : j = 1, \ldots, n\}$$

whose meaning is evident: among the points $\{t_{\pi(j)} : j = 1, \ldots, n\}$ only the $t_{\pi(j)}$'s falling inside $(a, b)$ are collected. If $n(a, b)$ is the total number of points $t_{\pi(j)}$'s belonging to $A(\pi, n, (a, b))$ and the relative frequency is introduced

$$\frac{1}{n(a, b)} \sum_{t_{\pi(j)} \in A(\pi, n, (a, b))} Y(t_{\pi(j)})$$

10 the a.s. convergence for (10) when $n \to \infty$ and then necessarily $n(a, b) \to \infty$, is stated by below theorem

**Theorem 4** The sequence of r.v.'s (10), when $n \to \infty$ is a strongly consistent estimate of $p_0(L)$ for some points $L \in [a, b]$.

Proof of Theorem By Corollary (1) to main Theorem (7) a permutation $\{t_{\pi(j)} : j \geq 1\}$ can be found such that the P.E.M.'s $P_{\pi n}$ are vaguely convergent to the uniform measure $P_U$ (with density function $f_U(t) = \frac{1}{T} \forall t \in [0, T]$), where
by Theorem (6), the convergences hold true

\[ P_{\pi n}(a, b) = \frac{n(a, b)}{n} \quad \text{and} \quad \lim_{n \to \infty} \frac{n(a, b)}{n} = P_U(a, b) = \frac{b - a}{T}, \]

and this because each \((a, b)\) is a \(P_U\)-continuity set. Now for a fixed \((a, b)\) let us denote by \(P_{(\pi n, a, b)}\) the probability measure giving mass \(\frac{1}{n(a, b)}\) to each point \(t_{\pi(j)} \in A(\pi, n, (a, b))\) in such a way that

\[ P_{(\pi n, a, b)}(c, d) = \frac{n(c, d)}{n(a, b)} \forall (c, d) \subset (a, b], \]

where \(n(c, d)\) is defined analogously to \(n(a, b)\). Let us observe that, because of the equality

\[ \frac{n(c, d)}{n(a, b)} \frac{n(a, b)}{n} = P_{\pi n}(c, d) \frac{1}{P_{\pi n}(a, b)}, \]

and the vague convergence \(P_{\pi n} \to P_U\), we have

\[ \lim_{n \to \infty} \frac{n(c, d)}{n(a, b)} = \lim_{n \to \infty} P_{\pi n}(c, d) \frac{1}{P_{\pi n}(a, b)} = \frac{d - c}{T} \frac{T}{b - a} = \frac{d - c}{b - a}, \]

i.e.

\[ \lim_{n \to \infty} P_{(\pi n, a, b)}(c, d) = \frac{d - c}{b - a}, \]

and the sequence of P.E.M.’s \(P_{(\pi n, a, b)}\) is vaguely convergent to uniform measure \(P_U(a, b)\) having density function \(U_{(a, b)}(t) = \frac{1}{b - a} \forall t \in (a, b]\). Denoting with \(p_0(P_{(\pi n, a, b)})\) and \(p_0(P_{U(a, b)})\) the induced measures by \(p_0\), i.e.

\[ p_0(P_{(\pi n, a, b)})(B) = P_{(\pi n, a, b)}(p_0^{-1}(B)) \quad \text{and} \quad p_0(P_{U(a, b)})(B) = P_{U(a, b)}(p_0^{-1}(B)), \]

\(\forall B \in B[0, 1]\); because of continuity of \(p_0\), the vague convergence of \(P_{(\pi n, a, b)}\) to \(P_{U(a, b)}\) implies the vague convergence of \(p_0(P_{(\pi n, a, b)})\) to \(p_0(P_{U(a, b)})\) and then, by Theorem (6), the convergences hold true

\[ \lim_{n \to \infty} \frac{1}{n(a, b)} \sum_{t_{\pi(j)} \in A(\pi, n, (a, b))} p_0(t_{\pi(j)}) = \int_0^1 pdp_0(P_{U(a, b)}) \]

and

\[ \frac{1}{n(a, b)} \sum_{t_{\pi(j)} \in A(\pi, n, (a, b))} Y(t_{\pi(j)}) \to \int_0^1 pdp_0(P_{U(a, b)}) a.s.; \]

finally, by standard analysis arguments,

\[ \int_0^1 pdp_0(P_{U(a, b)}) = \int_a^b p_0(t) dp_{U(a, b)} = \int_a^b p_0(t) \frac{1}{b - a} dt = \]
\[
\frac{1}{b-a} p_0(t)(b-a) = p_0(t)
\]
where \( \xi \) is a point whose existence is stated by the mean value Theorem for integral of the continuous \( p_0 \) function and proof is now complete.

4.2 PROBLEM 2

Our interest is now concerning a strongly consistent estimate of \( p_0(t) \) where \( t \in [0, T] \) is assigned. The elementary solution given by \( \frac{1}{n} \sum_{i=1}^{n} Y_i(t) \) and based on the observations \( Y_1(t), ..., Y_n(t) \) of the r.v. \( Y(t) \) has no meaning in our context; in fact we may suppose that, when \( n \) is big enough, taking \( n \) observations at the same instant \( t \) is not possible and then we necessarily need \( n \) observation instants \( t_1, t_2, ..., t_n \) with the respective r.v.’s \( Y(t_1), Y(t_2), ..., Y(t_n) \) and their expectations \( p_0(t_1), p_0(t_2), ..., p_0(t_n) \), and this because our urn model has a time dependent composition.

Our aim consists in proving the following result:

**Theorem 5** If \( \{t_j : j \geq 1\} \) is any convergent sequence to \( t \), then \( \frac{1}{n} \sum_{j=1}^{n} Y(t_j) \) is a strongly consistent estimate of \( p_0(t) \).

Proof of Theorem. A first elementary proof is given proving that the convergence \( p_0(t_j) \rightarrow p_0(t) \) implies convergence \( \frac{1}{n} \sum_{j=1}^{n} p_0(t_j) \rightarrow p_0(t) \). In fact, because of convergence \( p_0(t_j) \rightarrow p_0(t) \), for fixed \( \varepsilon > 0 \) there exists \( k \) such that

\[
|p_0(t_j) - p_0(t)| < \frac{\varepsilon}{2} \forall j > k
\]

and

\[
\frac{1}{n-k} \sum_{j=k+1}^{n} p_0(t_j) \rightarrow p_0(t)
\]

Finally the limits \( \frac{1}{n} \sum_{j=1}^{k} p_0(t_j) \rightarrow 0 \) and \( \frac{n-k}{n} \rightarrow 1 \), when \( n \rightarrow \infty \), allows us to state the existence of \( n_0 \) such that

\[
\left| \frac{1}{n} \sum_{j=1}^{n} p_0(t_j) - p_0(t) \right| < \varepsilon \forall n > n_0;
\]

proving that \( \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^{n} p_0(t_j) = p_0(t) \), which jointly with the almost sure convergence \( \frac{1}{n} \sum_{j=1}^{n} (Y(t_j) - p_0(t_j)) \rightarrow 0 \) (apply Rajchman Theorem) completes the proof.
The same result may be proved also via vague convergence of P.E.M.’s $P_n^0$ which assigns weight $\frac{1}{n}$ to each point $\{p_0(t_j) : j = 1, ..., n\}$ ∀n fixed. If $(a, b]$ is an internal interval having $t$ as an internal point, then there exists $k$ such that $p_0(t_j) \in (a, b]$ ∀ $j > k$ and $P_n^0(a, b] \to 1$, while if $t$ is internal to the complement of $(a, b]$ we have that $P_n^0(a, b] \to 0$, proving that $P_n^0$ is vaguely convergent to $P = \delta_t$ which assigns weight 1 to point $t$. Applying Theorem (6) the result is proved.

Applying again the above technique a consistent estimation is found for the difference $[p_0(t) - p_0(t^-)]$ where $p_0(t^-) = \lim_{s \to t^-} p_0(s)$ if the function $p_0$ is right continuous with left limits. In fact if $\{t_j : j \geq 1\}$ and $\{s_j : j \geq 1\}$ are two sequences satisfying

$$\lim_{j \to \infty} t_j = t^+ \text{ and } \lim_{j \to \infty} s_j = t^-$$

we have

$$\lim_{j \to \infty} p_0(t_j) = p_0(t) \text{ and } \lim_{j \to \infty} p_0(s_j) = p_0(t^-),$$

thus applying the above Theorem (5) we obtain

$$\frac{1}{n} \sum_{j=1}^{n} Y(t_j) \to p_0(t) \text{ and } \frac{1}{n} \sum_{j=1}^{n} Y(s_j) \to p_0(t^-) \text{ a.s.}$$

and then

$$\frac{1}{n} \sum_{j=1}^{n} Y(t_j) - \frac{1}{n} \sum_{j=1}^{n} Y(s_j) \to [p_0(t) - p_0(t^-)] \text{ a.s..}$$

5 A RIEMANN-DINI TYPE THEOREM FOR SLLN

The well known Riemann-Dini theorem for real numbers series is extended to strong laws of large numbers for real random variables. Namely if $\sum_{j=1}^{\infty} x_j$ is a simply but not an absolutely convergent series of real numbers and $\alpha \in \mathbb{R} \cup \{\infty, -\infty\}$ is an assigned value, then there exists a permutation (i.e. a bijection $\pi: \mathbb{N} \to \mathbb{N}$) such that $\sum_{j=1}^{\infty} x_{\pi(j)} = \alpha$. Analogously, given a sequence of real random variables $\{Y_j : j \geq 1\}$ having arbitrarily different and finite expectations $\{E(Y_j) : j \geq 1\}$, it is shown, under suitable assumptions, that for any fixed real number $\beta$ belonging to a wide class $B \subset \mathbb{R}$, there exists a permutation $\pi: \mathbb{N} \to \mathbb{N}$ such that the sequence $\frac{1}{n} \sum_{j=1}^{n} Y_{\pi(j)}$ is almost surely
convergent to $\beta$ when $n \to \infty$. The main technical tool is the study of convergence for the sequences of measures $P_n$ which assigns probability mass $\frac{1}{n}$ to each value $\{E(Y_j) : j = 1, \ldots, n\}$ and of the deep interplay between the possible limits of sequences $\{P_n : n \geq 1\}$ and the permutations of values $\{E(Y_{\pi(j)}) : j \geq 1\}$ where $\pi: N \to N$ is an assigned bijection.

5.1 PRELIMINARY ELEMENTS

As an introductory argument a simple but meaningful example may help in showing the goal of our analysis.

**EXAMPLE 1**

Let us suppose that there exists a partition for the sequence of real r.v.’s $\{Y_j : j \geq 1\}$ into two subsequences denoted by $\{Y_{l_k} : k \geq 1\}$ and $\{Y_{n_k} : k \geq 1\}$ satisfying

$$\{Y_j : j \geq 1\} = \{Y_{l_k} : k \geq 1\} \cup \{Y_{n_k} : k \geq 1\}$$ (11)

where $E(Y_{l_k}) = L_1$, $E(Y_{n_k}) = L_2$, $\forall k \geq 1$.

For each fixed natural $n$ let $C_n(L_1)$ and $C_n(L_2)$ denote respectively the total number of r.v.’s $Y_j$ with $1 \leq j \leq n$ which satisfy $E(Y_j) = L_1$ or $E(Y_j) = L_2$ in such a way that $n = C_n(L_1) + C_n(L_2)$ and then

$$\sum_{j=1}^{n} Y_j = \sum_{k=1}^{C_n(L_1)} Y_{l_k} + \sum_{K=1}^{C_n(L_2)} Y_{n_k}.$$ 

Consequently we obtain

$$\frac{1}{n} \sum_{j=1}^{n} Y_j = \frac{C_n(L_1)}{n} \sum_{k=1}^{C_n(L_1)} \frac{Y_{l_k}}{C_n(L_1)} + \frac{C_n(L_2)}{n} \sum_{K=1}^{C_n(L_2)} \frac{Y_{n_k}}{C_n(L_2)}$$ (12)

where

$$0 \leq \frac{C_n(L_1)}{n} \leq 1, 0 \leq \frac{C_n(L_2)}{n} \leq 1 \text{ and } \frac{C_n(L_1)}{n} + \frac{C_n(L_2)}{n} = 1.$$ (13)

Because of (12), the convergence for $\frac{1}{n} \sum_{j=1}^{n} Y_j$ can be shown if the following two steps procedure holds true:

1) applying the standard SLLN the convergences are stated

$$\frac{\sum_{k=1}^{C_n(L_1)} Y_{l_k}}{C_n(L_1)} \to L_1 \text{ and } \frac{\sum_{k=1}^{C_n(L_2)} Y_{n_k}}{C_n(L_2)} \to L_2$$
almost surely when \( n \to \infty \);
b) if
\[
\lim_{n \to \infty} \frac{C_n(L_1)}{n} = p_1, \quad \lim_{n \to \infty} \frac{C_n(L_2)}{n} = p_2,
\]
(14)
because of (13) \( p_1 + p_2 = 1 \) and then the pair \( (p_1, p_2) \) defines a probability distribution over the real values \( L_1, L_2 \). Then, under a) and b) above, we have
\[
\frac{1}{n} \sum_{j=1}^{n} y_j \to p_1 L_1 + p_2 L_2 \text{ almost surely.} \tag{15}
\]
Now this simple case allows us to detect the main elements of our analysis:
i) a class of limit values \( p_1 L_1 + p_2 L_2 \) can be introduced for fixed \( L_1 \) and \( L_2 \) when the pair \( (p_1, p_2) \) is arbitrarily chosen under conditions \( 0 \leq p_i \leq 1 \) for \( i = 1, 2 \) and \( p_1 + p_2 = 1 \) in such a way that for fixed \( L_1 \) and \( L_2 \) the set
\[
B(L_1, L_2) = \{ p_1 L_1 + p_2 L_2 : 0 \leq p_i \leq 1 (i = 1, 2), p_1 + p_2 = 1 \}
\]
defines all possible values which can be the almost sure limit for a sequence
\[
\frac{1}{n} \sum_{j=1}^{n} Y_{\pi(j)} \text{ where } \pi : N \to N \text{ is a permutation of } Y_j' \text{s.}
\]
ii) the existence is evident of a strict connection between any fixed value \( p_1 L_1 + p_2 L_2 \in B(L_1, L_2) \) and a permutation \( \pi \) such that
\[
\frac{1}{n} \sum_{j=1}^{n} Y_{\pi(j)} \to p_1 L_1 + p_2 L_2.
\]
iii) the almost sure limit \( p_1 L_1 + p_2 L_2 \) can be written as an integral
\[
\int_R I(v) d(p_1 \delta_{L_1} + p_2 \delta_{L_2})
\]
(17)
where \( I(.) \) is the identity map and \( p_1 \delta_{L_1} + p_2 \delta_{L_2} \) is the probability measure giving mass \( p_1 \) to \( L_1 \) and \( p_2 \) to \( L_2 \) respectively. This measure is defined through the strict interplay of two components:
c1) the values \( L_1 \) and \( L_2 \) which are assigned by the expectations \( E(Y_j)'s \);
c2) the probability distribution denoted with \( p_1 \) and \( p_2 \) which is the result of limits (14) and choosing a permutation of \( Y_j' \)s. Such a probability measure plays a central role in our approach: for any fixed pair \( (p_1, p_2) \) with \( 0 \leq p_i \leq 1 \) and \( p_1 + p_2 = 1 \) there exists some permutations \( \pi \) such that
\[
\frac{1}{n} \sum_{j=1}^{n} Y_{\pi(j)} \to \int_R I(v) d(p_1 \delta_{L_1} + p_2 \delta_{L_2}) \text{ a.s.;}
\]
thus the limit for the SLLN is assigned by measure \( p_1 \delta_{L_1} + p_2 \delta_{L_2} \). ♦
Our aim consists in extending the above example 1 to more general situations: for instance if the set \( \{ E(Y_j) : j \geq 1 \} \) contains arbitrarily different
values, including the case when \( E(Y_j) \neq E(Y_k) \ \forall j \neq k \). The main result deals with a sequence of r.v.’s \( \{Y_j : j \geq 1\} \) under the following assumptions:

**A 3** the \( Y_j \)’s are uniformly bounded i.e. there exists \( M > 0 \) such that \( |Y_j| \leq M \ \forall j \geq 1 \);

**A 4** the \( Y_j \)’s are pairwise uncorrelated;

**A 5** the \( Y_j \)’s have probability distributions and finite expectations which are arbitrarily different;

**A 6** the sequence of expectations \( \{E(Y_j) : j \geq 1\} \subset [-M, M] \) has at least two different limit points, i.e. there exist at least two different values \( x_1, x_2 \in [-M, M] \) which are the limits of some subsequences of \( \{E(Y_j) : j \geq 1\} \).

It will be shown below the existence of a wide class \( \mathcal{M} \) of probability measures \( P \) over the Borel \( \sigma \)-field \( B(-M, M) \) such that for any assigned \( P \in \mathcal{M} \) there exist some permutations \( \pi : \mathbb{N} \rightarrow \mathbb{N} \) satisfying

\[
\frac{1}{n} \sum_{j=1}^{n} Y_{\pi(j)} \rightarrow \int_{-M}^{M} I(v) dP(v) \text{ a.s..} \tag{18}
\]

The representation of limits given in (18) by integrals of type \( \int_{-M}^{M} I(v) dP(v) \) gives big evidence to measure \( P \); not only: the convergence stated by (18) and the approach here adopted are mainly based on measures defined over the interval \((-M, M)\). Namely: \( P \) is a probability measure which is the limit in some sense of the sequence of the P.E.M. \( P_{\pi n} \)’s which assigns weight \( \frac{1}{n} \) to each point \( \{E(Y_{\pi(j)}) : j = 1, \ldots, n\} \); moreover, if the permutation \( \pi \) is adopted, the set of mean values \( \{E(Y_{\pi(j)}) : j = 1, \ldots, n\} \), the P.E.M.’s \( P_{\pi n} \), and the possible limit \( P \) depend on \( \pi \). The detailed and rigorous definition of the class \( \mathcal{M} \) needs several technical elements which will be an argument of the below subsections.

A further intuitive argument may help in understanding the meaning of our aim; if the r.v.’s \( \{Y_j : j \geq 1\} \) satisfy above assumptions and have arbitrarily different expectations \( \{E(Y_j) : j \geq 1\} \) a SLLN can be easily given taking the differences \( \{(Y_j - E(Y_j)) : j \geq 1\} \) and then applying a well known result: see, for instance, theorem 5.1.2 at page 108 of Chung book [3]. In fact the \( Y_j \)'s are uncorrelated and with uniformly bounded second moments, then

\[
\frac{1}{n} \sum_{j=1}^{n}(Y_j - E(Y_j)) = \frac{1}{n} \sum_{j=1}^{n} Y_j - \frac{1}{n} \sum_{j=1}^{n} E(Y_j) \rightarrow 0 \text{ a.s..} \tag{19}
\]
Of course this is not a solution to our problem : the (19) in fact states the convergence to 0 for the differences and a convergence result for $\frac{1}{n} \sum_{j=1}^{n} E(Y_j)$ is not so easy to obtain. A law of large numbers cannot be applied to the deterministic sequence \( \{ E(Y_j) : j \geq 1 \} \) and also the convergence for the series $\sum_{j=1}^{\infty} E(Y_j)$, in order to apply the Kronecker lemma, is not an easy one if $\{ E(Y_j) : j \geq 1 \}$ is a general sequence in the interval $(-M, M]$. On the other hand the convergence for $\frac{1}{n} \sum_{j=1}^{n} E(Y_j)$ to some value $L$ jointly with the SLLN (19) implies that $\frac{1}{n} \sum_{j=1}^{n} Y_j \to L$ a.s. solving our problem. That of finding hypotheses under which the sequence $\frac{1}{n} \sum_{j=1}^{n} E(Y_j)$ is a convergent one is then a relevant tool in this context. Let us write $\frac{1}{n} \sum_{j=1}^{n} E(Y_j)$ as an integral, i.e.

$$\frac{1}{n} \sum_{j=1}^{n} E(Y_j) = \int_{-M}^{M} I(v) dP_n(v) = \int_{R} I_M(v) dP_n(v)$$

(20)

where $P_n$ is the P.E.M. giving mass $\frac{1}{n}$ to each point $\{ E(Y_j) : j = 1, ..., n \}$, $I(v)$ is the identity map and

$$I_M(v) = I(v) \text{ if } v \in (-M, M)$$

$$I_M(v) = M \text{ if } v \in [M, \infty)$$

$$I_M(v) = -M \text{ if } v \in (-\infty, -M].$$

Because of continuity and boundedness of $I_M$ over $R^1$ a favourable context for convergence of the integrals sequence

$$\{ \int_{R^1} I_M(v) dP_n(v) : n \geq 1 \}$$

is given by VAGUE CONVERGENCE for the sequence $\{ P_n \}$ of probability measures. Applying Theorem 4.4.2 at page 93 of Chung book [3] we have that if $P_n, P$ are probability measures, then $\{ P_n \}$ is vaguely convergent to $P$ if and only if

$$\int_{R^1} f(v) dP_n(v) \to \int_{R^1} f(v) dP(v)$$

for each continuous and bounded $f$.

Thus the vague convergence of $P_n$’s to $P$ implies convergence for integrals

$$\int_{R^1} I_M(v) dP_n(v) \to \int_{R^1} I_M(v) dP(v),$$

thus

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} E(Y_j) = \int_{-M}^{M} I(v) dP(v) \text{ and}$$

16
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} Y_j = \int_{-M}^{M} I(v) dP(v) \text{ a.s.}
\]

The vague convergence of P.E.M. \( P_n \)'s is the general setting adopted for our analysis: the centrality of its role, now evident for convergence of \( \frac{1}{n} \sum_{j=1}^{n} E(Y_j) \), will be shown below also for directly proving the convergence of \( \frac{1}{n} \sum_{j=1}^{n} Y_j \).

### 5.2 THE TECHNICAL BACKGROUND

For a fixed natural \( m \) let us denote by \( \mathcal{H}_m = \{ H_r : r = 1, ..., m \} \) a partition of the interval \((-M, M)\) into \( m \) subintervals where

\[
H_1 = (-M, t_1], H_2 = (t_1, t_2], ..., H_m = (t_{m-1}, M];
\]

the sequence of r.v.'s \( \{ Y_j : j \geq 1 \} \) is supposed to satisfy Assumption (3),..., Assumption (6) and a permutation \( \pi \), which is assigned for \( Y_j \)'s, is omitted in the notations in order to simplify formulas. A partition for \( \{ Y_j : j \geq 1 \} \) into a family of \( m \) subsequences is introduced on the base of the \( m \) sets \( \{ H_r : r = 1, ..., m \} \): for each fixed \( H_r \) we collect the \( Y_j \)'s having the respective \( E(Y_j) \in H_r \), i.e. the subsequence is introduced

\[
\{ Y_{jr_k} : k = 1, 2, ..., Q(H_r) \}
\]

where:

i) \( Q \) is the counting measure which assigns to each \( B \in B(-M, M) \) the respective value \( Q(B) \) i.e. the total number of values \( E(Y_j) \in B \). Thus the set of values taken by \( Q \) includes any natural \( n \) and also \(+\infty\).

ii) The index \( jr \) is a strictly increasing map \( jr : N \to N \) and any value \( jr_k = jr(k) \) means that \( Y_j \) with \( j = jr_k \) is the \( k \)-th element inside \( \{ Y_j : j \geq 1 \} \) such that \( E(Y_j) \in H_r \). Thus each of the \( m \) subsequences \( \{ Y_{jr_k} : k = 1, ..., Q(r) \} \) with \( r = 1, ..., m \) is characterized through the respective index, i.e. the strictly increasing map \( jr : N \to N \), satisfies the following properties:

I) the \( m \) sets of values \( \{ jr_k = jr(k) : k = 1, 2, ..., Q(r) \} \) for \( r = 1, ..., m \) are pairwise disjoint;

II) their union is equal to \( N \).

Then the \( m \) subsequences \( \{ Y_{jr_k} : k = 1, ..., Q(r) \} \) for \( r = 1, ..., m \) are a partition of \( \{ Y_j : j \geq 1 \} \). Now, for each fixed natural \( n \) and given \( \{ Y_j : j = 1, ..., n \} \) and \( \{ E(Y_j) : j = 1, ..., n \} \), let us define the quantities \( \{ c_n(r) : r = 1, ..., m \} \) as

\[
C_n(r) = \sum_{j=1}^{n} I_{H_r}(E(Y_j))
\]

where \( I_{H_r}(E(Y_j)) = 1 \) if \( E(Y_j) \in H_r \) and \( I_{H_r}(E(Y_j)) = 0 \) if \( E(Y_j) \notin H_r \); \( C_n(r) \) is then the total number of values in the set \( \{ E(Y_j) : j = 1, ..., n \} \)
falling inside the interval $H_r$. The following quantity is a generalization of (12) concerning EXAMPLE 1

$$\frac{1}{n} \sum_{j=1}^{n} Y_j = \sum_{r=1}^{m} \frac{C_n(r)}{n} \frac{\sum_{k=1}^{C_n(r)} Y_{jr_k}}{C_n(r)}.$$  \hspace{1cm} (24)

A technical tool for below proofs consists in studying the limit for the second member of (24) when $n \to \infty$. A two step procedure is pointed out dealing, for a fixed $r$, with the two sequences $\frac{C_n(r)}{n}$ and $\sum_{k=1}^{C_n(r)} Y_{jr_k}$. Of course the interesting case is when $H_r$ contains infinitely many $E(Y_j)$’s and then $\frac{C_n(r)}{n}$ may be convergent to a non zero limit.

**STEP 1** The convergence is assumed

$$P(H_r) = \lim_{n \to \infty} \frac{C_n(r)}{n} \forall r = 1, ..., m,$$

where $P$ is an assigned probability measure over the Borel $\sigma$-field $B(-M, M]$.

**STEP 2** If $H_r$ includes infinitely many values $E(Y_j)$’s, then the SLLN can be applied to the sequence

$$\frac{1}{n} \sum_{k=1}^{C_n(r)} (Y_{jr_k} - E(Y_{jr_k})) \to 0 \ a.s.$$  \hspace{1cm} (25)

Because of Assumptions (3) and (11) the SLLN (see Theorem 5.1.2 of Chung book [3]) is applied to the first term in second member of (24)

$$\frac{1}{n} \sum_{k=1}^{C_n(r)} Y_{jr_k} \to 0 \ a.s.$$  \hspace{1cm} (26)

The inclusion $E(Y_{jr_k}) \in H_r = (t_{r-1}, t_r]$ means $t_{r-1} < E(Y_{jr_k}) \leq t_r$ and then the below inequality\n
$$t_{r-1} < \frac{\sum_{k=1}^{C_n(r)} E(Y_{jr_k})}{C_n(r)} \leq t_r$$  \hspace{1cm} (27)

states that the oscillations of the sequence $\frac{1}{C_n(r)} \sum_{k=1}^{C_n(r)} E(Y_{jr_k})$ can be made arbitrarily small if the length of $H_r$ is small and the above steps imply that

$$\left| \frac{C_n(r)}{n} \sum_{k=1}^{C_n(r)} Y_{jr_k} - P(H_r) \frac{\sum_{k=1}^{C_n(r)} E(Y_{jr_k})}{C_n(r)} \right| \to 0 \ a.s.$$  \hspace{1cm} (28)

and

$$\left| \sum_{r=1}^{m} \frac{C_n(r)}{n} \frac{\sum_{k=1}^{C_n(r)} Y_{jr_k}}{C_n(r)} - \sum_{r=1}^{m} P(H_r) \frac{\sum_{k=1}^{C_n(r)} E(Y_{jr_k})}{C_n(r)} \right| \to 0 \ a.s.$$  \hspace{1cm} (29)

We are now ready for the below statement:
Lemma 1 If the sequence of r.v.’s \( \{Y_j : j \geq 1\} \) satisfies Assumptions (2), (4) and if, for \( \epsilon \) fixed, there exists a partition of \((-M, M]\) into subsets \( \{H_r : r = 1, \ldots, m\} \) such that:

i) the length of each \( H_r \) is not greater than \( \epsilon \);

ii) \( \lim_{n \to \infty} \frac{C_n(r)}{C_n} = P(H_r) \forall r = 1, \ldots, m \) where \( P \) is an assigned probability measure over \( B(-M, M] \), then there exists a set \( A \) with probability one such that for each \( \omega \in A \) the existence is proved of a natural value \( n_0(\epsilon, \omega) \) satisfying

\[
\left| \frac{1}{n} \sum_{j=1}^{n} Y_j(\omega) - \int_{-M}^{M} I(v) dP(v) \right| < 2\epsilon, \quad \forall n > n_0(\epsilon, \omega).
\]

PROOF OF LEMMA (1). The sequence of r.v.’s \( \{Y_{jrk} : k \geq 1\} \) satisfies Assumptions (3) and (4) and then, applying Theorem 5.1.2 of Chung book [3], the existence is proved for a set \( A_r \subset \Omega \) with \( \mu(A_r) = 1 \), where \( \mu \) is the probability measure defined over \( \Omega \), such that

\[
\frac{1}{C_n(r)} \sum_{k=1}^{C_n(r)} (Y_{jrk} - E(Y_{jrk})) \to 0
\]

over the set \( A_r \). Of course the above arguments are concerning a set \( H_r \) including infinitely many values \( E(Y_j) \)'s in such a way that \( \frac{C_n(r)}{C_n} \to \infty \) when \( n \to \infty \). Now using (25), the convergence (27) can be directly proved. Through iterations of above procedure for each \( r = 1, \ldots, m \) the existence is given of sets \( \{A_r : r = 1, \ldots, m\} \) with \( \mu(A_r) = 1 \) \( \forall r = 1, \ldots, m \) and then through the intersection \( A = \bigcap_{r=1}^{m} A_r \) we have that \( \mu(A) = 1 \) and (28) holds true. If the value

\[
\sum_{r=1}^{m} P(H_r) \frac{\sum_{k=1}^{C_n(r)} E(Y_{jrk})}{C_n(r)}
\]

is thought as the integral of a simple function taking a constant value over each interval \( H_r \), than its distance from \( \int_{-M}^{M} I(v) dP(v) \) can be estimated using standard arguments:

\[
\left| \sum_{r=1}^{m} P(H_r) \frac{\sum_{k=1}^{C_n(r)} E(Y_{jrk})}{C_n(r)} - \int_{-M}^{M} I(v) dP(v) \right| \leq \\
\leq \sum_{r=1}^{m} \int_{H_r} \left| \frac{\sum_{k=1}^{C_n(r)} E(Y_{jrk})}{C_n(r)} - I(v) \right| dP(v) \leq \\
\leq \sum_{r=1}^{m} \epsilon P(H_r) = \epsilon \sum_{r=1}^{m} P(H_r) = \epsilon
\]

and this recalling that \( \frac{\sum_{k=1}^{C_n(r)} E(Y_{jrk})}{C_n(r)} \in H_r \forall r = 1, \ldots, m \) and if the length of each \( H_r \) is at most \( \epsilon \). The result follows from (28) and the last inequalities.
5.3 THE MEASURES $P_n, P, Q$

Two types of measures introduced above have a central role:
1) the counting measure $Q$ (see (22)), whose values $Q(B)$ assigns the total number of $E(Y_j)$’s falling into $B$, $\forall B \in B(−M, M]$; $Q$ is based on the position of $E(Y_j)$’s inside $(-M, M]$ and may take any natural value and $+\infty$ too.

2) keeping on account of $Q$, and a fixed permutation $\pi$ for $Y_j$’s and $E(Y_j)$’s, the quantities $C_n(r)$ were introduced (see (23) and (24)) for each $H_r$ with $r = 1, \ldots, m$; $C_n(r)$ is the total number of values $E(Y_j)$’s belonging to $H_r$. If a different permutation $\pi'$ is chosen for $Y_j$’s, different values $C_n'(r)$ will be generated. Thus if the limit exists $\lim_{n \to \infty} C_n'(r) = P'(H_r) \ \forall r = 1, \ldots, m$, where $P'$ is a probability measure over $B(−M, M]$, the $P'$ depends on measure $Q$ and permutation $\pi'$. A more general way to define the quantities $C_n(r)$ is that of introducing the probability measure $P_n$ which assigns the mass $\frac{1}{n}$ to each value $\{E(Y_j) : j = 1, \ldots, n\}$ for $n$ fixed and then

$$P_n = \sum_{j=1}^{n} \frac{1}{n} \delta_{E(Y_j)}$$

(29)

defines a probability measure over $B(−M, M]$ which is referred as ”pseudoempiric measure” (P.E.M.) where the ”pseudo” means that $\{E(Y_j) : j \geq 1\}$ is a deterministic and not an i.i.d. sequence of observations. The above limits, if they exist, may be rewritten as

$$\lim_{n \to \infty} P_n(H_r) = P(H_r) \ \forall r = 1, \ldots, m$$

(30)

and the close interplay between permutation $\pi$ and measure $P$ is one of the interesting aspects which characterize the context with arbitrarily different expectations $E(Y_j)$’s, where the P.E.M. $P_n$’s and the possible limit measure $P$ are strictly dependent on $\pi$. If $E(Y_j) = v_0 \ \forall j \geq 1$, i.e. if we consider the classical case, then we have $P_n = P = \delta_{v_0}$, and this for any assigned permutation $\pi$ showing that the classical case is invariant with respect to permutations.

5.4 THE CONVERGENCE OF $P_n$’S TO $P$

This subsection deals mainly with the type of convergence to adopt for the sequence of P.E.M. $P_n$’s to $P$. Each $P_n$ and $P$ are defined over the Borel $\sigma$-field $B(−M, M]$ and then it may appear as a natural request to ask that the convergence $\lim_{n \to \infty} P_n(B) = P(B)$ holds true for each $B \in B(−M, M]$. The following example shows that convergence $P_n(B) \to P(B) \ \forall B \in B(−M, M]$
is a too restrictive request for our purposes.  

**EXAMPLE 2** Let us suppose that \( \{E(Y_j) : j \geq 1\} \) is a strictly decreasing sequence inside \((-M, M]\) such that \( L = \lim_{j \to \infty} E(Y_j) \) and then a sequence of intervals 

\[
A_j = (a_j, b_j] \subset (-M, M] \ \forall j \geq 1
\]

can be constructed in such a way that

i) \( A_j \) contains only one \( E(Y_j) \) as an internal point;

ii) \( A_j \cap A_l = \emptyset \ \forall j \neq l \). A permutation \( \pi \) is assigned and the corresponding sequence \( \{Y_{\pi(j)} : j \geq 1\} \) is considered; for each \( n \) fixed let \( P_n \) be the P.E.M. which assigns probability mass \( \frac{1}{n} \) to each point \( \{E(Y_{\pi(j)}) : j = 1, ..., n\} \) and then \( P_n(A_{\pi(j)}) = \frac{1}{n} \) if \( j = 1, ..., n \) and \( P_n(A_{\pi(j)}) = 0 \) if \( j > n \). Because of the equalities

\[
\sum_{j=1}^{\infty} P_n(A_{\pi(j)}) = 1 \ \forall n \text{ fixed and } \lim_{n \to \infty} P_n(A_{\pi(j)}) = 0 \ \forall j \text{ fixed,}
\]

the Steinhhaus Lemma (see Ash book [1] at page 44) ensures the existence of a subsequence \( \{A_{\pi(j)_k} : k \geq 1\} \) such that \( \{P_n(\cup_{k=1}^{\infty} A_{\pi(j)_k}) : n \geq 1\} \) is not a convergent sequence, proving that the convergence \( P_n(b) \to P(B) \) does not hold true over all sets of \( B(-M, M] \), and this for any assigned permutation \( \pi. \)\)

Now the above example 2 suggests to adopt a type of convergence \( P_n \to P \) which is based on a suitable subclass of \( B(-M, M] \); then the VAGUE CONVERGENCE of \( P_n \) to \( P \) is considered as a driving element for main results given below. The general definition (see Chung book [3] at page 85) is given when \( P_n, P \) are subprobability measures; nevertheless, in this context, we are dealing only with probability measures and then we prefer to consider this case. Moreover, as \( P_n(-M, M] = P(-M, M] = 1 \) we may suppose, without loss of generality, to handle probability measures \( P \) satisfying \( P(-M) = P(M) = 0 \) and \( Q(-M) = Q(M) = 0 \), where \( Q \) is the counting measure. The above elements suggest us to use a condition for vague convergence of probability measures which is equivalent to the general one over \( R^1 \) but using only the interval \((-M, M]\). Some preliminary notions are needed to introduce the definition of vague convergence given below. In [21] we denoted as \( \mathcal{H}_m = \{H_r : r = 1, ..., m\} \) a partition of \((-M, M]\) into \( m \) subintervals \( H_1 = (-M, t_1], H_2 = (t_1, t_2], ..., H_m = (t_{m-1}, M] \).

Here a sequence of partitions for \((-M, M]\) is introduced as it follows; \( \mathcal{H}_1 = (-M, M] \) contains \((-M, M] \) as its unique element. Then choosing arbitrarily a point \( t_3 \) satisfying \(-M < t_3 < M \) the partition \( \mathcal{H}_2 \) is obtained consisting of two intervals \( \mathcal{H}_2 = \{(-M, t_3], (t_3, M]\} \) and choosing \( t_4 \) such that \(-M < t_4 < t_3 \) the partition \( \mathcal{H}_3 \) consists of three intervals \( \mathcal{H}_3 = \{(-M, t_4], (t_4, t_3], (t_3, M]\}. If
$t_5$ is chosen with $t_3 < t_5 < M$ we have $\mathcal{H}_4 = \{-M, t_4], (t_4, t_3], (t_3, t_5], (t_5, M]\}$ and so on..... generating a sequence of partitions $\{\mathcal{H}_m : m \geq 1\}$.

**Definition 3** A sequence of partitions $\{\mathcal{H}_m : m \geq 1\}$ generated by above procedure is defined to be a "progressive sequence of partitions" (P.S.P. hereafter) if $\lim_{m \rightarrow \infty} l_m = 0$ where $l_m$ is the maximum length of the $m$ intervals included into $\mathcal{H}_m$.

**Definition 4** An interval $(a, b]$ is defined to be a continuity interval for the probability measure $P$ defined over the Borel $\sigma$-field $B(R^1)$ if $P(a) = P(b) = 0$.

**Definition 5** If $P_n, P$ are probability measures satisfying $P(-M) = P(M) = P_n(-M) = P_n(M) = 0$ and $P(-M, M] = P_n(-M, M] = 1$, the sequence $\{P_n\}$ is defined to be vaguely convergent to $P$ if there exists a P.S.P. $\{\mathcal{H}_m : m \geq 1\}$ such that each interval $H \in \bigcup_m \mathcal{H}_m$ is a continuity interval for $P$ and $\lim_{n \rightarrow \infty} P_n(H) = P(H)$.

### 5.5 THE MAIN RESULTS

Let us suppose that the sequence of P.E.M. $P_n$'s satisfying Definition (5), is vaguely convergent to $P$. Thus the convergence holds true if $\lim_{n \rightarrow \infty} P_n(H) = P(H)$ for each $H$ inside a P.S.P. $\bigcup_m \mathcal{H}_m$ and consequently if $Q(H) = k \in \mathbb{N}$ ($k < +\infty$), then

$$Q(H) = k \Rightarrow P(H) = 0$$

or equivalently

$$P(H) > 0 \Rightarrow Q(H) = +\infty$$

where $Q(H)$ is the counting measure which assigns the total number of values $E(Y_j) \in H$. Condition (31) seems to be very close to absolute continuity of $P$ with respect to $Q$; nevertheless the absolute continuity is defined over the Borel $\sigma$-field $B(-M, M]$ while (31) involves only intervals inside $\bigcup_m \mathcal{H}_m$ where $\{\mathcal{H}_m : m \geq 1\}$ is a P.S.P.. In our context conditions (31) or (32) are more general than absolute continuity $P << Q$; the evidence is reached via some simple examples, and this could be the case when $\{E(Y_j) : j \geq 1\}$ is a convergent sequence to $L$ and $E(Y_j) \neq L \forall j \geq 1$. If $L$ is an interior point of $(a, b]$ then $\lim_{n \rightarrow \infty} P_n(a, b] = 1$ and $\lim_{n \rightarrow \infty} P_n(a, b] = 0$ if $L$ is interior to the complement of $(a, b]$. Denoting as $P = \delta_L$ the probability measure giving mass 1 to $L$, a P.S.P. $\{\mathcal{H}_m : m \geq 1\}$ for $(-M, M]$ is easy to obtain such that each $H \in \bigcup_m \mathcal{H}_m$ is a $P$-continuity set and $\lim_{n \rightarrow \infty} P_n(H) = P(H)$. Now $P(H) > 0$ means $P(H) = 1$ and this implies $Q(H) = +\infty$ showing that (32) holds true. Nevertheless, being $E(Y_j) \neq L \forall j$ we have $Q(L) = 0$ and $P(L) = 1$.
showing that $P$ is not absolutely continuous with respect to $Q$ (over the Borel $\sigma$-field) and this even if (31) and (32) hold true over a P.S.P. $\{H_m : m \geq 1\}$ of $P$-continuity sets. Condition (31) or (32) has a central role in main results described by the following two statements.

Our interest is concerning an assigned sequence of r.v.’s $\{Y_j : j \geq 1\}$ with finite expectations $\{E(Y_j) : j \geq 1\}$ satisfying Assumptions (3)-(6); the P.E.M. $P_n$ gives mass $\frac{1}{n}$ to each of $n$ values $\{E(Y_j) : j = 1, ..., n\}$ and $Q$ is the counting measure defined above.

**Theorem 6** If the sequence of P.E.M. $P_n$’s is vaguely convergent to a probability measure $P$, then the convergence is satisfied

$$\frac{1}{n} \sum_{j=1}^{n} Y_j \to \int_{-M}^{M} I(v) dP(v) \text{ a.s.}$$

Of course $P$ satisfies (31) and (32) with respect to $Q$ because of vague convergence of $P_n$’s to $P$. Such a relationship shows its importance also in the main statement which is, in some sense, the converse of above Theorem (6): given a probability measure $P$ over $B(-M, M]$ does exist a condition which ensures the existence of a permutation $\pi : N \rightarrow N$ such that $\frac{1}{n} \sum_{j=1}^{n} Y_{\pi(j)} \to \int_{-M}^{M} I(v) dP(v)$ a.s.? The answer is (32): using such a condition the class $\mathcal{M}$ is introduced.

**Definition 6** Given the sequence of r.v.’s $\{Y_j : j \geq 1\}$ with finite expectations $\{E(Y_j) : j \geq 1\}$ satisfying Assumptions (3)-(6), let $\mathcal{M}$ denote the class of probability measures $P$ over $B(-M, M]$ having a P.S.P. $\{H_m : m \geq 1\}$ of $P$-continuity sets such that $P(H) > 0 \Rightarrow Q(H) = +\infty \forall H \in \cup_m H_m$.

**Theorem 7** For each assigned probability measure $P \in \mathcal{M}$ a permutation $\pi : N \rightarrow N$ can be computed such that the sequence of P.E.M. $P_{\pi n}$’s (which for each $n$ fixed assigns mass $\frac{1}{n}$ to each value $\{E(Y_{\pi(j)}) : j = 1, ..., n\}$) is vaguely convergent to $P$ and then (by Theorem (6))

$$\frac{1}{n} \sum_{j=1}^{n} Y_{\pi(j)} \to \int_{-M}^{M} I(v) dP(v) \text{ a.s.}$$

### 5.6 PROOF OF MAIN RESULTS

**PROOF OF THEOREM (6)** Applying definitions (3)-(6), there exists a P.S.P. $\{H_m : m \geq 1\}$ of $P$-continuity sets such that

i) $\lim_{n \to \infty} P_n(H) = P(H) \forall H \in \cup_m H_m$;

ii) $\lim_{m \to \infty} \epsilon_m = 0$ where $\epsilon_m$ is the maximum length of the set of intervals.
\{H_r : r = 1, \ldots, m\} = \mathcal{H}_m.

Then applying Lemma (1) to each fixed partition \(\mathcal{H}_m\), the existence is shown for a set \(A_m\) such that

a) \(\mu(A_m) = 1\) where \(\mu\) is the probability measure defined over \(\Omega\).

b) for each \(\omega \in A_m\), there exists an integer \(n_0(\epsilon_m, \omega)\) such that

\[
\left| \frac{1}{n} \sum_{j=1}^{n} Y_j(\omega) - \int_{-M}^{M} I(v) dP(v) \right| < 2\epsilon_m \ \forall n > n_0(\epsilon_m, \omega).
\]

Thus the \(\mu(\cap_{m=1}^{\infty} A_m) = 1\) and each \(\omega \in \cap A_m\) satisfying statement b) above for each \(m \geq 1\), shows the convergence \(\frac{1}{n} \sum_{j=1}^{n} Y_j(\omega) \to \int_{-M}^{M} I(v) dP(v)\) a.s.

PROOF OF THEOREM (7)
The starting point is a probability measure \(P\) over \(B([-M, M])\) which admits a P.S.P. \(\{H_m : m \geq 1\}\) of P-continuity sets \(H\)'s such that \(P(H) > 0 \Rightarrow Q(H) = \infty \ \forall H \in \cup H_m\). The below proof consists of several steps.

1) THE STRUCTURE OF PARTITIONS

Recalling the construction for the P.S.P. \(\{H_m : m \geq 1\}\), the partition \(H_m\) is a class of \(m\) right closed and left open intervals \(H_{rm} \subset (-M, M]\) indexed by \(rm\) i.e. \(H_m = \{H_{rm} : rm = 1, 2, \ldots, m\}\) and inside \(H_m\) we separate the sets \(H_{rm}\) having positive and null P-measure:

\[H_m^+ = \{H_{rm} : P(H_{rm}) > 0\} = \{H_{sm} : sm = 1, \ldots, m^+\},\]

where \(m^+ \leq m\) and \(H_{sm}\) is a relabeling of P-positive sets, and

\[H_m^0 = \{H_{rm} : P(H_{rm}) = 0\}.
\]

A sequence of partitions \(H_m, H_{m+1}, H_{m+2}, \ldots\) is used which is briefly denoted as \(\{H_{m+i} : i \geq 1\}\) where the notation is adopted

\[H_{m+i} = \{H_{r(m+i)} : r(m+i) = 1, 2, \ldots, m+i\}\]

and (see the construction of partitions in subsection 5.4) \(H_{m+i+1}\) is obtained partitioning only one interval \(H_{r(m+i)} \in H_{m+i}\) into two subintervals denoted as

\[H_{r(m+i+1)} \cup H_{r(m+i+1)+1} \in H_{m+i+1}\]

with

\[H_{r(m+i+1)} \cup H_{r(m+i+1)+1} = H_{r(m+i)}\]

and including into \(H_{m+i+1}\) all the remaining intervals \(H_{r(m+i)} \in H_{m+i}\) with \(r(m+i) \neq r(m+i+1)\). Our goal of finding a permutation may be performed
assigning to each fixed \( n \geq 1 \) a corresponding value \( E(Y_n) \in \{E(Y_j) : j \geq 1\} \) such that the convergence holds true

\[
\lim_{n \to \infty} P_n(H_{rm}) = \lim_{n \to \infty} \frac{C_n(H_{rm})}{n} = P(H_{rm}) \forall H_{rm} \in \cup_m \mathcal{H}_m. \quad (33)
\]

The idea of considering the difference

\[
\left| \frac{C_n(H_{rm})}{n} - P(H_{rm}) \right| \quad (34)
\]

is an intuitive one and the assigned value \( E(Y_n) \) corresponding to \( n \) will be found selecting a set \( H_{rm_0} \in \mathcal{H}_m \) and choosing a value \( E(Y_{jo}) \in H_{rm_0} \); thus we put \( E(Y_n) = E(Y_{jo}) \). Of course a permutation has to be found such that the convergence (33) holds true \( \forall H_{rm} \in \cup_m \mathcal{H}_m \), then the possibility is needed of selecting sets inside each \( \mathcal{H}_m \) for any fixed \( m \geq 1 \). Moreover the differences (34) are not meaningful if the sets \( H_{rm} \in \mathcal{H}_m \) are taken when \( m > n \); in fact the equality \( C_n(H_{rm}) = 0 \) is trivially satisfied for a large class of \( H_{rm} \in \mathcal{H}_m \). Thus a good policy suggests that the index \( m \) of partitions depends on \( n \), i.e. \( m(n) \) is increasing with \( m < n \). Recalling that a sequence of partitions \( \{\mathcal{H}_{m+i} : i \geq 0\} \) is used, we assume to work with a strictly increasing sequence of naturals

\[
\{n_{m+i} : i \geq 0\} \quad (35)
\]

and with the sequence of ”natural intervals”

\[
[n_{m+i}, n_{m+i+1}) = \{n \in N : n_{m+i} \leq n < n_{m+i+1}\} \quad (36)
\]

in such a way that for each fixed \( n \in [n_{m+i}, n_{m+i+1}) \) the selection is performed for a set \( H_{r_{m+i}j_0} \in \mathcal{H}_{m+i} \) and then we put \( E(Y_n) = E(Y_{jo}) \) where \( E(Y_{jo}) \) is a chosen value of \( H_{r_{m+i}j_0} \). Let us observe that when for each \( n \in [n_{m+h}, n_{m+h+1}) \) we select a set \( H_{r_{m+i}j_0} \in \mathcal{H}_{m+i} \) and at the same time, we still select a set \( H_{r_{m+i}j_0} \in \mathcal{H}_{m+i} \) for any \( i \leq h \); in fact each assigned set \( H_{r_{m+h}} \in \mathcal{H}_{m+h} \) is a subset, i.e. \( H_{r_{m+h}} \subseteq H_{r_{m+i}} \) for some \( H_{r_{m+i}} \in \mathcal{H}_{m+i} \) for any fixed \( i \leq h \).

2) THE P-NULL SETS

Given \( \mathcal{H}_m \) and its subclass \( \mathcal{H}_m^0 \) of P-null sets, the union is taken

\[
B_m^0 = \cup\{H_{rm} \in \mathcal{H}_m^0\}. \quad (37)
\]

Where \( n \in [n_{m+1}, n_{m+2}) \) and \( \mathcal{H}_{m+1} \) is taken, let us describe the set \( B_{m+1}^0 = \cup\{H_{r_{m+1}} \in \mathcal{H}_{m+1}^0\} \). The class \( \mathcal{H}_{m+1} \) contains the partition into two subsets \( H_{r_{m+1}}, H_{r_{m+1}+1} \) of only one set \( H_{rm} \in \mathcal{H}_m \) and all the remaining sets \( H_{rm} \in \mathcal{H}_m \) with \( rm \neq rm \). It is now useful to distinguish some cases:
i) if $P(H_{rm}) = 0$, i.e. $H_{rm} \in \mathcal{H}_m^0$, then $P(H_{r(m+1)}) = P(H_{r(m+1)+1}) = 0$ (because subset of the P-null set $H_{rm}$ and $B_{m+1}^0 = B_m^0$.

ii) if $P(H_{rm}), P(H_{r(m+1)}), P(H_{r(m+1)+1})$ are all positive, then we have too $B_{m+1}^0 = B_m^0$ because $\mathcal{H}_{m+1}^0$ and $\mathcal{H}_m^0$ contain the same sets.

iii) if $P(H_{rm}) > 0$ and $P(H_{r(m+1)}) > 0$, $P(H_{r(m+1)+1}) = 0$ (or vice versa $P(H_{r(m+1)}) = 0, P(H_{r(m+1)+1}) > 0$): the class $\mathcal{H}_{m+1}^0$ contains all sets of $\mathcal{H}_m^0$ and the new set $H_{r(m+1)+1}$. Thus $B_{m+1}^0 \supset B_m^0$ and, in the general case, we may write $B_{m+1}^0 \supseteq B_m^0$.

Of course, under iteration of above arguments, we have that $\{B_{m+1}^i : i \geq 0\}$ is a non decreasing sequence of P-null sets where the strict inclusion $B_{m+i+1}^0 \supset B_{m+i}^0$ holds true if the set $H_{r(m+i)} \in \mathcal{H}_{m+i}$, which is partitioned into two subsets $H_{r(m+i+1)}, H_{r(m+i+1)+1} \in \mathcal{H}_{m+i+1}$, satisfies the same conditions of iii) above, i.e. $P(H_{r(m+i)}) > 0$ and $P(H_{r(m+i)+1}) > 0$, $P(H_{r(m+i)+1}+1) = 0$.

3) THE SELECTION TECHNIQUE

The technique we consider deals with selection of ”next term” $E(Y_{n+1})$ of the permutation, when the first $n$ values $E(Y_1), E(Y_2), ..., E(Y_n)$ are assigned and $n$ satisfies $n_m + h \leq n \leq n_m + h + 1 - 2$, where $h$ is a fixed natural. Our purpose is that of selecting a set $H_{r(m+h)} \in \mathcal{H}_{m+h}$ and then to choose the $(n+1)$-th value of permutation taking $E(Y_{n+1}) = E(Y_j)$, where $E(Y_j) \in H_{r(m+h)}$. The selection technique is based on two different procedures for P-null and P-positive sets. We assume here that any P-null set $H_{rm}$ contains infinitely many values $E(Y_j)$’s; in fact the case of a P-null set $H_{rm}$ with finitely many values $E(Y_j)$’s is a trivial one: the convergence $\lim_{n \to \infty} C_n(H_{rm}) = 0$ holds true under any permutation.

We consider all partitions $\mathcal{H}_{m+i}$ with $0 \leq i \leq h$, starting with $\mathcal{H}_m$, its subclasses $\mathcal{H}_m^+, \mathcal{H}_m^0$ of P-positive and P-null sets respectively and $B_m^0$, the union of all sets in $\mathcal{H}_m^0$. A subset is selected

$$N_{m+h} \subset [n_{m+h}, n_{m+h+1}]$$

(38)

and for each $n \in N_{m+h}$ we put $E(Y_n) = E(Y_j)$ where $E(Y_j)$ is a value belonging to $B_m^0$. The choice of $N_{m+h}$ satisfying some conditions which will be discussed later, gives the index values inside $[n_{m+h}, n_{m+h+1}]$ where to place the elements $E(Y_j) \in B_m^0$. Thus if $(n+1) \in N_{m+h}$ we put $E(Y_{n+1}) = E(Y_j) \in B_m^0$, while if $(n+1) \notin N_{m+h}$ we select a subset $H_{rm+1}^0 \in \mathcal{H}_m^+$ using the below method. For each assigned index $sm_0 = 1, 2, ..., m^+$ let us write

$$a_{sm_0} = \frac{C_n(H_{sm_0}) + 1}{n+1} - P(H_{sm_0}) + \sum_{sm_0, sm_0 \neq sm_0} \frac{C_n(H_{sm})}{n+1} - P(H_{sm})$$

(39)
and define as $a_{sm0}$ the index satisfying

$$a_{sm0} = \min\{a_{sm0} : sm0 = 1, ..., m^+\}. \quad (40)$$

Recalling that our goal consists in choosing a set inside $\mathcal{H}_{m+h}$, if the selected set $H_{sm0} \in \mathcal{H}_m$ is too included into $\mathcal{H}_{m+h}$ we may put $E(Y_{n+1}) = E(Y_j)$ where $E(Y_j)$ is a not previously chosen value of $H_{sm0}$. But if $H_{sm0} \notin \mathcal{H}_{m+h}$, this implies that inside $\mathcal{H}_{m+h}$ there exists a family of sets defining a partition of $H_{sm0}$. A first partition of $H_{sm0}$ into two subsets may be found inside a class $\mathcal{H}_{m+i1}$ denoting two sets denoted by $H_{r(m+i1)}$ and $H_{r(m+i1)+1}$ such that $H_{r(m+i1)} \cup H_{r(m+i1)+1} = H_{sm0}$ and afterwards a partition of $H_{r(m+i1)}$ into two subsets may exist inside a class $\mathcal{H}_{m+i2}$ (where $i1 < i2 \leq h$) including two sets $H_{r(m+i2)}$, $H_{r(m+i2)+1}$ in such a way that $H_{r(m+i1)} = H_{r(m+i2)} \cup H_{r(m+i2)+1}$. For sake of simplification, and without loss of generality, we may suppose that $\mathcal{H}_{m-h}$ contains no further subsets of $H_{sm0}$ than the three subsets $H_{r(m+i1)}$, $H_{r(m+i2)+1}$, $H_{r(m+i2)+1}$ and the selection of one of the three above subsets is performed below when all the three subsets have positive P-measure.

The first partition of $H_{sm0}$ into two subsets is introduced by $\mathcal{H}_{m+i1}$; then, after selection of $H_{sm0}$, one of the two subsets $H_{r(m+i1)}$ or $H_{r(m+i1)+1}$ is chosen using a method which is the analogous of above (39) and (40) when there are only two alternatives. Thus, given the two quantities

$$b_1 = \left| \frac{C_n(H_{r(m+i1)}) + 1}{n + 1} - P(H_{r(m+i1)}) \right| + \left| \frac{C_n(H_{r(m+i1)+1})}{n + 1} - P(H_{r(m+i1)+1}) \right| \quad (41)$$

and

$$b_2 = \left| \frac{C_n(H_{r(m+i1)})}{n + 1} - P(H_{r(m+i1)}) \right| + \left| \frac{C_n(H_{r(m+i1)+1}) + 1}{n + 1} - P(H_{r(m+i1)+1}) \right| \quad (42)$$

let us denote by $\overline{k}_0$ the index satisfying

$$b_{\overline{k}_0} = \min\{b_1, b_2\}. \quad (43)$$

If $b_{\overline{k}_0} = b_2$ then $H_{r(m+i1)+1}$ is selected and we put $E(Y_{n+1}) = E(Y_j)$, where $E(Y_j)$ is a not previously chosen value of $H_{r(m+i1)+1}$ and this because of the inclusion $H_{r(m+i1)+1} \in \mathcal{H}_{m+h}$. Vice versa, if $b_{\overline{k}_0} = b_1$ the selected set is $H_{r(m+i1)}$ which is not included into $\mathcal{H}_{m+h}$; in fact $\mathcal{H}_{m+h}$ contains the two subsets $H_{r(m+i2)}$ and $H_{r(m+i2)+1}$ of $H_{r(m+i1)}$. Then, applying again (41), (42) and (43) to $H_{r(m+i2)}$ and $H_{r(m+i2)+1}$, one of the two sets will be selected; thus we put $E(Y_{n+1}) = E(Y_j) \in \mathcal{H}_{r(m+i2)}$ if $H_{r(m+i2)}$ is selected or $E(Y_{n+1}) = E(Y_j) \in \mathcal{H}_{r(m+i2)+1}$ if $H_{r(m+i2)+1}$ is selected.

27
such that $\forall n \in N_{m+i}$ a value $E(Y_n)$ is selected in such a way that $E(Y_n) = E(Y_j)$ where $E(Y_j)$ is a not previously chosen value belonging to the set $B_{m+i}^0$ which is the union of all P-null sets inside the partition $H_{m+i}$. As a choice criterion for the set $N_{m+i}$ the following elements are introduced.

Let us consider the family of quotients $\frac{C_n(B_{m+i}^0)}{n}$ for each $n \geq n_m$ where $m+i = m+i \forall n \in [n_{m+i}, n_{m+i+1})$ and $C_n(B_{m+i}^0)$ gives the total number of values in the set $\{E(Y_j): j = 1, ..., n\} \cap B_{m+i}^0$.

The selection of the subset $N_{m+i} \subset [n_{m+i}, n_{m+i+1}) \forall i \geq 0$ is performed in such a way that:

$$\lim_{n \to \infty} \frac{C_n(B_{m+i}^0)}{n} = 0$$

and all values $E(Y_j) \in \bigcup_{i=0}^{\infty} B_{m+i}^0$ are selected. (45)

The limit (44) above implies the convergence $\lim_{n \to \infty} \frac{C_n(H)}{n} = 0$ for each P-null set $H \in \bigcup_{i=0}^{\infty} H_{m+i}$; in fact, if $H_{r(m+i)}^0 \in H_{m+i}$ and $P(H_{r(m+i)}^0) = 0$, we have $H_{r(m+i)}^0 \subset B_{m+i}^0 \subset B_{m+i}^0 \forall i \geq h$ and then $\frac{C_n(H_{r(m+i)}^0)}{n} \leq \frac{C_n(B_{m+i}^0)}{n} \forall i \geq h$; and thus, by limit (44), $\lim_{n \to \infty} \frac{C_n(H_{r(m+i)}^0)}{n} = 0$.

We are now ready to choose the next term $E(Y_{n+1})$; if $(n+1) \in N_{m+i}$ we put $E(Y_{n+1}) = E(Y_j)$ where $E(Y_j)$ is a not previously chosen value belonging to $B_{m+i}^0$, while if $(n+1) \notin N_{m+i}$ we select a P-positive set following the above procedure and the proof is now complete. ♦

As an example/application the extension of Theorem (7) is suggested to the case of an arbitrary real bounded and dense sequence $\{t_j: j \geq 1\} \subset [0, T]$ where each $t_j$ is not necessarily the expectation $E(Y_j)$ of an assigned random variable. Thus the basic elements concerning Theorem (7) are shown:

i) each $t_j$ denotes an observation time of the process $\{Y(t), \forall t \in [0, T]\}$ under the assumption that $t_j \neq t_k, \forall j \neq k$;

ii) $Q$ is the counting measure defined over the Borel $\sigma$-field $B[0, T]$ such that $Q(A)$ is the total number of $t_j$’s belonging to $A$, for each fixed $A \in B[0, T]$.

Then $Q((a, b]) = +\infty$ for each $(a, b] \subset [0, T]$ and $Q(A)$ is a natural value if $A$ is a finite union of points $t \in [0, T]$.
iii) The class $\mathcal{M}$ is defined in close connection with $Q$: position and density of $t_j$’s inside $[0, T]$ are elements having a strong impact on $\mathcal{M}$: for instance each absolutely continuous P.M. over $[0, T]$ belongs to $\mathcal{M}$. In fact, if $P$ is a P.M.
over $[0, T]$ with density function $f_P(t)$, each interval $H$ belonging to any P.S.P. 
$\{H_m : m \geq 1\}$ of $(0, T]$ is a P-continuity set and if $P(H) > 0 \Rightarrow Q(H) = +\infty$
because of the density of $t_j$’s. Thus Theorem (7) may be applied to any
absolutely continuous measure $P$ over $[0, T]$.

**Corollary 1** If $\{t_j : j \geq 1\}$ is a dense subset of $[0, T]$, then for each assigned absolutely continuous probability measure $P$ over $[0, T]$ some permutation $\pi$
can be computed such that the sequence of P.E.M.’s $P_{\pi n}$, which assigns weight $\frac{1}{n}$ to each point $\{t_{\pi(j)} : j = 1, ..., n\}$, is vaguely convergent to $P$.

**References**

[1] Ash,R.B.;Doleans-Dade,C.A. *Probability & measure theory*, second edition, Academic Press:London,2000.

[2] Chandra,T.K. *Laws of large numbers*, Narosa Publishing House:New Delhi,2012.

[3] Chung,K.L. *A course in probability theory*, third edition, Academic Press:London,2001.

[4] Fiorin,S. *Laws of large numbers for random variables with arbitrarily different and finite expectations via regression method*, Topics in statistical simulation, Volume 114, Springer Proceedings in Mathematics & Statistics, page 167-181: New York, 2014.

[5] Priestley,M.B. *Spectral analysis and time series*, Volume 1, univariate series, Academic Press:London,1981.

[6] Priestley,M.B. *Non-linear and non-stationary time series analysis*, Academic Press:London,1988.