I. Crimaldi, F. Leisen

Asymptotic results for a generalized Pòlya urn with delay and an application

2007/5

UNIVERSITÀ DELL'INSUBRIA
FACOLTÀ DI ECONOMIA

http://eco.uninsubria.it
In questi quaderni vengono pubblicati i lavori dei docenti della Facoltà di Economia dell’Università dell’Insubria. La pubblicazione di contributi di altri studiosi, che abbiano un rapporto didattico o scientifico stabile con la Facoltà, può essere proposta da un professore della Facoltà, dopo che il contributo sia stato discusso pubblicamente. Il nome del proponente è riportato in nota all'articolo. I punti di vista espressi nei quaderni della Facoltà di Economia riflettono unicamente le opinioni degli autori, e non rispecchiano necessariamente quelli della Facoltà di Economia dell'Università dell'Insubria.

These Working papers collect the work of the Faculty of Economics of the University of Insubria. The publication of work by other Authors can be proposed by a member of the Faculty, provided that the paper has been presented in public. The name of the proposer is reported in a footnote. The views expressed in the Working papers reflect the opinions of the Authors only, and not necessarily the ones of the Economics Faculty of the University of Insubria.
Asymptotic results for a generalized Pólya urn and applications to clinical trials

Irene Crimaldi* - Fabrizio Leisen†

April 26, 2007

Abstract

In this paper a new Pólya urn model is introduced and studied; in particular, a strong law of large numbers and two central limit theorems are proven. This urn generalizes a model studied in Berti et al. (2004), May et al. (2005) and in Crimaldi (2007) and it has natural applications in clinical trials. Indeed, the model include both delayed and missing (or null) responses. Moreover, a connection with the conditional identity in distribution of Berti et al. (2004) is given.

1 Introduction

We consider the following experiment. An urn contains $b \in \mathbb{N}^*$ black and $r \in \mathbb{N}^*$ red balls. Let us suppose given two sequences $(r_i)_{i \geq 0}$ and $(u_i)_{i \geq 0}$ of integers such that

$$r_0 = u_0 = 0 < r_1 \leq u_1 < r_2 \leq u_2 < r_3 \leq u_3 < \ldots$$

At each time $n \geq 1$, a ball is drawn from the urn and then it is put again in the urn. Moreover, at each time $u_i$ the urn is updated in the following way: for each $j$ with $u_{i-1} + 1 \leq j \leq r_i$, we put in the urn other $N_j$ balls of the same color as the ball drawn at time $j$. The numbers $N_j$ are randomly chosen in $\mathbb{N}^*$. The way in which the number $N_j$ is chosen may depend on $j$ but it must be suitably independent of the results of the choices for the preceding numbers and of the preceding drawings (see sec. 2). The special case in which $r_i = u_i = i$ for all $i$ is just the case of the generalized Pólya urn studied in Berti et al. (2004) and in Crimaldi (2007). Moreover, if we take $r_i = u_i = i$ for all $i$ and the random variables $N_j$ identically distributed, then we fall in the case considered in May et al. (2005).

In clinical trials this urn can be used to allocate patients to two different treatments. The black balls represent the first treatment, while the red balls represent the second; at each time $n \geq 1$ a patient is allocated to a treatment by picking a ball and observing its color. The introductions $N_j$ represent, according to the interpretation of May et al. (2005), the responses. At time $u_i$, a part of these responses, precisely those associated to an index $j$ with $u_{i-1} + 1 \leq j \leq r_i$, arrives with delay. The responses associated to an index $j$ with $r_i + 1 \leq j \leq u_i$ are considered null or missing because of various facts: for example, decease of

*Department of Mathematics, University of Bologna, Piazza di Porta San Donato 5, 40126 Bologna, Italy. crimaldi@dm.unibo.it
†Department of Economics, University of Insubria, Via Monte generoso, 71, 21100 Varese, Italy. leisen.fabrizio@unimore.it
the patient for reasons that we can’t connect to the treatments, responses that are missed from the analysis laboratory, or responses that the doctor considers irrelevant for future allocations. Hu-Zhang (2004) and Zhang et al. (2007) have introduced interesting urn models with delayed responses which differs from ours in the structure and in the mechanism of updating so that they can be applied to different situations. On the contrary, the purposes and the type of the given results are similar. In the last section of this paper, the reader can find another experiment that can be formalized by the model we study.

Let us denote by $Y_n$ the indicator function of the event \{black ball at time $n$\}, that is, in the language of clinical trials, the indicator function of the event \{first treatment to patient $n$\}, then the random variable $C_n = \sum_{i=1}^n Y_i$ counts the number of patients assigned to the first treatment in the first $n$ trials. In Section 3, we prove a Strong law of large numbers for $(Y_n)_{n \geq 1}$, i.e.

$$\frac{1}{n} \sum_{i=1}^n Y_i \xrightarrow{\text{a.s.}} V,$$

where $V$ is also the almost sure limit of $V_n = E[Y_{n+1} | \mathcal{F}_n]$ (with $(\mathcal{F}_n)_{n \geq 0}$ the natural filtration associated with the model). In the language of clinical trials, this means

$$\frac{C_n}{n} \xrightarrow{\text{a.s.}} V. \quad (1)$$

Moreover, we prove two central limit theorems: precisely, under suitable conditions, we obtain that

$$\sqrt{n}(E[Y_{n+1} | \mathcal{G}_n] - V) \xrightarrow{\mathcal{D}} \nu_1, \quad (2)$$

and

$$\sqrt{r_i(n)} \left( \frac{C_{r_i(n)}}{r_i(n)} - V_n \right) \xrightarrow{\mathcal{D}} \nu_2, \quad (3)$$

where $\mathcal{D}$ means “convergence in distribution” and $\nu_1$ and $\nu_2$ are suitable “mixtures” of Gaussian distributions that are formally defined in Sections 4 and 5. Actually, we show that stronger convergences hold for the two above sequences: almost sure conditional convergence (in the sense of Crimaldi, 2007) for the first sequence and stable convergence (see, for instance, Jacod-Memin, 1981) for the second one. The proof of (2) is based on a limit theorem for martingales which has been proved in Crimaldi (2007) and it employs the same technique used in that paper; while, in order to prove (3), we apply a classical result regarding the stable convergence. Moreover, for the first central limit theorem, we illustrate also an example; while, for the second one, we give for the particular case $r_i = u_i$ (but not necessarily equal to $i$) for all $i$, a set of conditions, which are less difficult to be verified in practice than the general conditions of the theorem.

Finally we can note that, if we consider the proposed model by a more deeper theoretical point of view, then we can say that, with respect to a suitable filtration $(\mathcal{G}_n)_{n \geq 0}$, the sequence $(Y_n)_{n \geq 1}$ has all the property of conditionally identically distributed (cid, abbreviated) sequences, introduced in Berti et al. (2004), except the adaption to the filtration. The study of non adapted sequences of random variables is very interesting because sometimes the request
of adaptation can be restrictive. Thus it could be a fertile ground for further researches. To the best of our knowledge the only paper on this argument is Jayte (2002), which deals with non adapted martingale.

The literature on urn models is very wide. For instance, in addition to the above cited papers, the reader may look at Hill et al. (1980), Gouet (1993), Dirienzo (2000), Kotz et al. (2000), Janson (2006), Muliere et al. (2006).

2 The model and preliminary results

Let us set

\[ l_i = r_i - u_{i-1} \quad \text{for each } i \geq 1, \]

\[ i(n) = \sup\{i \geq 0 : u_i \leq n\} \quad \text{for each } n \geq 0. \]

Given a sequence \((\mu_i)_{i \geq 1}\) of probability measures on \((\mathbb{N}^*)^\infty\), it is possible to build a probability space \((\Omega, \mathcal{A}, P)\) and, on it, a sequence \((Y_n)_{n \geq 1}\) of random variables with values in \(\{0, 1\}\) and a sequence \((L_i)_{i \geq 1}\) of random vectors of the form

\[ L_i = [N_j : u_{i-1} + 1 \leq j \leq r_i] \]

such that the following conditions are satisfied:

(a) For each \(n \geq 0\), a version of the conditional distribution of \(Y_{n+1}\) given the

---

![Figure 1: An example of sequences \((r_i), (u_i)\)](image)
Throughout this paper we use the convention that prove that \( \bar{\mu} \) is the kernel of the sequence \((L_1, V_n(\omega))_{\omega \in \Omega} \) where \( B(1, p) \) denotes the Bernoulli distribution with parameter \( p \) and \( V_n \) is the random variable defined by

\[
V_n := \left( b + \sum_{i=1}^{\ell(n)} \sum_{j=u_{i-1}+1}^{r_i} Y_j N_j \right) \left( b + r + \sum_{i=1}^{\ell(n)} \sum_{j=u_{i-1}+1}^{r_i} N_j \right)^{-1}.
\]

(b) For each \( i \geq 1 \), the random vector \( L_i = [N_j : u_{i-1} + 1 \leq j \leq r_i] \) has distribution \( \mu_i \) and it is independent of the sub-\( \sigma \)-field

\[
\mathcal{F}_{u_{i-1}} \vee \sigma(Y_{u_{i-1}+1}, \ldots, Y_{u_i}).
\]

With this formalization, for each \( n \geq 1 \), the random variable \( Y_n \) denotes the indicator function of the event \{black ball at time \( n \)\} and the random variable \( V_n \) represents the proportion of black ball in the urn at time \( n \).

By condition (a), we have \( E[Y_{n+1} | \mathcal{F}_n] = V_n \) for each \( n \geq 0 \). Moreover, if we set

\[
\mathcal{H}_n := \sigma(Y_j : 1 \leq j \leq r_{i(n)}) \vee \sigma(L_i : 1 \leq i \leq i(n)) \quad \text{(where } \mathcal{H}_0 := \{\emptyset, \Omega\})
\]

we also have \( E[Y_{n+1} | \mathcal{H}_n] = V_n \). Finally, by this equality and condition (b), if we set

\[
\mathcal{G}_n := \mathcal{H}_n \vee \sigma(L_{i(n)+1}),
\]

we also have \( E[Y_{n+1} | \mathcal{G}_n] = V_n \). Indeed, for each \( n \geq 0 \), only the two following cases are possible:

1) \( i(n+1) = i(n) \) and so \( n+1 < u_{i(n)+1} \);
2) \( i(n+1) = i(n) + 1 \) and so \( n+1 = u_{i(n)+1} \).

In both cases, since \( i(u_{i(n)}) = i(n) \), the sub-\( \sigma \)-field \( \mathcal{H}_n \vee \sigma(Y_{n+1}) \) is contained in the sub-\( \sigma \)-field

\[
\mathcal{F}_{u_{i(n)}} \vee \sigma(Y_{u_{i(n)}+1}, \ldots, Y_{u_{i(n)}+1}).
\]

Thus, by assumption (b), the random variable \( L_{i(n)+1} \) is independent of the sub-\( \sigma \)-field \( \mathcal{H}_n \vee \sigma(Y_{n+1}) \).

**Proposition 2.1.** The sequence \((V_n)_{n \geq 0}\) is a martingale with respect to the filtration \( \mathcal{G} = (\mathcal{G}_n)_{n \geq 0} \) (and the filtration \( \mathcal{H} = (\mathcal{H}_n)_{n \geq 0} \)).

**Proof.** Since \((V_n)_{n} \) is \( \mathcal{H} \)-adapted and \( \mathcal{H}_n \subset \mathcal{G}_n \) for each \( n \), then it suffices to prove that \((V_n)_{n}\) is a \( \mathcal{G} \)-martingale. To this end, we observe as above that, for each \( n \geq 0 \), only the two following cases are possible:

1) \( i(n+1) = i(n) \);
2) \( i(n+1) = i(n) + 1 \).

\(^1\)Throughout this paper we use the convention that \( \sum_{a}^{b} = 0 \) if \( b < a \).
In the first case, we have $V_{n+1} = V_n$ and so $E[V_{n+1} | G_n] = V_n$. In the second case, if we set
\[ S_n := (b + r + \sum_{i=1}^{i(n)} \sum_{j=u_i-1+1}^{r_i} N_j), \]
then we can write
\[ V_{n+1} = S_{n+1}^{-1} (V_n S_n + \sum_{j=u_{i(n)}+1}^{r_{i(n)+1}} Y_j N_j) = S_{n+1}^{-1} (V_n S_n + \sum_{j=u_{i(n)}+1}^{r_{i(n)+1}} Y_j N_j). \]
It follows that
\[ E[V_{n+1} | G_n] = S_{n+1}^{-1} (V_n S_n + \sum_{j=u_{i(n)}+1}^{r_{i(n)+1}} N_j E[Y_j | G_n]). \]
On the other hand, for each $j$ with $u_{i(n)} + 1 \leq j \leq r_{i(n)+1}$, we have $i(j-1) = i(n)$ and so we have
\[ E[Y_j | G_n] = E[Y_{i-1} | G_{i-1}] = V_{i-1} = V_n. \]
Thus, we obtain $E[V_{n+1} | G_n] = V_n$. \hfill \Box

**Remark 2.2.** Since each random variable $Y_n$ takes values in $\{0, 1\}$, the above proposition implies that, for each real function $f$ on $\{0, 1\}$, the sequence of conditional expectations $\{E[f(Y_{n+1}) | G_n]\}_{n \geq 0}$ is a $G$-martingale. However, we cannot conclude that the sequence $(Y_n)_{n \geq 1}$ is $G$-conditionally identically distributed in the sense of Berti et al. (2004) because it is generally not $G$-adapted. On the other hand, the sequence $(Y_n)_{n \geq 1}$ is adapted with respect to the filtration $\mathcal{F}_n = (\mathcal{F}_n)_{n \geq 0}$ but $(V_n)_{n \geq 0}$ can not be an $\mathcal{F}$-martingale. For example, if we consider the particular case in which the random variables $N_j$ are deterministic, we have
\[ E[V_{u_k} | \mathcal{F}_{u_k-1}] = S_{u_k}^{-1} (V_{u_k-1} S_{u_k-1} + \sum_{j=u_{k-1}+1}^{u_k-1} Y_j N_j + \sum_{j=u_k}^{r_k} N_j E[Y_j | \mathcal{F}_{u_k-1}]) = S_{u_k}^{-1} (V_{u_k-1} S_{u_k-1} + \sum_{j=u_{k-1}+1}^{u_k-1} Y_j N_j + \sum_{j=u_k}^{r_k} V_{u_k-1}). \]
which is equal to $V_{u_k-1}$ if and only if $u_{k-1} + 1 = u_k = r_k$, that is $u_k = r_k = k$ for all $k \geq 0$. This is the case of the generalized Pólya urn studied in Berti et al. (2004) and in Crimaldi (2007).

### 3 The strong law of large numbers

The sequence $(V_n)_{n \geq 0}$ is a uniformly bounded martingale and so it converges almost surely and in $L^1$ to a bounded random variable $V$. This random variable $V$ is also the limit of the sequence of the empirical means
\[ M_n = \frac{C_n}{n} = \frac{1}{n} \sum_{j=1}^{n} Y_j. \]
More precisely, we have the following proposition.

**Proposition 3.1.** The sequence $(M_n)_{n \geq 1}$ converges in $L^1$ and almost surely to the random variable $V$.

**Proof.** The sequence $(M_n)_{n}$ is uniformly bounded and so it suffices to prove only the almost sure convergence. To this end, we start with observing that, by definition, we have $V_n = E[Y_{n+1} | \mathcal{F}_n]$ and the sequence
\[ Z_n = \sum_{j=1}^{n} j^{-1} (Y_j - V_{j-1}) = \sum_{j=1}^{n} j^{-1} (Y_j - E[Y_j | \mathcal{F}_{j-1}]). \]
is obviously an \( \mathcal{F} \)-martingale. Moreover, since each random variable \( Y_j \) takes values in \( \{0, 1\} \), we have \( \sup_n E[Z_n^2] < \infty \). Hence, the martingale \( (Z_n)_n \) converges almost surely. Kronecker’s lemma ensures that
\[
\frac{1}{n} \sum_{j=1}^{n} (Y_j - V_{j-1}) \xrightarrow{a.s.} 0.
\]
Now, we recall that, if \( (a_n)_n \) and \( (b_n)_n \) are any real sequences, then \( \frac{1}{n} \sum_{k=1}^{n} a_k b_k \rightarrow ab \) whenever \( a_n \geq 0 \) for each \( n \), \( \frac{1}{n} \sum_{k=1}^{n} a_k \rightarrow a \) and \( b_n \rightarrow b \). Therefore, since \( V_{j-1} \) converges almost surely to \( V \), we obtain
\[
\frac{1}{n} \sum_{j=1}^{n} V_{j-1} \xrightarrow{a.s.} V
\]
and so
\[
M_n = \frac{1}{n} \sum_{j=1}^{n} (Y_j - V_{j-1}) + \frac{1}{n} \sum_{j=1}^{n} V_{j-1} \xrightarrow{a.s.} V.
\]

**Remark 3.2.** Since each random variable \( Y_n \) takes values in \( \{0, 1\} \), the above result implies that, for each real function \( f \) on \( \{0, 1\} \), the sequence
\[
\frac{1}{n} \sum_{j=1}^{n} f(Y_j)
\]
converges in \( L^1 \) and almost surely to the random variable \( V_f = f(0)(1 - V) + f(1)V \). Indeed, we have
\[
\frac{1}{n} \sum_{j=1}^{n} f(Y_j) = \frac{1}{n} \sum_{j=1}^{n} f(0)(1 - Y_j) + \frac{1}{n} \sum_{j=1}^{n} f(1)Y_j
\]
\[
= f(0)(1 - \frac{1}{n} \sum_{j=1}^{n} Y_j) + \frac{f(1)}{n} \sum_{j=1}^{n} Y_j.
\]

## 4 A central limit theorem

We are going to prove the following limit theorem.

**Theorem 4.1.** Let us set
\[
Q_k := \begin{cases} 
0 & \text{if } 0 \leq k < u_1 - 1 \\
(\sum_{i=1}^{r_i} \sum_{h=u_{i-1}+1}^{u_i} N_h)^{-1} \sum_{j=u_{i(k)+1}}^{r_{i(k)+1}} N_j & \text{if } k \geq u_1 - 1
\end{cases}
\]
and
\[
Q_{k,j} := \begin{cases} 
0 & \text{if } 0 \leq k < u_1 - 1 \\
N_j (\sum_{i=1}^{r_i} \sum_{h=u_{i-1}+1}^{u_i} N_h)^{-1} & \text{if } k \geq u_1 - 1
\end{cases}
\]
Moreover, let us set
\[
W_n := \sqrt{n}(V_n - V).
\]
Further, let us denote by \( K_n \) a version of the conditional distribution of \( W_n \) given \( G_n \).

Suppose that the following conditions are satisfied:

(i) \( n \sum_{k \geq n} Q_k^2 \xrightarrow{a.s.} H \), where \( H \) is a positive real variable.

(ii) \( \sum_{k \geq 0} k^2 E[Q_k^4] < \infty \).

(iii) \( n \sum_{k \geq n} \sum_{j=u_{i(k)+1}}^{r_{i(k)+1}} Q_{k,j} \sum_{h=u_{i(k)+1}}^{h=u_{i(k)+1}+1} Q_{k,h} \xrightarrow{a.s.} 0 \).
Then, for almost every $\omega$ in $\Omega$, the sequence $\{K_n(\omega, \cdot)\}_n$ of probability measures converges weakly to the Gaussian distribution

$$\mathcal{N}(0, H(\omega)(V(\omega) - V^2(\omega))).$$

In other words, for each bounded continuous function $f$ on $\mathbb{R}$, the conditional expectation $\mathbb{E}[f(W_n)|G_n]$ converges almost surely to the random variable

$$\omega \mapsto \int f(x) \mathcal{N}(0, H(\omega)(V(\omega) - V^2(\omega)))(dx).$$

More briefly, the statement of the above theorem can be so reformulated: with respect to the conditioning system $G = (G_n)_n$, the sequence $(W_n)_n$ converges to the Gaussian kernel

$$\mathcal{N}(0, H(V - V^2)) = \left(\mathcal{N}(0, H(\omega)(V(\omega) - V^2(\omega)))\right)_{\omega \in \Omega}$$

in the sense of the almost sure conditional convergence (see Crimaldi, 2007, Sec. 2). In particular, it follows that the sequence $(W_n)_n$ converges $\mathcal{A}$-stably to the kernel $\mathcal{N}(0, H(V - V^2))$. It is well known that this fact implies that the sequence $(W_n)_n$ converges in distribution to the probability measure $\nu_1$ on $\mathbb{R}$ defined by

$$\nu_1(B) = \int \mathcal{N}(0, H(\omega)(V(\omega) - V^2(\omega)))(B) \, P(d\omega).$$

**Remark 4.2.** Note that the random variables $Q_k$ have been defined in such a way that $Q_k = 0$ when $i(k) = i(k + 1)$.

**Remark 4.3.** If we are in the case $u_{i-1} + 1 = r_i$ for all $i$, then assumption (iii) is obviously satisfied since the third sum is zero.

**Proof.** It will be sufficient to prove that the $G$-martingale $(V_n)_n$ satisfies conditions (a) and (b) of Proposition 2.2 in Crimaldi (2007) with $U = H(V - V^2)$ (see the appendix). To this end, we recall firstly that we can have only two cases $i(k + 1) = i(k)$ or $i(k + 1) = i(k) + 1$. Then, after some calculations, we get

$$V_k - V_{k+1} = \left(V_k \sum_{j=u_i(k)+1}^{r_i(k+1)} N_j - \sum_{j=u_i(k)+1}^{r_i(k+1)} Y_j N_j \right) (b + r + \sum_{i=1}^{i(k+1)} \sum_{h=u_i-1}^{u_i+1} N_h)^{-1}. \quad (5)$$

Moreover, it is immediate to verify that

$$\sum_{j=u_i(k)+1}^{r_i(k+1)} Q_{k,j} = Q_k \quad (6)$$

(Note that, if $i(k + 1) = i(k)$, then $r_i(k+1) < u_{i(k)} + 1$ and the sums in the above relations are equal to zero. On the contrary, if $i(k + 1) = i(k) + 1$, then $r_i(k+1) = r_i(k+1) \geq u_{i(k)} + 1$.) Thus, from (5) and (6), we have

$$|V_k - V_{k+1}| = \left|\sum_{j=u_i(k)+1}^{r_i(k+1)} (V_k - Y_j) N_j \right| (b + r + \sum_{i=1}^{i(k+1)} \sum_{h=u_i-1}^{u_i+1} N_h)^{-1} \leq \sum_{j=u_i(k)+1}^{r_i(k+1)} |V_k - Y_j| Q_{k,j} \leq \sum_{j=u_i(k)+1}^{r_i(k+1)} Q_{k,j} = Q_k,$$
and so, using assumption (ii), we find

$$\sup_k k^2 |V_k - V_{k+1}|^4 \leq \sum_{k \geq 0} k^2 Q_k^4 \in L^1.$$ 

Furthermore, we have

$$\sum_{j=u_k(k)+1}^{r_j(k+1)} N_j \left( b + r + \sum_{i=1}^{r_j(k+1)} \sum_{h=u_{i-1}(k)+1}^{r_i} N_h \right)^{-1} \sim Q_k \quad \text{for } k \to +\infty,$$

and hence, by (5),

$$\sum_{k \geq n} (V_k - V_{k+1})^2 \sim \sum_{k \geq n} \left( V_k Q_k - \sum_{j=u_k(k)+1}^{r_j(k+1)} Y_j Q_{k,j} \right)^2 \quad \text{for } n \to +\infty.$$ 

Therefore, in order to complete the proof, it suffices to prove, for \( n \to +\infty \), the following convergence:

$$n \sum_{k \geq n} \left( V_k Q_k - \sum_{j=u_k(k)+1}^{r_j(k+1)} Y_j Q_{k,j} \right)^2 \xrightarrow{a.s.} H(V - V^2).$$

Since we have \( Y_j^2 = Y_j \), the above sum can be rewritten as

$$n \sum_{k \geq n} \left[ V_k Q_k^2 + \left( \sum_{j=u_k(k)+1}^{r_j(k+1)} Y_j Q_{k,j} \right)^2 - 2V_k Q_k \sum_{j=u_k(k)+1}^{r_j(k+1)} Y_j Q_{k,j} \right] =$$

$$n \sum_{k \geq n} V_k Q_k^2 + n \sum_{k \geq n} \sum_{j=u_k(k)+1}^{r_j(k+1)} Y_j Q_{k,j}^2 + 2n \sum_{k \geq n} \sum_{j=u_k(k)+1}^{r_j(k+1)} \sum_{h=u_{j-1}(k)+1}^{r_{j-1}(k+1)} Y_h Q_{k,h} - 2n \sum_{k \geq n} V_k Q_k \sum_{j=u_k(k)+1}^{r_j(k+1)} Y_j Q_{k,j}.$$ 

Now, by assumption (i) and the almost sure convergence of \((V_k)_k\) to \(V\), we have

$$n \sum_{k \geq n} V_k Q_k^2 \xrightarrow{a.s.} V^2 H.$$ (7)

In the sequel, we are going to prove the following convergences for \( n \to +\infty \):

(c1) \( n \sum_{k \geq n} \sum_{j=u_k(k)+1}^{r_j(k+1)} Y_j Q_{k,j}^2 \xrightarrow{a.s.} VH; \)

(c2) \( n \sum_{k \geq n} V_k Q_k \sum_{j=u_k(k)+1}^{r_j(k+1)} Y_j Q_{k,j} \xrightarrow{a.s.} V^2 H; \)

(c3) \( n \sum_{k \geq n} \sum_{j=u_k(k)+1}^{r_j(k+1)} Y_j Q_{k,j} \sum_{h=u_{j-1}(k)+1}^{r_{j-1}(k+1)} Y_h Q_{k,h} \xrightarrow{a.s.} 0. \)

Let us start with convergence (c1). By assumptions (i) and (iii), we have

$$n \sum_{k \geq n} \sum_{j=u_k(k)+1}^{r_j(k+1)} Q_{k,j}^2 \xrightarrow{a.s.} \left( \sum_{j=u_k(k)+1}^{r_j(k+1)} Q_{k,j} \right)^2 - 2n \sum_{k \geq n} \sum_{j=u_k(k)+1}^{r_j(k+1)} Q_{k,j} \sum_{h=u_{j-1}(k)+1}^{r_{j-1}(k+1)} Q_{k,h} \xrightarrow{a.s.} H.$$ (8)

Thus, by the almost sure convergence of \((V_j)_j\) to \(V\), we have

$$n \sum_{k \geq n} \sum_{j=u_k(k)+1}^{r_j(k+1)} Y_j Q_{k,j}^2 \xrightarrow{a.s.} VH.$$ (9)

Therefore, it will be enough to prove the following convergence:

$$n \sum_{k \geq n} \sum_{j=u_k(k)+1}^{r_j(k+1)} (Y_j - V_j - 1) Q_{k,j}^2 \xrightarrow{a.s.} 0.$$ (10)
Indeed, from this and (9), we obtain convergence (c1).

In order to prove (10), we consider the process \((Z_n)_{n \in \mathbb{N}}\) defined by

\[
Z_n := \sum_{k=0}^{n-1} k \sum_{j=u(k)+1}^{r_i(k+1)} (Y_j - V_{j-1})Q_{k,j}^2.
\]

The random variable \(Z_n\) is \(G_n\)-measurable and we have

\[
Z_{n+1} = \begin{cases} 
Z_n & \text{if } i(n+1) = i(n) \\
Z_n + n \sum_{j=u_i(n)+1}^{r_i(n)+1} (Y_j - V_{j-1})Q_{n,j}^2 & \text{if } i(n+1) = i(n) + 1
\end{cases}
\]

where

\[
E[(Y_j - V_{j-1})Q_{n,j}^2 | G_n] = E[(Y_j - V_{j-1}) | G_n] Q_{n,j}^2 = 0.
\]

Indeed, for \(u_i(n) + 1 \leq j \leq r_i(n)+1\), we have \(G_{j-1} = G_n\) and so

\[
E[(Y_j - V_{j-1}) | G_n] = E[Y_j | G_{j-1} - V_{j-1} = 0.
\]

We have so proved that \((Z_n)_n\) is a martingale with respect to the filtration \(\mathcal{G} = (G_n)_{n \in \mathbb{N}}\). Moreover, by assumption (ii), we have

\[
E[Z_n^2] = \sum_{k=0}^{n-1} k^2 E \left[ \left( \sum_{j=u_i(k)+1}^{r_i(k+1)} (Y_j - V_{j-1})Q_{k,j}^2 \right)^2 \right]
\]

\[
\leq \sum_{k=0}^{n-1} k^2 E \left[ \left( \sum_{j=u_i(k)+1}^{r_i(k+1)} Q_{k,j}^2 \right)^2 \right]
\]

\[
\leq \sum_{k=0}^{n-1} k^2 E \left[ \left( \sum_{j=u_i(k)+1}^{r_i(k+1)} Q_{k,j} \right)^2 \right] = \sum_{k=0}^{n-1} k^2 E[Q_k^4]
\]

\[
\leq \sum_{k=0}^{n-1} k^2 E[Q_k^4] < \infty.
\]

Hence, the martingale \((Z_n)_n\) is bounded in \(L^2\) and so it converges almost surely; that is, the series

\[
\sum_{k \geq 0} k \sum_{j=u_i(k)+1}^{r_i(k+1)} (Y_j - V_{j-1})Q_{k,j}^2
\]

is almost surely convergent. On the other hand, by a well-known Abel’s result, the convergence of a series \(\sum k a_k\), with \(a_k \in \mathbb{R}\), implies the convergence of the series \(\sum k^{-1} a_k\) and the relation \(n \sum_{k \geq n} k^{-1} a_k \to 0\) for \(n \to +\infty\). Applying this result, we find (10).

From (c1), we obtain (c2). Indeed, we have

\[
n \sum_{k \geq n} V_k Q_k \sum_{j=u_i(k)+1}^{r_i(k+1)} Y_j Q_{k,j}^2 = n \sum_{k \geq n} V_k \sum_{j=u_i(k)+1}^{r_i(k+1)} Y_j Q_{k,j}^2 + n \sum_{k \geq n} V_k \sum_{j=u_i(k)+1}^{r_i(k+1)} Y_j (Q_k - Q_{k,j}) Q_{k,j}.
\]

From (c1) and the almost sure convergence of \((V_k)_k\) to \(V\), we get that

\[
n \sum_{k \geq n} V_k \sum_{j=u_i(k)+1}^{r_i(k+1)} Y_j Q_{k,j}^2 \overset{a.s.}{\longrightarrow} V^2 H.
\]

Moreover, from (6), (8) and (i), we get

\[
n \sum_{k \geq n} V_k \sum_{j=u_i(k)+1}^{r_i(k+1)} Y_j (Q_k - Q_{k,j}) Q_{k,j} \leq
\]

\[
n \sum_{k \geq n} \sum_{j=u_i(k)+1}^{r_i(k+1)} (Q_k - Q_{k,j}) Q_{k,j} =
\]

\[
n \sum_{k \geq n} Q_k^2 - n \sum_{k \geq n} \sum_{j=u_i(k)+1}^{r_i(k+1)} Q_{k,j}^2 \overset{a.s.}{\longrightarrow} 0.
\]
Finally, we observe that, by assumption (iii), we have
\[ n \sum_{k \geq n} \sum_{j=\mu(k)+1}^{r(k+1)} Y_{j} Q_{k,j} \sum_{h=\mu(k)+1}^{j-1} Y_{h} Q_{k,h} \leq \]
\[ n \sum_{k \geq n} \sum_{j=\mu(k)+1}^{r(k+1)} Q_{k,j} \sum_{h=\mu(k)+1}^{j-1} Q_{k,h} \xrightarrow{a.s.} 0. \]
The proof is so concluded. \( \square \)

**Example 4.4.** Let us suppose that \( r_i = 2i-1 \) and \( u_i = 2i \) for each \( i \geq 1 \). Then we have
\[ l_i = 1 \quad \text{and} \quad L_i = N_{2i-1} \quad \text{for each} \quad i \geq 1. \]
Let us assume that the random variables \( N_i \) are identically distributed (that is \( \mu_i = \mu \) for each \( i \geq 1 \)) with \( E[N_i^4] < +\infty \). If we set
\[ m := E[N_i], \quad \delta := E[N_i^2], \quad h := \frac{\delta}{m^2}, \]
then, using the same notation as in the previous theorem, we get that \( W_n \) converges \( G \)-stably in the strong sense to the Gaussian kernel \( N(0, 2h(V - V^2)) \).
In order to prove this fact, we have to verify conditions (i), (ii) and (iii) of the previous theorem. We firstly observe that \( u_{i-1} + 1 = r_i \) for all \( i \geq 1 \) and so assumption (iii) is obviously fulfilled. Moreover, for each \( k \geq 0 \), we have \( i(k) = [k/2] \) where the simbol \( [\cdot] \) denotes the integer part. Therefore, we have
\[ Q_k := \left\{ \begin{array}{ll}
0 & \text{if } k \text{ is even} \\
(\sum_{i=1}^{i(k+1)} N_{2i-1})^{-1} N_{2i(k+1)-1} = (\sum_{i=1}^{i(k)+1} N_{2i-1})^{-1} N_{2i(k)+1} & \text{if } k \text{ is odd}
\end{array} \right. \]
and so
\[ \sum_{k \geq 0} k^2 E[Q_k^2] \leq \sum_{j \geq 1} (2j - 1)^2 E[Q_{2j-1}^2] \leq E[N_i^4] \sum_{j \geq 1} (2j - 1)^2 j^{-4} \leq 4 E[N_i^4] \sum_{j \geq 1} j^{-2} < +\infty. \]
Further, we have for \( n \to +\infty \)
\[ n \sum_{k \geq n} Q_k^2 = n \sum_{j \in \mathbb{N}, j \geq (n+1)/2} Q_{2j-1}^2 \sim 2n \sum_{j \geq n} Q_{2j-1}^2 = 2n \sum_{j \geq n} N_{2j-1}^2 (\sum_{i=1}^{j} N_{2i-1})^{-2}. \]
Since the random variables \( N_i \) are independent, identically distributed and integrable, then, by the strong law of large numbers, we get
\[ \sum_{i=1}^{j} N_{2i-1} \overset{a.s.}{\sim} jm \quad \text{for } j \to +\infty \]
and so we obtain
\[ n \sum_{k \geq n} Q_k^2 \overset{a.s.}{\sim} 2m^{-2}n \sum_{j \geq n} j^{-2} N_{2j-1}^2. \]
Now, for each \( j \geq 1 \), let us set
\[ X_j := \frac{(N_{2j-1}^2 - \delta)}{j}. \]
The random variables \( X_j \) are independent, with mean equal to zero and variance \( \text{Var}[X_j] = j^{-2} \text{Var}[N_i^2] \). Thus, the series \( \sum_{j \geq 1} X_j \) converges almost surely and so we obtain
\[ n \sum_{j \geq n} j^{-1} X_j \xrightarrow{a.s.} 0. \]
This implies that
\[ n \sum_{j \geq n} j^{-2} N_j \overset{a.s.}{\sim} \delta n \sum_{j \geq n} j^{-2} \to \delta \]
and we can conclude that assumption (i) is satisfied with \( H = 2h \).

5 Another central limit theorem

We have the following result.

**Theorem 5.1.** For each \( n \geq u_1 \), let us set
\[
S_n = \sqrt{r_i(n)}(M_{r_i(n)} - V_n)
\]
and
\[
X_{n,j} = \frac{1}{\sqrt{r_i(n)}} \left( \sum_{k=r_i(j-1)+1}^{r_i(j)} Y_k + (u_i(j) - r_i(j))V_{u_i(j)-1} - \min(r_i(n), u_i(j))V_{u_i(j)} - (u_i(j-1) - r_i(j-1))V_{u_i(j-1)-1} + \min(r_i(n), u_i(j-1))V_{u_i(j-1)} \right)
\]
for \( 1 \leq j \leq n \). Suppose:

(a) \( U_n = \sum_{j=1}^{n} X_{n,j}^2 \overset{P}{\rightarrow} U \).

(b) \( X_n^* = \sup_{1 \leq j \leq n} |X_{n,j}| \overset{L^1}{\rightarrow} 0 \).

Then the sequence \( (S_n)_{n \geq 1} \) converges \( A \)-stably to the Gaussian kernel \( \mathcal{N}(0, U) \).

**Remark 5.2.** It is worthwhile to note that, for each \( n \), we have \( X_{n,j} = 0 \) when \( i(j-1) = i(j) \).

**Remark 5.3.** If \( r_j = u_j = j \), the above conditions become the same conditions as in Berti et al. (2004) or in Berti et al. (2005).

**Proof.** We will use Theorem A.1 in appendix. For each \( n \geq u_1 \), let us set
\[
D_n = \sqrt{r_i(n)}(M_{r_i(n)} - V),
\]
and for \( 0 \leq j \leq n \)
\[
L_{n,j} = \mathbb{E}[D_n \mid \mathcal{G}_j] \quad \mathcal{F}_{n,j} = \mathcal{G}_j.
\]
Then, for each \( n \geq u_1 \), the sequence \( (L_{n,j})_{0 \leq j \leq n} \) is a martingale with respect to \( (\mathcal{F}_{n,j})_{0 \leq j \leq n} \) such that \( L_{n,0} = \mathbb{E}[D_n \mid \mathcal{G}_0] = 0 \) and
\[
L_{n,j} - L_{n,j-1} = \mathbb{E}[D_n \mid \mathcal{G}_j] - \mathbb{E}[D_n \mid \mathcal{G}_{j-1}] = X_{n,j} \quad \text{for } 1 \leq j \leq n.
\]
Indeed we have
\[
\begin{align*}
E[D_n \mid G_j] &= E[D_n \mid G_{j-1}] \\
&= \frac{1}{\sqrt{r_{(n)}}} \left( \sum_{k=1}^{r_{(j)}} Y_k + \sum_{k=r_{(j)}+1}^{u_{(j)}} E[Y_k \mid G_j] + \sum_{k=u_{(j)}+1}^{r_{(n)}} E[Y_k \mid G_j] - r_{(n)} V_j \right) \\
&= \sum_{k=r_{(j)}+1}^{r_{(j+1)}} Y_k \left( E[Y_k \mid G_{j+1}] - E[Y_k \mid G_j] \right) - \sum_{k=r_{(j)}+1}^{r_{(n)}} E[Y_k \mid G_j] - E[Y_k \mid G_{j-1}] + \sum_{k=u_{(j)}+1}^{r_{(n)}} E[Y_k \mid G_j] - r_{(n)} V_j \\
&= \frac{1}{\sqrt{r_{(n)}}} \left( \sum_{k=r_{(j)}+1}^{r_{(j+1)}} Y_k + \sum_{k=r_{(j)}+1}^{u_{(j)}} V_{u_{(j)}} + \sum_{k=u_{(j)}+1}^{r_{(n)}} E[Y_k \mid G_j] - r_{(n)} V_j \right) \\
&= \frac{1}{\sqrt{r_{(n)}}} \left( \sum_{k=r_{(j)}+1}^{r_{(j+1)}} Y_k + (u_{(j)} - r_{(j)}) V_{u_{(j)}} - (r_{(n)} - u_{(j)}) V_{u_{(j)} - 1} + (r_{(n)} - u_{(j)}) V_{u_{(j)} - 1} + r_{(n)} V_{u_{(j)} - 1} \right) \\
&= \frac{1}{\sqrt{r_{(n)}}} \left( \sum_{k=r_{(j)}+1}^{r_{(j+1)}} Y_k + (u_{(j)} - r_{(j)}) V_{u_{(j)}} - (r_{(n)} - u_{(j)}) V_{u_{(j)} - 1} + r_{(n)} V_{u_{(j)} - 1} \right) \\
&= \frac{1}{\sqrt{r_{(n)}}} \left( \sum_{k=r_{(j)}+1}^{r_{(j+1)}} Y_k + (u_{(j)} - r_{(j)}) V_{u_{(j)}} - (r_{(n)} - u_{(j)}) V_{u_{(j)} - 1} + r_{(n)} V_{u_{(j)} - 1} \right) \\
&= \frac{1}{\sqrt{r_{(n)}}} \left( \sum_{k=r_{(j)}+1}^{r_{(j+1)}} Y_k + (u_{(j)} - r_{(j)}) V_{u_{(j)}} - (r_{(n)} - u_{(j)}) V_{u_{(j)} - 1} + r_{(n)} V_{u_{(j)} - 1} \right) \\
&= X_{n,j}.
\end{align*}
\]

Moreover, we have
\[
S_n = E[D_n \mid G_n] = L_{n,n} = \sum_{j=1}^{n} X_{n,j}.
\]

Finally, if \( \mathcal{N} \) denotes the sub-\( \sigma \)-field generated by the \( P \)-negligible events, then
\[
V_j = \lim \inf_{n} \mathcal{F}_{n,j \wedge n} = \lim \inf_{n} G_{j \wedge n} = G_j
\]
and
\[
\mathcal{V} = \mathcal{N} \vee \bigvee_{j \geq 0} V_j = \mathcal{N} \vee \bigvee_{j \geq 0} G_j
\]
and so the random variable \( U \) is measurable with respect to the \( \sigma \)-field \( \mathcal{V} \). At this point we can apply Theorem A.1 together with Remark A.2 and the proof of the first assertion is concluded.

If conditions (a1) and (b1) hold, then condition (a) is obviously verified and we have
\[
X_{n,j} = \frac{1}{\sqrt{r_{(n)}}} Z_j
\]
where
\[
Z_j = \sum_{k=r_{(j+1)}}^{r_{(n+1)}} Y_k - u_{(j)} V_{u_{(j)}} + u_{(j-1)} V_{u_{(j-1)}}.
\]
Therefore, since \( i(u_{(n)}) = i(n) \) and \( i(u_{(n)} - 1) = i(n) - 1 \), we can write
\[
\begin{align*}
\frac{1}{r_{(n)}} Z_j^2 &= X^2_{u_{(n)}-u_{(n)} - 1} \sum_{j=1}^{u_{(n)}-1} X^2_{u_{(n)}-j} - \sum_{j=1}^{u_{(n)}-1} X^2_{u_{(n)}-j} \\
&= \sum_{j=1}^{u_{(n)}-1} X^2_{u_{(n)}-j} - \frac{1}{r_{(n)}} \sum_{j=1}^{u_{(n)}-1} Z_j^2 \\
&= U_{u_{(n)}} - \frac{r_{(n)}-1}{r_{(n)}} U_{u_{(n)}-1} \overset{a.s.}{\to} 0,
\end{align*}
\]
This fact implies that
\[
X_n^* = \sup_{1 \leq j \leq n} \left| X_{n,j} \right| \overset{a.s.}{\to} 0,
\]
12
Indeed,
\[
\sup_{0 \leq j \leq n} X_{n,j}^2 = \sup_{0 \leq j \leq n} \frac{1}{r_i(n)} Z_j^2
= \sup_{0 \leq j \leq r_i(n)} \frac{1}{r_i(n)} Z_j^2 \xrightarrow{a.s.} 0.
\]
(Note that the second equality holds because \( Z_j = 0 \) for \( r_i(n) = u_i(n) < j \leq n \) since \( i(j-1) = i(j) \).) Further, we have
\[
E[(X_n^*)^2] = E[\sup_{1 \leq j \leq n} X_{n,j}^2] \leq \sum_{j=1}^{n} E[X_{n,j}^2]
= \sum_{j=1}^{n} E[(L_{n,j} - L_{n,j-1})^2]
= \sum_{j=1}^{n} E[L_{n,j}^2 - L_{n,j-1}^2]
= E[L_{n,n}^2] = E[S_n^2].
\]
From (c1) we obtain that the sequence \((X_n^*)\) is bounded in \(L^2\) and so we get condition (b). \(\square\)

6 Other interpretation

The proposed model can be employed also for the following experiment. At time 0 an urn contains \( b \in \mathbb{N}^* \) black and \( r \in \mathbb{N}^* \) red balls. At each time \( i \geq 1 \), a sample of \( u_i - u_{i-1} \) patients are assigned to a treatment by this procedure: for each patient we pick a ball from the urn, we observe its color and we put it again in the urn. Then \( r_i - u_{i-1} \) “significant” responses arrive, we give for convenience number \( j = u_{i-1} + 1, \ldots, r_i \) to the corresponding patients and the urn is so updated: for each \( j = u_{i-1} + 1, \ldots, r_i \), we add \( N_j \) balls of the color corresponding to the treatment assigned to patient \( j \) (\( Y_j = 1 \) means black ball and first treatment and \( Y_j = 0 \) means red ball and second treatment). In this context, the random variable \( C_u = \sum_{k=1}^{u_i} Y_k \) represents the number of patients allocated to the first treatment until time \( i \).

A Appendix

For the reader’s convenience, we state some results used above. For more details on the stable convergence or on the almost sure conditional convergence, we refer to Jacod-Memin (1981) and Crimaldi (2007), respectively.

**Theorem A.1.** Let \((l_n)_{n \geq 1}\) be a sequence of strictly positive integers. On a probability space \((\Omega, \mathcal{A}, P)\), for each \( n \geq 1 \), let \((\mathcal{F}_{n,j})_{0 \leq j \leq l_n}\) be a filtration and \((L_{n,j})_{n \geq 1, 0 \leq j \leq l_n}\) be a triangular array of real random variables on \((\Omega, \mathcal{A}, P)\) with values such that, for each \( n \), the family \((L_{n,j})_{0 \leq j \leq l_n}\) is a martingale with respect to \((\mathcal{F}_{n,j})_{0 \leq j \leq l_n}\) and \( L_{n,0} = 0 \). For each pair \((n, j)\), with \( n \geq 1, 1 \leq j \leq l_n \), let us set \( X_{n,j} = L_{n,j} - L_{n,j-1} \) and
\[
S_n = \sum_{j=1}^{l_n} X_{n,j} = L_{n,l_n}, \quad U_n = \sum_{j=1}^{l_n} X_{n,j}^2, \quad X_n^* = \sup_{1 \leq j \leq l_n} |X_{n,j}|.
\]
Let us suppose that the sequence \((U_n)_{n \geq 1}\) converges in probability to a positive random variable \(U\). Further, let us suppose \( X_n^* \overset{L^1}{\longrightarrow} 0 \). Finally, let \( \mathcal{N} \) be the
sub-σ-field generated by the \( P \)-negligible events and let us set

\[
\mathcal{V}_j = \lim \inf_n \mathcal{F}_{n,j \wedge n} \quad \text{for } j \geq 0, \quad \mathcal{V} = \mathcal{N} \vee \bigvee_{j \geq 0} \mathcal{V}_j.
\]

If \( U \) is measurable with respect to the σ-field \( \mathcal{V} \), then \((S_n)_{n \geq 1}\) converges \( \mathcal{V} \)-stably to the Gaussian kernel \( \mathcal{N}(0, U) \).

**Remark A.2.** We recall that, if the random variable \( S_n \) is \( \mathcal{V} \)-measurable for each \( n \), then the \( \mathcal{V} \)-stable convergence implies the \( \mathcal{A} \)-stable convergence.

For a proof of this theorem, the reader may look at Th. 5 and Cor. 7 in sec. 7 of Crimaldi et al. (2007). It maybe worthwhile to note that in Crimaldi et al. (2007) there exists a stronger version of the previous result and so also Theorem 5.1 could be enunciated in a stronger way.

**Proposition A.3.** (see Prop. 2.2 in Crimaldi (2007))

On a probability space \((\Omega, \mathcal{A}, \mathbb{P})\), let \((V_n)_{n \in \mathbb{N}}\) be a real martingale with respect to a filtration \( \mathcal{G} = (\mathcal{G}_n)_{n \in \mathbb{N}} \). Suppose that \((V_n)_{n}\) converges in \( L^1 \) to a random variable \( V \). Moreover, setting

\[
U_n := n \sum_{k \geq n} (V_k - V_{k+1})^2, \quad Y := \sup_k \sqrt{k} |V_k - V_{k+1}|,
\]

assume that the following conditions hold:

(a) The random variable \( Y \) is integrable.

(b) The sequence \((U_n)_{n \geq 1}\) converges almost surely to a positive real random variable \( U \).

Then, with respect to \( \mathcal{G} \), the sequence \((W_n)_{n \geq 1}\) defined by

\[
W_n := \sqrt{n}(V_n - V)
\]

converges to the Gaussian kernel \( \mathcal{N}(0, U) \) in the sense of the almost sure conditional convergence.

In particular, the sequence \((W_n)_{n}\) converges \( \mathcal{A} \)-stably to \( \mathcal{N}(0, U) \).

**References**


