Berti, Patrizia; Crimaldi, Irene; Pratelli, Luca; Rigo, Pietro

Working Paper

Central Limit Theorems For Multicolor Urns With Dominated Colors

Quaderni di Dipartimento, No. 106

Provided in Cooperation with:
University of Pavia, Department of Economics and Quantitative Methods (EPMQ)

Suggested Citation: Berti, Patrizia; Crimaldi, Irene; Pratelli, Luca; Rigo, Pietro (2009) : Central Limit Theorems For Multicolor Urns With Dominated Colors, Quaderni di Dipartimento, No. 106, Università degli Studi di Pavia, Dipartimento di Economia Politica e Metodi Quantitativi (EPMQ), Pavia

This Version is available at:
https://hdl.handle.net/10419/95302

Quaderni di Dipartimento

Central Limit Theorems For Multicolor Urns With Dominated Colors

Patrizia Berti
(Università di Modena e Reggio Emilia)

Irene Crimaldi
(Università di Bologna)

Luca Pratelli
(Accademia Navale di Livorno)

Pietro Rigo
(Università di Pavia)

106 (12-09)

Dipartimento di economia politica
e metodi quantitativi
Università degli studi di Pavia
Via San Felice, 5
I-27100 Pavia

Dicembre 2009

CENTRAL LIMIT THEOREMS FOR MULTICOLOR URNS WITH DOMINATED COLORS

PATRIZIA BERTI, IRENE CRIMALDI, LUCA PRATELLI, AND PIETRO RIGO

ABSTRACT. An urn contains balls of $d \geq 2$ colors. At each time $n \geq 1$, a ball is drawn and then replaced together with a random number of balls of the same color. Let $\mathbf{A_n} = \mathrm{diag}\big(A_{n,1}, \ldots, A_{n,d}\big)$ be the n-th reinforce matrix. Assuming $EA_{n,j} = EA_{n,1}$ for all n and j, a few CLT's are available for such urns. In real problems, however, it is more reasonable to assume

$$EA_{n,j} = EA_{n,1} \quad \text{whenever } n \geq 1 \text{ and } 1 \leq j \leq d_0,$$
$$\liminf_n EA_{n,1} > \limsup_n EA_{n,j} \quad \text{whenever } j > d_0,$$

for some integer $1 \leq d_0 \leq d$. Under this condition, the usual weak limit theorems may fail, but it is still possible to prove CLT's for some slightly different random quantities. These random quantities are obtained neglecting dominated colors, i.e., colors from $d_0 + 1$ to d, and allow the same inference on the urn structure. The sequence $(\mathbf{A_n} : n \geq 1)$ is independent but need not be identically distributed. Some statistical applications are given as well.

1. The problem

An urn contains $a_j > 0$ balls of color $j \in \{1, \ldots, d\}$ where $d \geq 2$. At each time $n \geq 1$, a ball is drawn and then replaced together with a random number of balls of the same color. Say that $A_{n,j} \geq 0$ balls of color j are added to the urn in case $X_{n,j} = 1$, where $X_{n,j}$ is the indicator of $\{\text{ball of color } j \text{ at time } n\}$. Let

$$N_{n,j} = a_j + \sum_{k=1}^{n} X_{k,j} A_{k,j}$$

be the number of balls of color j in the urn at time n and

$$Z_{n,j} = \frac{N_{n,j}}{\sum_{i=1}^{d} N_{n,i}}, \quad M_{n,j} = \frac{\sum_{k=1}^{n} X_{k,j}}{n}.$$

Fix j and let $n \to \infty$. Then, under various conditions, $Z_{n,j} \xrightarrow{a.s.} Z_{(j)}$ for some random variable $Z_{(j)}$. This typically implies $M_{n,j} \xrightarrow{a.s.} Z_{(j)}$. A CLT is available as well. Define in fact

$$C_{n,j} = \sqrt{n}\,\big(M_{n,j} - Z_{n,j}\big) \quad \text{and} \quad D_{n,j} = \sqrt{n}\,\big(Z_{n,j} - Z_{(j)}\big).$$

As shown in [4], under reasonable conditions one obtains

$$(C_{n,j},\, D_{n,j}) \longrightarrow \mathcal{N}(0, U_j) \times \mathcal{N}(0, V_j) \quad \text{stably}$$

Date: September 2, 2009.

2000 *Mathematics Subject Classification.* 60F05, 60G57, 60B10.

Key words and phrases. Central limit theorem – Clinical trials – Random probability measure – Stable convergence – Urn model.

for certain random variables U_j and V_j. A nice consequence is

$$\sqrt{n}\,\big(M_{n,j} - Z_{(j)}\big) = C_{n,j} + D_{n,j} \longrightarrow \mathcal{N}(0, U_j + V_j) \quad \text{stably.}$$

Stable convergence, in the sense of Aldous and Renyi, is a strong form of convergence in distribution. The definition is recalled in Section 3.

For $(C_{n,j},\, D_{n,j})$ to converge, it is fundamental that $EA_{n,j} = EA_{n,1}$ for all n and j. In real problems, however, it is more sound to assume that

$$\begin{aligned} EA_{n,j} &= EA_{n,1} \quad \text{whenever } n \geq 1 \text{ and } 1 \leq j \leq d_0, \\ \liminf_n EA_{n,1} &> \limsup_n EA_{n,j} \quad \text{whenever } j > d_0, \end{aligned}$$

for some integer $1 \leq d_0 \leq d$. Roughly speaking, when $d_0 < d$ some colors (those labelled from $d_0 + 1$ to d) are dominated by the others. In this framework, for $j \in \{1, \ldots, d_0\}$, meaningful quantities are

$$C^*_{n,j} = \sqrt{n}\,\big(M^*_{n,j} - Z^*_{n,j}\big) \quad \text{and} \quad D^*_{n,j} = \sqrt{n}\,\big(Z^*_{n,j} - Z_{(j)}\big) \quad \text{where}$$

$$M^*_{n,j} = \frac{\sum_{k=1}^n X_{k,j}}{1 + \sum_{i=1}^{d_0} \sum_{k=1}^n X_{k,i}}\,, \quad Z^*_{n,j} = \frac{N_{n,j}}{\sum_{i=1}^{d_0} N_{n,i}}.$$

If $d_0 = d$, then $D^*_{n,j} = D_{n,j}$ and $|C^*_{n,j} - C_{n,j}| \leq \frac{1}{\sqrt{n}}$. If $d_0 < d$, in a sense, dealing with $(C^*_{n,j},\, D^*_{n,j})$ amounts to neglecting dominated colors.

Our problem is to determine the limiting distribution of $(C^*_{n,j},\, D^*_{n,j})$, under reasonable conditions, when $d_0 < d$.

2. Motivations

Possibly, when $d_0 < d$, $Z_{n,j}$ and $M_{n,j}$ have a more transparent meaning than their counterparts $Z^*_{n,j}$ and $M^*_{n,j}$. Accordingly, a CLT for $(C_{n,j},\, D_{n,j})$ is more intriguing than a CLT for $(C^*_{n,j},\, D^*_{n,j})$. So, why dealing with $(C^*_{n,j},\, D^*_{n,j})$?

The main reason is that $(C_{n,j},\, D_{n,j})$ merely fails to converge in case

$$\liminf_n EA_{n,j} > \frac{1}{2} \liminf_n EA_{n,1} \quad \text{for some } j > d_0. \tag{1}$$

Fix in fact $j \leq d_0$. Under some conditions, $Z_{n,j} \xrightarrow{a.s.} Z_{(j)}$ with $Z_{(j)} > 0$ a.s.; see Lemma 3. Furthermore, condition (1) yields $\sqrt{n} \sum_{i=d_0+1}^d Z_{n,i} \xrightarrow{a.s.} \infty$. (This follows from Corollary 2 of [9] for $d = 2$, but it can be shown in general). Hence,

$$D^*_{n,j} - D_{n,j} \geq Z_{n,j}\, \sqrt{n} \sum_{i=d_0+1}^{d} Z_{n,i} \xrightarrow{a.s.} \infty.$$

Since $D^*_{n,j}$ converges stably, as proved in Theorem 4, $D_{n,j}$ fails to converge in distribution under (1).

A CLT for $D_{n,j}$, thus, is generally not available. A way out could be looking for the right norming factors, that is, investigating whether $\frac{\alpha_n}{\sqrt{n}}\, D_{n,j}$ converges stably for suitable constants α_n. This is a reasonable solution but we discarded it. In fact, as proved in Corollary 5, $(C_{n,j},\, D_{n,j})$ converges stably whenever

$$\limsup_n EA_{n,j} < \frac{1}{2} \liminf_n EA_{n,1} \quad \text{for all } j > d_0. \tag{1*}$$

So, the choice of α_n depends on whether (1) or (1*) holds, and this is typically unknown in applications (think to clinical trials). In addition, dealing with

$(C^*_{n,j},\, D^*_{n,j})$ looks natural (to us). Loosely speaking, as the problem occurs because there are some dominated colors, the trivial solution is just to neglect dominated colors.

A next point to be discussed is the practical utility (if any) of a CLT for $(C^*_{n,j},\, D^*_{n,j})$ or $(C_{n,j},\, D_{n,j})$. To fix ideas, we refer to $(C^*_{n,j},\, D^*_{n,j})$ but the same comments apply to $(C_{n,j},\, D_{n,j})$ provided a CLT for the latter is available. It is convenient to distinguish two situations. With reference to a real problem, suppose the subset of *non dominated* colors is some $J \subset \{1, \ldots, d\}$ and not necessarily $\{1, \ldots, d_0\}$.

If J is known, the main goal is to make inference on $Z_{(j)}$, $j \in J$. To this end, the limiting distribution of $D^*_{n,j}$ is useful. Knowing such distribution, for instance, asymptotic confidence intervals for $Z_{(j)}$ are easily obtained. An example (cf. Example 6) is given in Section 4.

But in various frameworks, J is actually unknown (think to clinical trials again). Then, the main focus is to identify J and the limiting distribution of $C^*_{n,j}$ can help. If such distribution is known, the hypothesis

$$H_0 : J = J^*$$

can be (asymptotically) tested for any $J^* \subset \{1, \ldots, d\}$ with $\mathrm{card}(J^*) \geq 2$. Details are in Examples 7 and 8.

A last remark is that our results become trivial for $d_0 = 1$. On one hand, this is certainly a gap, as $d_0 = 1$ is important in applications. On the other hand, $d_0 = 1$ is itself a trivial case. Indeed, $Z_{(1)} = 1$ a.s., so that no inference on $Z_{(1)}$ is required.

This paper is the natural continuation of [4]. While the latter deals with $d_0 = d$, the present paper focus on $d_0 < d$. Indeed, our results hold for $d_0 \leq d$, but they are contained in Corollary 9 of [4] in the particular case $d_0 = d$. In addition to [4], a few papers which inspired and affected the present one are [1] and [9]. Other related references are [2], [3], [5], [7], [8], [10], [12].

The paper is organized as follows. Section 3 recalls some basic facts on stable convergence. Section 4 includes the main results (Theorem 4 and Corollary 5). Precisely, conditions for

$$(C^*_{n,j},\, D^*_{n,j}) \longrightarrow \mathcal{N}(0, U_j) \times \mathcal{N}(0, V_j) \quad \text{stably and}$$
$$(C_{n,j},\, D_{n,j}) \longrightarrow \mathcal{N}(0, U_j) \times \mathcal{N}(0, V_j) \quad \text{stably under (1*)}$$

are given, U_j and V_j being the same random variables mentioned in Section 1. As a consequence,

$$\sqrt{n}\left(M^*_{n,j} - Z_{(j)}\right) = C^*_{n,j} + D^*_{n,j} \longrightarrow \mathcal{N}(0, U_j + V_j) \quad \text{stably and}$$
$$\sqrt{n}\left(M_{n,j} - Z_{(j)}\right) = C_{n,j} + D_{n,j} \longrightarrow \mathcal{N}(0, U_j + V_j) \quad \text{stably under (1*)}.$$

Also, it is worth noting that $D^*_{n,j}$ and $D_{n,j}$ actually converge in a certain stronger sense.

Finally, our proofs are admittedly long. To make the paper more readable, they have been confined in Section 5 and in a final Appendix.

3. Stable convergence

Let $(\Omega, \mathcal{A}, P)$ be a probability space and S a metric space. A *kernel* on S (or a *random probability measure* on S) is a measurable collection $N = \{N(\omega) : \omega \in \Omega\}$

of probability measures on the Borel σ-field on S. Measurability means that

$$N(\omega)(f) = \int f(x)\, N(\omega)(dx)$$

is $\mathcal{A}$-measurable, as a function of $\omega \in \Omega$, for each bounded Borel map $f : S \to \mathbb{R}$.

Let (Y_n) be a sequence of S-valued random variables and N a kernel on S. Both (Y_n) and N are defined on $(\Omega, \mathcal{A}, P)$. Say that Y_n converges *stably* to N in case

$$P\big(Y_n \in \cdot \mid H\big) \longrightarrow E\big(N(\cdot) \mid H\big) \quad \text{weakly}$$
$$\text{for all } H \in \mathcal{A} \text{ such that } P(H) > 0.$$

Clearly, if $Y_n \to N$ stably, then Y_n converges in distribution to the probability law $E\big(N(\cdot)\big)$ (just let $H = \Omega$). We refer to [5] and references therein for more on stable convergence. Here, we mention a strong form of stable convergence, introduced in [5]. Let $\mathcal{F} = (\mathcal{F}_n)$ be any sequence of sub-σ-fields of $\mathcal{A}$. Say that Y_n converges *$\mathcal{F}$-stably in strong sense* to N in case

$$E\big(f(Y_n) \mid \mathcal{F}_n\big) \xrightarrow{P} N(f) \quad \text{for all bounded continuous functions } f : S \to \mathbb{R}.$$

Finally, we give two lemmas from [4]. In both, $\mathcal{G} = (\mathcal{G}_n)$ is an increasing filtration. Given kernels M and N on S, let $M \times N$ denote the kernel on $S \times S$ defined as

$$\big(M \times N\big)(\omega) = M(\omega) \times N(\omega) \quad \text{for all } \omega \in \Omega.$$

Lemma 1. *Let Y_n and Z_n be S-valued random variables and M and N kernels on S, where S is a separable metric space. Suppose $\sigma(Y_n) \subset \mathcal{G}_n$ and $\sigma(Z_n) \subset \mathcal{G}_\infty$ for all n, where $\mathcal{G}_\infty = \sigma(\cup_n \mathcal{G}_n)$. Then,*

$$(Y_n, Z_n) \longrightarrow M \times N \quad \textit{stably}$$

provided $Y_n \to M$ stably and $Z_n \to N$ $\mathcal{G}$-stably in strong sense.

Lemma 2. *Let (Y_n) be a $\mathcal{G}$-adapted sequence of real random variables. If $\sum_{n=1}^{\infty} \frac{EY_n^2}{n^2} < \infty$ and $E\big(Y_{n+1} \mid \mathcal{G}_n\big) \xrightarrow{a.s.} Y$, for some random variable Y, then*

$$n \sum_{k \geq n} \frac{Y_k}{k^2} \xrightarrow{a.s.} Y \quad \textit{and} \quad \frac{1}{n} \sum_{k=1}^{n} Y_k \xrightarrow{a.s.} Y.$$

4. Main results

In the sequel, $X_{n,j}$ and $A_{n,j}$, $n \geq 1$, $1 \leq j \leq d$, are real random variables on the probability space $(\Omega, \mathcal{A}, P)$ and $\mathcal{G} = (\mathcal{G}_n : n \geq 0)$, where

$$\mathcal{G}_0 = \{\emptyset, \Omega\}, \quad \mathcal{G}_n = \sigma\big(X_{k,j},\, A_{k,j} : 1 \leq k \leq n,\, 1 \leq j \leq d\big).$$

Let $N_{n,j} = a_j + \sum_{k=1}^{n} X_{k,j} A_{k,j}$ where $a_j > 0$ is a constant. We assume that

$$X_{n,j} \in \{0,1\}, \quad \sum_{j=1}^{d} X_{n,j} = 1, \quad 0 \leq A_{n,j} \leq \beta \quad \text{for some constant } \beta, \tag{2}$$

$$\big(A_{n,j} : 1 \leq j \leq d\big) \text{ independent of } \mathcal{G}_{n-1} \vee \sigma\big(X_{n,j} : 1 \leq j \leq d\big),$$

$$Z_{n,j} = P\big(X_{n+1,j} = 1 \mid \mathcal{G}_n\big) = \frac{N_{n,j}}{\sum_{i=1}^{d} N_{n,i}} \quad \text{a.s.}.$$

Given an integer $1 \leq d_0 \leq d$, let us define

$$\lambda_0 = 0 \text{ if } d_0 = d \text{ and } \lambda_0 = \max_{d_0 < j \leq d} \limsup_n EA_{n,j} \text{ if } d_0 < d.$$

We also assume that

$$EA_{n,j} = EA_{n,1} \quad \text{for } n \geq 1 \text{ and } 1 \leq j \leq d_0, \tag{3}$$
$$m := \lim_n EA_{n,1}, \quad m > \lambda_0, \quad q_j := \lim_n EA_{n,j}^2 \quad \text{for } 1 \leq j \leq d_0.$$

A few useful consequences are collected in the following lemma. Define

$$S_n^* = \sum_{i=1}^{d_0} N_{n,i} \quad \text{and} \quad S_n = \sum_{i=1}^{d} N_{n,i}.$$

Lemma 3. *Under conditions* (2)-(3)*, as* $n \to \infty$*,*

$$\frac{S_n^*}{n} \xrightarrow{a.s.} m \quad \textit{and} \quad \frac{S_n}{n} \xrightarrow{a.s.} m,$$
$$n^{1-\lambda} \sum_{i=d_0+1}^{d} Z_{n,i} \xrightarrow{a.s.} 0 \quad \textit{whenever } d_0 < d \textit{ and } \lambda > \frac{\lambda_0}{m},$$
$$Z_{n,j} \xrightarrow{a.s.} Z_{(j)} \quad \textit{for each } 1 \leq j \leq d_0,$$

where each $Z_{(j)}$ *is a random variable such that* $Z_{(j)} > 0$ *a.s..*

For $d = 2$, Lemma 3 follows from results in [9] and [10]. For arbitrary d, it is possibly known but we do not know of any reference. Accordingly, a proof of Lemma 3 is given in the Appendix. We also note that, apart from a few particular cases, the probability distribution of $Z_{(j)}$ is not known (even if $d_0 = d$).

We aim to settle the asymptotic behavior of

$$C_{n,j} = \sqrt{n}\,\big(M_{n,j} - Z_{n,j}\big), \quad D_{n,j} = \sqrt{n}\,\big(Z_{n,j} - Z_{(j)}\big),$$
$$C_{n,j}^* = \sqrt{n}\,\big(M_{n,j}^* - Z_{n,j}^*\big), \quad D_{n,j}^* = \sqrt{n}\,\big(Z_{n,j}^* - Z_{(j)}\big),$$

where $j \in \{1, \ldots, d_0\}$ and

$$M_{n,j} = \frac{\sum_{k=1}^n X_{k,j}}{n}, \quad M_{n,j}^* = \frac{\sum_{k=1}^n X_{k,j}}{1 + \sum_{k=1}^n \sum_{i=1}^{d_0} X_{k,i}}, \quad Z_{n,j}^* = \frac{N_{n,j}}{\sum_{i=1}^{d_0} N_{n,i}}.$$

Let $\mathcal{N}(a, b)$ denote the one-dimensional Gaussian law with mean a and variance $b \geq 0$ (where $\mathcal{N}(a, 0) = \delta_a$). Note that $\mathcal{N}(0, L)$ is a kernel on $\mathbb{R}$ for each real non negative random variable L. We are in a position to state our main result.

Theorem 4. *If conditions* (2)-(3) *hold, then*

$$C_{n,j}^* \longrightarrow \mathcal{N}(0, U_j) \textit{ stably and}$$
$$D_{n,j}^* \longrightarrow \mathcal{N}(0, V_j) \ \mathcal{G}\textit{-stably in strong sense}$$
$$\textit{for each } j \in \{1, \ldots, d_0\}, \textit{ where } U_j = V_j - Z_{(j)}(1 - Z_{(j)})$$
$$\textit{and } V_j = \frac{Z_{(j)}}{m^2} \Big\{ q_j\,(1 - Z_{(j)})^2 + Z_{(j)} \sum_{i \leq d_0, i \neq j} q_i\, Z_{(i)} \Big\}.$$

In particular (by Lemma 1),

$$(C_{n,j}^*,\, D_{n,j}^*) \longrightarrow \mathcal{N}(0, U_j) \times \mathcal{N}(0, V_j) \textit{ stably.}$$

As noted in Section 2, Theorem 4 has been thought for the case $d_0 < d$, and it reduces to Corollary 9 of [4] in the particular case $d_0 = d$. We also remark that some assumptions can be stated in a different form. In particular, under

suitable extra conditions, Theorem 4 works even if $(A_{n,1}, \ldots, A_{n,d})$ independent of $\mathcal{G}_{n-1} \vee \sigma(X_{n,1}, \ldots, X_{n,d})$ is weakened into

$$(A_{n,1}, \ldots, A_{n,d}) \text{ conditionally independent of } (X_{n,1}, \ldots, X_{n,d}) \text{ given } \mathcal{G}_{n-1};$$

see Remark 8 of [4].

The proof of Theorem 4 is deferred to Section 5. Here, we stress a few of its consequences.

We already know (from Section 2) that $(C_{n,j},\, D_{n,j})$ may fail to converge when $d_0 < d$. There is a remarkable exception, however.

Corollary 5. *Under conditions* (2)-(3)*, if* $2\,\lambda_0 < m$ *(that is,* (1*) *holds) then*

$$C_{n,j} \longrightarrow \mathcal{N}(0, U_j) \text{ stably and } D_{n,j} \longrightarrow \mathcal{N}(0, V_j) \ \mathcal{G}\text{-stably in strong sense}$$

for each $j \in \{1, \ldots, d_0\}$*. In particular (by Lemma 1),*

$$(C_{n,j},\, D_{n,j}) \longrightarrow \mathcal{N}(0, U_j) \times \mathcal{N}(0, V_j) \text{ stably.}$$

Proof. By Theorem 4, it is enough to prove $D^*_{n,j} - D_{n,j} \xrightarrow{P} 0$ and $C^*_{n,j} - C_{n,j} \xrightarrow{P} 0$. It can be assumed $d_0 < d$. Note that

$$\Big|D^*_{n,j} - D_{n,j}\Big| = \sqrt{n}\, Z_{n,j}\, \big(\frac{S_n}{S^*_n} - 1\big) \le \frac{S_n}{S^*_n}\, \sqrt{n} \sum_{i=d_0+1}^{d} Z_{n,i},$$

$$C^*_{n,j} - C_{n,j} = D_{n,j} - D^*_{n,j} \,+\, M_{n,j}\, \sqrt{n}\, \frac{\sum_{i=d_0+1}^{d} M_{n,i} - \frac{1}{n}}{\frac{1}{n} + \sum_{i=1}^{d_0} M_{n,i}}.$$

By Lemma 3 and $2\,\lambda_0 < m$, there is $\alpha > \frac{1}{2}$ such that $n^\alpha \sum_{i=d_0+1}^{d} Z_{n,i} \xrightarrow{a.s.} 0$. Thus, it remains only to see that $\sqrt{n}\, M_{n,i} \xrightarrow{a.s.} 0$ for each $i > d_0$. Fix $i > d_0$ and define $L_{n,i} = \sum_{k=1}^{n} \frac{X_{k,i} - Z_{k-1,i}}{\sqrt{k}}$. Since $(L_{n,i} : n \ge 1)$ is a $\mathcal{G}$-martingale and

$$\sum_n E\big\{(L_{n+1,i} - L_{n,i})^2 \mid \mathcal{G}_n\big\} = \sum_n \frac{Z_{n,i}(1 - Z_{n,i})}{n+1} \le \sum_n \frac{n^\alpha Z_{n,i}}{n^{1+\alpha}} < \infty \quad \text{a.s.},$$

then $L_{n,i}$ converges a.s.. By Kronecker lemma,

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_{k,i} - Z_{k-1,i}) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sqrt{k}\, \frac{X_{k,i} - Z_{k-1,i}}{\sqrt{k}} \xrightarrow{a.s.} 0.$$

Since $\frac{1}{\sqrt{n}} \sum_{k=1}^{n} k^{-\alpha} \longrightarrow 0$ and $Z_{k,i} = o(k^{-\alpha})$ a.s., it follows that

$$\sqrt{n}\, M_{n,i} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_{k,i} - Z_{k-1,i}) \,+\, \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} Z_{k,i} \xrightarrow{a.s.} 0.$$

□

Theorem 4 has some statistical implications as well.

Example 6. (A statistical use of $D^*_{n,j}$). Suppose $d_0 > 1$, conditions (2)-(3) hold, and fix $j \le d_0$. Let $(V_{n,j} : n \ge 1)$ be a sequence of consistent estimators of V_j, that is, $V_{n,j} \xrightarrow{P} V_j$ and $\sigma(V_{n,j}) \subset \mathcal{D}_n$ for each n where

$$\mathcal{D}_n = \sigma\big(X_{k,i}A_{k,i},\, X_{k,i} : 1 \le k \le n,\, 1 \le i \le d\big)$$

is the σ-field corresponding to the "available data". Since $(V_{n,j})$ is $\mathcal{G}$-adapted, Theorem 4 yields

$$(D^*_{n,j},\, V_{n,j}) \longrightarrow \mathcal{N}(0, V_j) \times \delta_{V_j} \quad \mathcal{G}\text{-stably in strong sense.}$$

Since $d_0 > 1$, then $0 < Z_{(j)} < 1$ a.s., or equivalently $V_j > 0$ a.s.. Hence,

$$I_{\{V_{n,j}>0\}}\, \frac{D^*_{n,j}}{\sqrt{V_{n,j}}} \longrightarrow \mathcal{N}(0,1) \quad \mathcal{G}\text{-stably in strong sense.}$$

For large n, this fact allows to make inference on $Z_{(j)}$. For instance,

$$Z^*_{n,j} \pm \frac{u_\alpha}{\sqrt{n}}\, \sqrt{V_{n,j}}$$

provides an asymptotic confidence interval for $Z_{(j)}$ with (approximate) level $1-\alpha$, where u_α is such that $\mathcal{N}(0,1)(u_\alpha, \infty) = \frac{\alpha}{2}$.

An obvious consistent estimator of V_j is

$$V_{n,j} = \frac{1}{m_n^2}\, \Big\{\, Q_{n,j}\,(1 - Z_{n,j})^2 \,+\, Z_{n,j}^2 \sum_{i \le d_0, i \neq j} Q_{n,i} \,\Big\} \quad \text{where}$$

$$m_n = \frac{\sum_{k=1}^n \sum_{i=1}^d X_{k,i} A_{k,i}}{n} \;\text{ and }\; Q_{n,i} = \frac{\sum_{k=1}^n X_{k,i} A_{k,i}^2}{n}.$$

In fact, $E(X_{n+1,i} A_{n+1,i}^2 \mid \mathcal{G}_n) = Z_{n,i}\, E A_{n+,i}^2 \xrightarrow{a.s.} Z_{(i)}\, q_i$ for all $i \le d_0$, so that Lemma 2 implies $Q_{n,i} \xrightarrow{a.s.} Z_{(i)}\, q_i$. Similarly, $m_n \xrightarrow{a.s.} m$. Therefore, $V_{n,j} \xrightarrow{a.s.} V_j$.

Finally, Theorem 4 also implies $\sqrt{n}\,\big(M^*_{n,j} - Z_{(j)}\big) = C^*_{n,j} + D^*_{n,j} \longrightarrow \mathcal{N}(0, U_j + V_j)$ stably. So, another asymptotic confidence interval for $Z_{(j)}$ is $M^*_{n,j} \pm \frac{u_\alpha}{\sqrt{n}} \sqrt{G_{n,j}}$ where $G_{n,j}$ is a consistent estimator of $U_j + V_j$. One merit of the latter interval is that it does not depend on the initial composition a_i, $i = 1, \ldots, d_0$ (provided this is true for $G_{n,j}$ as well).

Example 7. (A statistical use of $C^*_{n,j}$). Suppose

$$E A_{n,j} = \mu_j \;\text{ and }\; \mathrm{var}(A_{n,j}) = \sigma_j^2 > 0 \;\text{ for all } n \ge 1 \text{ and } 1 \le j \le d.$$

Suppose also that conditions (2)-(3) hold with some $J \subset \{1, \ldots, d\}$ in the place of $\{1, \ldots, d_0\}$, where $\mathrm{card}(J) > 1$, that is

$$\mu_r = m > \mu_s \quad \text{whenever } r \in J \text{ and } s \notin J.$$

Both J and $\mathrm{card}(J)$ are unknown, and we aim to test the hypothesis $H_0 : J = J^*$ where $J^* \subset \{1, \ldots, d\}$ and $\mathrm{card}(J^*) > 1$. Note that U_j can be written as

$$U_j = \frac{Z_{(j)}}{m^2}\, \big\{(1 - Z_{(j)})^2 \sigma_j^2 + Z_{(j)} \sum_{i \in J, i \neq j} Z_{(i)}\, \sigma_i^2\big\}, \quad j \in J.$$

Fix $j \in J^*$. Under H_0, a consistent estimator of U_j is

$$U_{n,j} = \frac{Z_{n,j}}{\widehat{m}_n^2 \left(\sum_{i \in J^*} Z_{n,i}\right)^4}\, \big\{(1 - Z_{n,j})^2 \widehat{\sigma}_{n,j}^2 + Z_{n,j} \sum_{i \in J^*, i \neq j} Z_{n,i}\, \widehat{\sigma}_{n,i}^2\big\} \quad \text{where}$$

$$\widehat{m}_n = \frac{1}{\mathrm{card}(J^*)} \sum_{i \in J^*} \widehat{m}_{n,i},\; \widehat{m}_{n,i} = \frac{\sum_{k=1}^n X_{k,i} A_{k,i}}{\sum_{k=1}^n X_{k,i}},\; \widehat{\sigma}_{n,i}^2 = \frac{\sum_{k=1}^n X_{k,i} \big(A_{k,i} - \widehat{m}_{n,i}\big)^2}{\sum_{k=1}^n X_{k,i}}.$$

A couple of remarks are in order. First,

$$F_n := \sum_{i \in J^*} Z_{n,i} \xrightarrow{a.s.} 1 \quad \text{under } H_0.$$

Indeed, the factor F_n^{-4} has been inserted into the definition of $U_{n,j}$ in order that $K_{n,j}$ fails to converge in distribution to $\mathcal{N}(0,1)$ when H_0 is false, where $K_{n,j}$ is defined a few lines below. Second, $\sum_{k=1}^n X_{k,i} > 0$, eventually a.s., so that $\widehat{m}_{n,i}$ and $\widehat{\sigma}^2_{n,i}$ are well defined. Similarly, $\widehat{m}_n > 0$ eventually a.s..

Next, defining $C^*_{n,j}$ in the obvious way (i.e., with J^* in the place of $\{1, \ldots, d_0\}$), Theorem 4 implies

$$K_{n,j} := I_{\{U_{n,j}>0\}} \frac{C^*_{n,j}}{\sqrt{U_{n,j}}} \longrightarrow \mathcal{N}(0,1) \quad \text{stably under } H_0.$$

The converse is true as well, i.e., $K_{n,j}$ fails to converge in distribution to $\mathcal{N}(0,1)$ when H_0 is false. (This can be proved arguing as in Remark 10; we omit a formal proof). Thus, an asymptotic critical region for H_0, with approximate level α, is $\big\{|K_{n,j}| \geq u_\alpha\big\}$ with u_α satisfying $\mathcal{N}(0,1)(u_\alpha, \infty) = \frac{\alpha}{2}$. In real problems, sometimes, it is known in advance that $j_0 \in J$ for some $j_0 \in J^*$. Then, $j = j_0$ is a natural choice in the previous test. Otherwise, an alternative option is a critical region of the type $\bigcup_{i \in J^*} \big\{|K_{n,i}| \geq u_i\big\}$ for suitable u_i. This results in a more powerful test but requires the joint limit distribution of $\big(K_{n,i} : i \in J^*\big)$ under H_0. Such a distribution is given in [4] when $J^* = \{1, \ldots, d\}$, and can be easily obtained for arbitrary J^* using the techniques of this paper.

Example 8. (Another statistical use of $C^*_{n,j}$). As in Example 7 (and under the same assumptions), we aim to test $H_0 : J = J^*$. Contrary to Example 7, however, we are given observations $A_{k,j}$, $1 \leq k \leq n$, $1 \leq j \leq d$, but no urn is explicitly assigned. This is a main problem in statistical inference, usually faced by the ANOVA techniques and their very many ramifications. A solution to this problem is using $C^*_{n,j}$, as in Example 7, after *simulating* the $X_{n,j}$. The simulation is not hard. Take in fact an i.i.d. sequence $(Y_n : n \geq 0)$, independent of the $A_{k,j}$, with Y_0 uniformly distributed on $(0,1)$. Let $a_i = 1$, $Z_{0,i} = \frac{1}{d}$ for $i = 1, \ldots, d$, and

$$X_{1,j} = I_{\{F_{0,j-1} < Y_0 \leq F_{0,j}\}} \quad \text{where } F_{0,j} = \sum_{i=1}^{j} Z_{0,i} \text{ and } F_{0,0} = 0.$$

By induction, for each $n \geq 1$,

$$X_{n+1,j} = I_{\{F_{n,j-1} < Y_n \leq F_{n,j}\}} \quad \text{where } F_{n,j} = \sum_{i=1}^{j} Z_{n,i},$$

$$F_{n,0} = 0 \text{ and } Z_{n,i} = \frac{1 + \sum_{k=1}^n X_{k,i} A_{k,i}}{d + \sum_{r=1}^d \sum_{k=1}^n X_{k,r} A_{k,r}}.$$

Now, H_0 can be asymptotically tested as in Example 7. In addition, since $A_{k,i}$ is actually observed (unlike Example 7, where only $X_{k,i}A_{k,i}$ is observed), $\widehat{m}_{n,i}$ and $\widehat{\sigma}^2_{n,i}$ can be taken as

$$\widehat{m}_{n,i} = \frac{\sum_{k=1}^n A_{k,i}}{n} \quad \text{and} \quad \widehat{\sigma}^2_{n,i} = \frac{\sum_{k=1}^n \big(A_{k,i} - \widehat{m}_{n,i}\big)^2}{n}.$$

Clearly, this procedure needs to be much developed and investigated. By now, however, it looks (to us) potentially fruitful.

5. Proof of Theorem 4

Next result, of possible independent interest, is inspired by ideas in [4] and [5].

Proposition 9. *Let $\mathcal{F} = (\mathcal{F}_n)$ be an increasing filtration and (Y_n) an $\mathcal{F}$-adapted sequence of real integrable random variables. Suppose $Y_n \xrightarrow{a.s.} Y$ for some random variable Y and $H_n \in \mathcal{F}_n$ are events satisfying $P(H_n^c \text{ i.o.}) = 0$. Then,*

$$\sqrt{n}\,(Y_n - Y) \longrightarrow \mathcal{N}(0, V) \quad \mathcal{F}\text{-stably in strong sense},$$

for some random variable V, whenever

$$E\big\{I_{H_n}\,\big(E(Y_{n+1} \mid \mathcal{F}_n) - Y_n\big)^2\big\} = o(n^{-3}), \tag{4}$$

$$\sqrt{n}\,E\big\{I_{H_n}\,\sup_{k\geq n}|E(Y_{k+1} \mid \mathcal{F}_k) - Y_{k+1}|\big\} \longrightarrow 0, \tag{5}$$

$$n\sum_{k\geq n}(Y_k - Y_{k+1})^2 \xrightarrow{P} V. \tag{6}$$

Proof. We base on the following result, which is a consequence of Corollary 7 of [5]. *Let (L_n) be an $\mathcal{F}$-martingale such that $L_n \xrightarrow{a.s.} L$. Then, $\sqrt{n}\,(L_n - L) \longrightarrow \mathcal{N}(0, V)$ $\mathcal{F}$-stably in strong sense whenever*

(i) $\lim_n \sqrt{n}\,E\big\{I_{H_n}\,\sup_{k\geq n}|L_k - L_{k+1}|\big\} = 0$; **(ii)** $n\sum_{k\geq n}(L_k - L_{k+1})^2 \xrightarrow{P} V$.

Next, define the $\mathcal{F}$-martingale

$$L_0 = Y_0, \quad L_n = Y_n - \sum_{k=0}^{n-1} E\big(Y_{k+1} - Y_k \mid \mathcal{F}_k\big).$$

Define also $T_n = E\big(Y_{n+1} - Y_n \mid \mathcal{F}_n\big)$. By (4),

$$\sqrt{n}\sum_{k\geq n} E|I_{H_k}\,T_k| \leq \sqrt{n}\sum_{k\geq n}\sqrt{E(I_{H_k}\,T_k^2)} = \sqrt{n}\sum_{k\geq n} o(k^{-3/2}) \longrightarrow 0. \tag{7}$$

In particular, $\sum_{k=0}^{\infty} E|I_{H_k}\,T_k| < \infty$ so that $\sum_{k=0}^{n-1} I_{H_k}\,T_k$ converges a.s.. Since Y_n converges a.s. and $P(I_{H_n} \neq 1 \text{ i.o.}) = 0$,

$$L_n = Y_n - \sum_{k=0}^{n-1} T_k \xrightarrow{a.s.} L \quad \text{for some random variable } L.$$

Next, write

$$(L_n - L) - (Y_n - Y) = \sum_{k\geq n}(L_k - L_{k+1}) - \sum_{k\geq n}(Y_k - Y_{k+1}) = \sum_{k\geq n} T_k.$$

Recalling $\sqrt{n}\,\sum_{k\geq n}|I_{H_k}\,T_k| \xrightarrow{P} 0$ (thanks to (7)), one obtains

$$\Big|\sqrt{n}\,(L_n - L) - \sqrt{n}\,(Y_n - Y)\Big| = \sqrt{n}\,\Big|\sum_{k\geq n} T_k\Big|$$

$$\leq \sqrt{n}\sum_{k\geq n}|I_{H_k}\,T_k| + \sqrt{n}\sum_{k\geq n}|(1 - I_{H_k})\,T_k| \xrightarrow{P} 0.$$

Thus, it suffices to prove $\sqrt{n}\,(L_n - L) \longrightarrow \mathcal{N}(0,V)$ $\mathcal{F}$-stably in strong sense, that is, to prove conditions (i) and (ii). Condition (i) reduces to (5) after noting that $L_k - L_{k+1} = E(Y_{k+1} \mid \mathcal{F}_k) - Y_{k+1}$.

As to (ii), since $L_k - L_{k+1} = Y_k - Y_{k+1} + T_k$, condition (6) yields

$$n \sum_{k\ge n} (L_k - L_{k+1})^2 = V + n \sum_{k\ge n} \big\{T_k^2 + 2\,T_k(Y_k - Y_{k+1})\big\} + o_P(1).$$

By (4), $E\big\{n \sum_{k\ge n} I_{H_k} T_k^2\big\} = n \sum_{k\ge n} o(k^{-3}) \longrightarrow 0$. Since $P(I_{H_n} \neq 1 \text{ i.o.}) = 0$, then $n \sum_{k\ge n} T_k^2 \xrightarrow{P} 0$. Because of (6), this also implies

$$\big\{ n \sum_{k\ge n} T_k(Y_k - Y_{k+1})\big\}^2 \le n \sum_{k\ge n} T_k^2 \cdot n \sum_{k\ge n} (Y_k - Y_{k+1})^2 \xrightarrow{P} 0.$$

Therefore, condition (ii) holds and this concludes the proof. □

We next turn to Theorem 4. From now on, it is assumed $d_0 < d$ (the case $d_0 = d$ has been settled in [4]). Recall the notations $S_n^* = \sum_{i=1}^{d_0} N_{n,i}$ and $S_n = \sum_{i=1}^{d} N_{n,i}$. Note also that, by a straightforward calculation,

$$Z_{n+1,j}^* - Z_{n,j}^* = \frac{X_{n+1,j}\,A_{n+1,j}}{S_n^* + A_{n+1,j}} \;-\; Z_{n,j}^* \sum_{i=1}^{d_0} \frac{X_{n+1,i}\,A_{n+1,i}}{S_n^* + A_{n+1,i}}.$$

Proof of Theorem 4. The proof is split into two steps.

(i) $D_{n,j}^* \longrightarrow \mathcal{N}(0,V_j)$ $\mathcal{G}$-stably in strong sense.

By Lemma 3, $Z_{n,j}^* = \frac{Z_{n,j}}{\sum_{i=1}^{d_0} Z_{n,i}} \xrightarrow{a.s.} Z_{(j)}$. Further, $P(2\,S_n^* < n\,m \text{ i.o.}) = 0$ since $\frac{S_n^*}{n} \xrightarrow{a.s.} m$. Hence, by Proposition 9, it suffices to prove conditions (4)-(5)-(6) with

$$\mathcal{F}_n = \mathcal{G}_n, \quad Y_n = Z_{n,j}^*, \quad Y = Z_{(j)}, \quad H_n = \{2\,S_n^* \ge n\,m\}, \quad V = V_j.$$

Conditions (4) and (5) trivially hold. As to (4), note that

$$Z_{n,j}^* \sum_{i=1}^{d_0} Z_{n,i} = Z_{n,j} \sum_{i=1}^{d_0} Z_{n,i}^* = Z_{n,j}.$$

Therefore,

$$E\big\{Z_{n+1,j}^* - Z_{n,j}^* \mid \mathcal{G}_n\big\} = Z_{n,j}\,E\big\{\frac{A_{n+1,j}}{S_n^* + A_{n+1,j}} \mid \mathcal{G}_n\big\} \;-\; Z_{n,j}^* \sum_{i=1}^{d_0} Z_{n,i}\,E\big\{\frac{A_{n+1,i}}{S_n^* + A_{n+1,i}} \mid \mathcal{G}_n\big\}$$

$$= -Z_{n,j}\,E\big\{\frac{A_{n+1,j}^2}{S_n^*(S_n^* + A_{n+1,j})} \mid \mathcal{G}_n\big\} \;+\; Z_{n,j}^* \sum_{i=1}^{d_0} Z_{n,i}\,E\big\{\frac{A_{n+1,i}^2}{S_n^*(S_n^* + A_{n+1,i})} \mid \mathcal{G}_n\big\},$$

$$\text{so that} \quad I_{H_n} \Big| E\big\{Z_{n+1,j}^* - Z_{n,j}^* \mid \mathcal{G}_n\big\}\Big| \le I_{H_n}\,\frac{d_0\,\beta^2}{(S_n^*)^2} \le \frac{4\,d_0\,\beta^2}{m^2}\,\frac{1}{n^2}.$$

As to (5),

$$\Big|E\big(Z^*_{k+1,j}\mid\mathcal{G}_k\big)-Z^*_{k+1,j}\Big|\le\frac{2\,\beta}{S^*_k}+N_{k,j}\,\Big|E\big(\frac{1}{S^*_{k+1}}\mid\mathcal{G}_k\big)-\frac{1}{S^*_{k+1}}\Big|$$
$$\le\frac{2\,\beta}{S^*_k}+N_{k,j}\,\big(\frac{1}{S^*_k}-\frac{1}{S^*_k+\beta}\big)\le\frac{3\,\beta}{S^*_k},$$
$$\text{so that}\quad I_{H_n}\sup_{k\ge n}|E(Z^*_{k+1,j}\mid\mathcal{G}_k)-Z^*_{k+1,j}|\le I_{H_n}\,\frac{3\,\beta}{S^*_n}\le\frac{6\,\beta}{m}\,\frac{1}{n}.$$

Finally, let us turn to (6). For every $i\in\{1,\dots,d_0\}$,

$$n^2E\big\{\frac{A^2_{n+1,i}}{(S^*_n+A_{n+1,i})^2}\mid\mathcal{G}_n\big\}\le n^2\frac{EA^2_{n+1,i}}{(S^*_n)^2}\xrightarrow{a.s.}\frac{q_i}{m^2}\quad\text{and}$$
$$n^2E\big\{\frac{A^2_{n+1,i}}{(S^*_n+A_{n+1,i})^2}\mid\mathcal{G}_n\big\}\ge n^2\frac{EA^2_{n+1,i}}{(S^*_n+\beta)^2}\xrightarrow{a.s.}\frac{q_i}{m^2}.$$

Since $X_{n+1,r}\,X_{n+1,s}=0$ for $r\ne s$, it follows that

$$n^2E\big\{(Z^*_{n+1,j}-Z^*_{n,j})^2\mid\mathcal{G}_n\big\}=n^2Z_{n,j}\,(1-Z^*_{n,j})^2E\big\{\frac{A^2_{n+1,j}}{(S^*_n+A_{n+1,j})^2}\mid\mathcal{G}_n\big\}+$$
$$+\,n^2(Z^*_{n,j})^2\sum_{i\le d_0,i\ne j}Z_{n,i}\,E\big\{\frac{A^2_{n+1,i}}{(S^*_n+A_{n+1,i})^2}\mid\mathcal{G}_n\big\}$$
$$\xrightarrow{a.s.}Z_{(j)}(1-Z_{(j)})^2\frac{q_j}{m^2}+Z^2_{(j)}\sum_{i\le d_0,i\ne j}Z_{(i)}\frac{q_i}{m^2}=V_j.$$

Let $R_{n+1}=(n+1)^2I_{H_n}\,(Z^*_{n+1,j}-Z^*_{n,j})^2$. Since $H_n\in\mathcal{G}_n$ and $P(I_{H_n}\ne1\text{ i.o.})=0$, then $E(R_{n+1}\mid\mathcal{G}_n)\xrightarrow{a.s.}V_j$. On noting that $|Z^*_{n+1,j}-Z^*_{n,j}|\le\frac{d_0\,\beta}{S^*_n}$,

$$\frac{ER^2_n}{n^2}\le(d_0\,\beta)^4n^2E\big(\frac{I_{H_{n-1}}}{(S^*_{n-1})^4}\big)\le\big(\frac{2\,d_0\,\beta}{m}\big)^4\frac{n^2}{(n-1)^4}.$$

By Lemma 2 (applied with $Y_n=R_n$),

$$n\sum_{k\ge n}I_{H_k}(Z^*_{k+1,j}-Z^*_{k,j})^2=\frac{n}{n+1}\,(n+1)\sum_{k\ge n+1}\frac{R_k}{k^2}\xrightarrow{a.s.}V_j.$$

Since $P(I_{H_n}\ne1\text{ i.o.})=0$ then $n\sum_{k\ge n}(Z^*_{k+1,j}-Z^*_{k,j})^2\xrightarrow{a.s.}V_j$, that is, condition (6) holds.

(ii) $C^*_{n,j}\longrightarrow\mathcal{N}(0,U_j)$ **stably.**

Define $T_{n,i}=\sum_{k=1}^n X_{k,i}$, $T_{0,i}=0$, and note that

$$C^*_{n,j}=-\frac{\sqrt{n}\,Z^*_{n,j}}{1+\sum_{i=1}^{d_0}T_{n,i}}+\frac{n}{1+\sum_{i=1}^{d_0}T_{n,i}}\,\frac{T_{n,j}-Z^*_{n,j}\sum_{i=1}^{d_0}T_{n,i}}{\sqrt{n}}\quad\text{and}$$
$$T_{n,j}-Z^*_{n,j}\sum_{i=1}^{d_0}T_{n,i}=\sum_{k=1}^n\big\{X_{k,j}-Z^*_{k,j}\sum_{i=1}^{d_0}T_{k,i}+Z^*_{k-1,j}\sum_{i=1}^{d_0}T_{k-1,i}\big\}$$
$$=\sum_{k=1}^n\big\{X_{k,j}-Z^*_{k-1,j}\sum_{i=1}^{d_0}X_{k,i}-\sum_{i=1}^{d_0}T_{k,i}(Z^*_{k,j}-Z^*_{k-1,j})\big\}.$$

Define also $H_n = \{2S_n^* \geq n\,m\}$ and

$$C_{n,j}^{**} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} I_{H_{k-1}} \big\{ X_{k,j} - Z_{k-1,j}^* \sum_{i=1}^{d_0} X_{k,i} + \sum_{i=1}^{d_0} T_{k-1,i} \big(E(Z_{k,j}^* \mid \mathcal{G}_{k-1}) - Z_{k,j}^* \big) \big\}.$$

Recalling (from point (i)) that $P(I_{H_n} \neq 1 \text{ i.o.}) = 0$, $\lim_n \frac{\sum_{i=1}^{d_0} T_{n,i}}{n} = 1$ a.s., and $I_{H_{k-1}} \Big| E\big\{ Z_{k,j}^* - Z_{k-1,j}^* \mid \mathcal{G}_{k-1} \big\} \Big| \leq \frac{c}{(k-1)^2}$ a.s. for some constant c, it is not hard to see that $C_{n,j}^* \longrightarrow N$ stably if and only if $C_{n,j}^{**} \longrightarrow N$ stably for any kernel N.

We next prove $C_{n,j}^{**} \longrightarrow \mathcal{N}(0, U_j)$ stably. For $k = 1, \ldots, n$, let $\mathcal{F}_{n,k} = \mathcal{G}_k$ and

$$Y_{n,k} = \frac{I_{H_{k-1}} \big\{ X_{k,j} \; - \; Z_{k-1,j}^* \sum_{i=1}^{d_0} X_{k,i} + \sum_{i=1}^{d_0} T_{k-1,i} \, (E(Z_{k,j}^* \mid \mathcal{G}_{k-1}) - Z_{k,j}^*) \, \big\}}{\sqrt{n}}.$$

Since $E(Y_{n,k} \mid \mathcal{F}_{n,k-1}) = 0$ a.s., the martingale CLT (see Theorem 3.2 of [6]) applies. As a consequence, $C_{n,j}^{**} = \sum_{k=1}^{n} Y_{n,k} \longrightarrow \mathcal{N}(0, U_j)$ stably provided

$$\sup_n E\big(\max_{1 \leq k \leq n} Y_{n,k}^2 \big) < \infty; \quad \max_{1 \leq k \leq n} |Y_{n,k}| \xrightarrow{P} 0; \quad \sum_{k=1}^{n} Y_{n,k}^2 \xrightarrow{P} U_j.$$

As shown in point (i), $I_{H_{k-1}} \Big| E(Z_{k,j}^* \mid \mathcal{G}_{k-1}) - Z_{k,j}^* \Big| \leq \frac{d}{k-1}$ a.s. for a suitable constant d. Hence, the first two conditions follow from

$$Y_{n,k}^2 \leq \frac{2}{n} + \frac{2}{n} I_{H_{k-1}} (k-1)^2 \big(E(Z_{k,j}^* \mid \mathcal{G}_{k-1}) - Z_{k,j}^* \big)^2 \leq \frac{2\,(1+d^2)}{n} \quad \text{a.s..}$$

To conclude the proof, it remains to see that $\sum_{k=1}^{n} Y_{n,k}^2 \xrightarrow{P} U_j$. After some (long) algebra, the latter condition is shown equivalent to

$$\frac{1}{n} \sum_{k=1}^{n} I_{H_{k-1}} \big\{ X_{k,j} - Z_{k-1,j}^* + k \, \big(Z_{k-1,j}^* - Z_{k,j}^* \big) \big\}^2 \xrightarrow{P} U_j. \tag{8}$$

Let $R_{n+1} = (n+1)^2 I_{H_n} \, (Z_{n+1,j}^* - Z_{n,j}^*)^2$. Since $E(R_{n+1} \mid \mathcal{G}_n) \xrightarrow{a.s.} V_j$, as shown in point (i), Lemma 2 implies

$$\frac{1}{n} \sum_{k=1}^{n} I_{H_{k-1}} k^2 \big(Z_{k-1,j}^* - Z_{k,j}^* \big)^2 \xrightarrow{a.s.} V_j.$$

A direct calculation shows that

$$\frac{1}{n} \sum_{k=1}^{n} I_{H_{k-1}} (X_{k,j} - Z_{k-1,j}^*)^2 \xrightarrow{a.s.} Z_{(j)}(1 - Z_{(j)}).$$

Finally, observe the following facts

$$\big(Z_{n,j}^* - Z_{n+1,j}^* \big) \, (X_{n+1,j} - Z_{n,j}^*) = -(1 - Z_{n,j}^*) \, \frac{X_{n+1,j} \, A_{n+1,j}}{S_n^* + A_{n+1,j}} - Z_{n,j}^* (Z_{n,j}^* - Z_{n+1,j}^*),$$

$$(n+1) \, Z_{n,j}^* \, I_{H_n} \Big| E\big(Z_{n,j}^* - Z_{n+1,j}^* \mid \mathcal{G}_n \big) \Big| \leq \frac{c\,(n+1)}{n^2} \xrightarrow{a.s.} 0,$$

$$(n+1) \, E\big\{ \frac{X_{n+1,j} \, A_{n+1,j}}{S_n^* + A_{n+1,j}} \mid \mathcal{G}_n \big\} \leq \frac{n+1}{S_n^*} \, Z_{n,j} \, E A_{n+1,j} \xrightarrow{a.s.} Z_{(j)},$$

$$(n+1) \, E\big\{ \frac{X_{n+1,j} \, A_{n+1,j}}{S_n^* + A_{n+1,j}} \mid \mathcal{G}_n \big\} \geq \frac{n+1}{S_n^* + \beta} \, Z_{n,j} \, E A_{n+1,j} \xrightarrow{a.s.} Z_{(j)}.$$

Therefore,

$$(n+1)\, I_{H_n} E\big\{ \,(Z^*_{n,j} - Z^*_{n+1,j})\,(X_{n+1,j} - Z^*_{n,j}) \mid \mathcal{G}_n \big\} \xrightarrow{a.s.} -Z_{(j)}(1 - Z_{(j)})$$

and Lemma 2 again implies

$$\frac{2}{n} \sum_{k=1}^{n} I_{H_{k-1}} k\,(Z^*_{k-1,j} - Z^*_{k,j})\,(X_{k,j} - Z^*_{k-1,j}) \xrightarrow{a.s.} -2\, Z_{(j)}(1 - Z_{(j)}).$$

Thus condition (8) holds, and this concludes the proof.

□

Remark 10. Point (ii) admits a simpler proof in case $EA_{k,j} = m$ for all $k \geq 1$ and $1 \leq j \leq d_0$. This happens, in particular, if the sequence $(A_{n,1}, \ldots, A_{n,d})$ is i.i.d..

Given the real numbers $b_1, \ldots, b_{d_0}$, define

$$Y_{n,k} = \frac{1}{\sqrt{n}} \sum_{j=1}^{d_0} b_j\, X_{k,j}\,(A_{k,j} - EA_{k,j}), \quad \mathcal{F}_{n,k} = \mathcal{G}_k, \quad k = 1, \ldots, n.$$

By Lemma 2, $\sum_{k=1}^{n} Y_{n,k}^2 \xrightarrow{a.s.} \sum_{j=1}^{d_0} b_j^2\,(q_j - m^2)\, Z_{(j)} := L$. Thus, the martingale CLT implies $\sum_{k=1}^{n} Y_{n,k} \longrightarrow \mathcal{N}(0, L)$ stably. Since $b_1, \ldots, b_{d_0}$ are arbitrary constants,

$$\Big(\frac{\sum_{k=1}^{n} X_{k,j}\,(A_{k,j} - EA_{k,j})}{\sqrt{n}} \,:\, j = 1, \ldots, d_0 \Big) \longrightarrow \mathcal{N}_{d_0}(0, \Sigma) \quad \text{stably}$$

where Σ is the diagonal matrix with $\sigma_{j,j} = (q_j - m^2) Z_{(j)}$. Let $T_{n,j} = \sum_{k=1}^{n} X_{k,j}$. Since $EA_{k,j} = m$ and $\frac{T_{n,j}}{n} \xrightarrow{a.s.} Z_{(j)} > 0$ for all $j \leq d_0$, one also obtains

$$\Big(\sqrt{n}\, \big\{ \frac{\sum_{k=1}^{n} X_{k,j}\, A_{k,j}}{T_{n,j}} - m \big\} \,:\, j = 1, \ldots, d_0 \Big) \longrightarrow \mathcal{N}_{d_0}(0, \Gamma) \quad \text{stably}$$

where Γ is diagonal with $\gamma_{j,j} = \frac{(q_j - m^2)}{Z_{(j)}}$. Next, write

$$\begin{aligned}
\widetilde{C}_{n,j} &:= \sqrt{n}\,\big(\frac{T_{n,j}}{\sum_{i=1}^{d_0} T_{n,i}} - \frac{\sum_{k=1}^{n} X_{k,j} A_{k,j}}{\sum_{i=1}^{d_0} \sum_{k=1}^{n} X_{k,i} A_{k,i}} \big) \\
&= \frac{T_{n,j}}{\sum_{i=1}^{d_0} \sum_{k=1}^{n} X_{k,i} A_{k,i}} \, \frac{\sum_{i \leq d_0, i \neq j} T_{n,i}}{\sum_{i=1}^{d_0} T_{n,i}} \, \sqrt{n}\,\big(m - \frac{\sum_{k=1}^{n} X_{k,j} A_{k,j}}{T_{n,j}} \big) + \\
&+ \frac{T_{n,j}}{\sum_{i=1}^{d_0} \sum_{k=1}^{n} X_{k,i} A_{k,i}} \, \frac{1}{\sum_{i=1}^{d_0} T_{n,i}} \sum_{i \leq d_0, i \neq j} T_{n,i}\, \sqrt{n}\,\big(\frac{\sum_{k=1}^{n} X_{k,i} A_{k,i}}{T_{n,i}} - m \big).
\end{aligned}$$

Clearly, $C^*_{n,j} - \widetilde{C}_{n,j} \xrightarrow{a.s.} 0$. To conclude the proof, it suffices noting that $\widetilde{C}_{n,j}$ converges stably to the Gaussian kernel with mean 0 and variance

$$\Big(\frac{Z_{(j)}(1 - Z_{(j)})}{m} \Big)^2 \frac{q_j - m^2}{Z_{(j)}} + \frac{Z_{(j)}^2}{m^2} \sum_{i \leq d_0, i \neq j} Z_{(i)}^2 \, \frac{q_i - m^2}{Z_{(i)}} = U_j.$$

APPENDIX

Proof of Lemma 3. We first note that $N_{n,j} \xrightarrow{a.s.} \infty$ for each $j \le d_0$. Arguing as in the proof of Proposition 2.3 of [9], in fact, $\sum_{n=1}^{\infty} X_{n,j} = \infty$ a.s.. Hence, $\sum_{k=1}^{n} X_{k,j}\, EA_{k,j} \xrightarrow{a.s} \infty$, and $N_{n,j} \xrightarrow{a.s.} \infty$ follows from

$$L_n = N_{n,j} - \big\{a_j + \sum_{k=1}^{n} X_{k,j}\, EA_{k,j}\big\} = \sum_{k=1}^{n} X_{k,j}\,\big(A_{k,j} - EA_{k,j}\big)$$

is a $\mathcal{G}$-martingale such that $|L_{n+1} - L_n| \le \beta$ for all n.

We also need the following fact.

CLAIM: $\tau_{n,j} = \frac{N_{n,j}}{(S_n^*)^\lambda}$ converges a.s. for all $j > d_0$ and $\lambda \in (\frac{\lambda_0}{m}, 1)$.

On noting that $(1-x)^\lambda \le 1 - \lambda\, x$ for $0 \le x \le 1$ and $\sum_{i=1}^{d_0} Z_{n,i} = \frac{S_n^*}{S_n}$, one can estimate as follows

$$\begin{aligned}
E\big\{\frac{\tau_{n+1,j}}{\tau_{n,j}} - 1 \mid \mathcal{G}_n\big\} &= E\big\{\frac{N_{n,j} + X_{n+1,j}\, A_{n+1,j}}{N_{n,j}}\,(\frac{S_n^*}{S_{n+1}^*})^\lambda \mid \mathcal{G}_n\big\} - 1 \\
&\le \frac{Z_{n,j}\, EA_{n+1,j}}{N_{n,j}} \;+\; E\big\{(\frac{S_n^*}{S_{n+1}^*})^\lambda - 1 \mid \mathcal{G}_n\big\} \\
&\le \frac{EA_{n+1,j}}{S_n} \;-\; \lambda \sum_{i=1}^{d_0} E\big\{\frac{X_{n+1,i}\, A_{n+1,i}}{S_{n+1}^*} \mid \mathcal{G}_n\big\} \\
&\le \frac{EA_{n+1,j}}{S_n} \;-\; \lambda \sum_{i=1}^{d_0} \frac{Z_{n,i}\, EA_{n+1,i}}{S_n^* + \beta} \\
&= \frac{EA_{n+1,j}}{S_n} \;-\; \lambda\, EA_{n+1,1} \frac{S_n^*}{S_n(S_n^* + \beta)} \\
&= \frac{1}{S_n}\,\big(EA_{n+1,j} \;-\; \lambda\, EA_{n+1,1} \frac{S_n^*}{S_n^* + \beta}\big) \quad \text{a.s..}
\end{aligned}$$

Since $\limsup_n \big(EA_{n+1,j} - \lambda\, EA_{n+1,1}\big) \le \lambda_0 - \lambda\, m < 0$, there are $\epsilon > 0$ and $n_0 \ge 1$ such that $EA_{n+1,j} - \lambda\, EA_{n+1,1} \le -\epsilon$ whenever $n \ge n_0$. Thus,

$$E\big\{\tau_{n+1,j} - \tau_{n,j} \mid \mathcal{G}_n\big\} = \tau_{n,j}\, E\big\{\frac{\tau_{n+1,j}}{\tau_{n,j}} - 1 \mid \mathcal{G}_n\big\} \le 0 \quad \text{a.s. whenever } n \ge n_0 \text{ and } S_n^* \ge c$$

for a suitable constant c. Since $S_n^* \ge N_{n,1} \xrightarrow{a.s.} \infty$, thus, $(\tau_{n,j})$ is eventually a non negative $\mathcal{G}$-super-martingale. Hence, $\tau_{n,j}$ converges a.s..

Let $\lambda \in (\frac{\lambda_0}{m}, 1)$. A first consequence of the Claim is that $Z_{n,j} \le \frac{\tau_{n,j}}{S_n^{1-\lambda}} \xrightarrow{a.s.} 0$ for each $j > d_0$. Letting $Y_n = \sum_{i=1}^{d_0} X_{n,i}\, A_{n,i}$, this implies

$$E(Y_{n+1} \mid \mathcal{G}_n) = \sum_{i=1}^{d_0} Z_{n,i}\, EA_{n+1,i} = EA_{n+1,1}\,(1 - \sum_{i=d_0+1}^{d} Z_{n,i}) \xrightarrow{a.s.} m.$$

Thus, Lemma 2 yields $\frac{S_n^*}{n} \xrightarrow{a.s.} m$. Similarly, $\frac{S_n}{n} \xrightarrow{a.s.} m$. Applying the Claim again,

$$n^{1-\lambda} Z_{n,j} = (\frac{n}{S_n})^{1-\lambda}\,(\frac{S_n^*}{S_n})^\lambda\, \tau_{n,j} \quad \text{converges a.s. for each } j > d_0.$$

Since $j > d_0$ and $\lambda \in (\frac{\lambda_0}{m}, 1)$ are arbitrary, it follows that $n^{1-\lambda} \sum_{j=d_0+1}^{d} Z_{n,j} \xrightarrow{a.s.} 0$ for each $\lambda > \frac{\lambda_0}{m}$.

Next, fix $j \le d_0$. For $Z_{n,j}$ to converge a.s., it suffices that

$$\sum_n E\big\{Z_{n+1,j} - Z_{n,j} \mid \mathcal{G}_n\big\} \text{ and } \sum_n E\big\{(Z_{n+1,j} - Z_{n,j})^2 \mid \mathcal{G}_n\big\} \quad \text{converge a.s.;}$$

see Lemma 3.2 of [11]. Since

$$Z_{n+1,j} - Z_{n,j} = \frac{X_{n+1,j}\, A_{n+1,j}}{S_n + A_{n+1,j}} \;-\; Z_{n,j} \sum_{i=1}^{d} \frac{X_{n+1,i}\, A_{n+1,i}}{S_n + A_{n+1,i}},$$

then $|Z_{n+1,j} - Z_{n,j}| \le \frac{d\,\beta}{S_n}$. Hence,

$$\sum_n E\big\{(Z_{n+1,j} - Z_{n,j})^2 \mid \mathcal{G}_n\big\} \le d^2\beta^2 \sum_n \frac{1}{n^2}\,(\frac{n}{S_n})^2 < \infty \quad \text{a.s..}$$

Moreover,

$$E\big\{Z_{n+1,j} - Z_{n,j} \mid \mathcal{G}_n\big\} = Z_{n,j}\, E\big\{\frac{A_{n+1,j}}{S_n + A_{n+1,j}} \mid \mathcal{G}_n\big\} \;-\; Z_{n,j} \sum_{i=1}^{d} Z_{n,i}\, E\big\{\frac{A_{n+1,i}}{S_n + A_{n+1,i}} \mid \mathcal{G}_n\big\}$$

$$= -Z_{n,j}\, E\big\{\frac{A_{n+1,j}^2}{S_n(S_n + A_{n+1,j})} \mid \mathcal{G}_n\big\} \;+\; Z_{n,j} \sum_{i=1}^{d} Z_{n,i}\, E\big\{\frac{A_{n+1,i}^2}{S_n(S_n + A_{n+1,i})} \mid \mathcal{G}_n\big\} +$$

$$+ Z_{n,j}\, \frac{EA_{n+1,j}}{S_n} \;-\; Z_{n,j} \sum_{i=1}^{d} Z_{n,i}\, \frac{EA_{n+1,i}}{S_n} \quad \text{a.s., and}$$

$$EA_{n+1,j} \;-\; \sum_{i=1}^{d} Z_{n,i}\, EA_{n+1,i} = EA_{n+1,1} \sum_{i=d_0+1}^{d} Z_{n,i} \;-\; \sum_{i=d_0+1}^{d} Z_{n,i} EA_{n+1,i}.$$

Therefore, $\sum_n E\big\{Z_{n+1,j} - Z_{n,j} \mid \mathcal{G}_n\big\}$ converges a.s. since

$$\Big|E\big\{Z_{n+1,j} - Z_{n,j} \mid \mathcal{G}_n\big\}\Big| \le \frac{d\,\beta^2}{S_n^2} + 2\,\beta\,\frac{\sum_{i=d_0+1}^{d} Z_{n,i}}{S_n} = \mathrm{o}(n^{\lambda-2}) \quad \text{a.s. for each } \lambda \in (\frac{\lambda_0}{m},\, 1).$$

Thus, $Z_{n,j} \xrightarrow{a.s.} Z_{(j)}$ for some random variable $Z_{(j)}$. To conclude the proof, we let $Y_{n,i} = \log \frac{Z_{n,i}}{Z_{n,1}}$ and prove that

$$\sum_n E\big\{Y_{n+1,i} - Y_{n,i} \mid \mathcal{G}_n\big\} \text{ and } \sum_n E\big\{(Y_{n+1,i} - Y_{n,i})^2 \mid \mathcal{G}_n\big\} \text{ converge a.s. whenever } i \le d_0.$$

In this case, in fact, $\log \frac{Z_{n,i}}{Z_{n,1}}$ converges a.s. for each $i \le d_0$ and this implies $Z_{(i)} > 0$ a.s. for each $i \le d_0$.

Since $Y_{n+1,i} - Y_{n,i} = X_{n+1,i} \log\big(1 + \frac{A_{n+1,i}}{N_{n,i}}\big) - X_{n+1,1} \log\big(1 + \frac{A_{n+1,1}}{N_{n,1}}\big)$, then

$$E\big\{Y_{n+1,i} - Y_{n,i} \mid \mathcal{G}_n\big\} = Z_{n,i} E\big\{\log\big(1 + \frac{A_{n+1,i}}{N_{n,i}}\big) \mid \mathcal{G}_n\big\} - Z_{n,1} E\big\{\log\big(1 + \frac{A_{n+1,1}}{N_{n,1}}\big) \mid \mathcal{G}_n\big\} \quad \text{a.s..}$$

Since $EA_{n+1,i} = EA_{n+1,1}$, a second order Taylor expansion of $x \mapsto \log(1+x)$ yields

$$\Big|E\big\{Y_{n+1,i} - Y_{n,i} \mid \mathcal{G}_n\big\}\Big| \le \frac{\beta^2}{S_n}\,\big(\frac{1}{N_{n,i}} + \frac{1}{N_{n,1}}\big) \quad \text{a.s..}$$

A quite similar estimate holds for $E\big\{(Y_{n+1,i} - Y_{n,i})^2 \mid \mathcal{G}_n\big\}$. Thus, it suffices to see

$$\sum_n \frac{1}{S_n\, N_{n,i}} < \infty \quad \text{a.s. for each } i \le d_0.$$

Define $R_{n,i} = \frac{(S_n^*)^u}{N_{n,i}}$ where $u \in (0,1)$ and $i \le d_0$. Since $(1+x)^u \le 1 + u\,x$ for $x \ge 0$, one can estimate as

$$E\Big\{\frac{R_{n+1,i}}{R_{n,i}} - 1 \mid \mathcal{G}_n\Big\} = E\Big\{(\frac{S_{n+1}^*}{S_n^*})^u - 1 \mid \mathcal{G}_n\Big\} - E\Big\{(\frac{S_{n+1}^*}{S_n^*})^u \frac{X_{n+1,i}\,A_{n+1,i}}{N_{n,i}+A_{n+1,i}} \mid \mathcal{G}_n\Big\}$$
$$\le u\,E\Big\{\frac{S_{n+1}^* - S_n^*}{S_n^*} \mid \mathcal{G}_n\Big\} - E\Big\{\frac{X_{n+1,i}\,A_{n+1,i}}{N_{n,i}+\beta} \mid \mathcal{G}_n\Big\}$$
$$= \frac{u}{S_n^*}\sum_{p=1}^{d_0} Z_{n,p}\,EA_{n+1,p} - \frac{Z_{n,i}EA_{n+1,i}}{N_{n,i}+\beta} = \frac{EA_{n+1,1}}{S_n}\Big\{u - \frac{N_{n,i}}{N_{n,i}+\beta}\Big\} \text{ a.s..}$$

As in the proof of the Claim, thus,

$$E\big\{R_{n+1,i} - R_{n,i} \mid \mathcal{G}_n\big\} = R_{n,i}\,E\Big\{\frac{R_{n+1,i}}{R_{n,i}} - 1 \mid \mathcal{G}_n\Big\} \le 0 \quad \text{a.s. whenever } N_{n,i} \ge c$$

for a suitable constant c. Since $N_{n,i} \xrightarrow{a.s.} \infty$, then $(R_{n,i})$ is eventually a non negative $\mathcal{G}$-super-martingale, so that $R_{n,i}$ converges a.s.. Hence,

$$\sum_n \frac{1}{S_n\,N_{n,i}} = \sum_n \frac{R_{n,i}}{S_n\,(S_n^*)^u} = \sum_n R_{n,i}\,\frac{n}{S_n}\,(\frac{n}{S_n^*})^u\,\frac{1}{n^{1+u}} < \infty \quad \text{a.s..}$$

This concludes the proof.

□

References

[1] Aletti G., May C. and Secchi P. (2008) A central limit theorem, and related results, for a two-color randomly reinforced urn, *Preprint*, currently available at: `ArXiv:math.PR/0811.2097v1`

[2] Bay Z.D. and Hu F. (2005) Asymptotics in randomized urn models, *Ann. Appl. Probab.*, 15, 914-940.

[3] Berti P., Pratelli L. and Rigo P. (2004) Limit theorems for a class of identically distributed random variables, *Ann. Probab.*, 32, 2029-2052.

[4] Berti P., Crimaldi I., Pratelli L. and Rigo P. (2009) A central limit theorem and its applications to multicolor randomly reinforced urns, *submitted*, currently available at: `http://arxiv.org/abs/0904.0932`

[5] Crimaldi I., Letta G. and Pratelli L. (2007) A strong form of stable convergence, *Sem. de Probab. XL*, LNM, 1899, 203-225.

[6] Hall P. and Heyde C.C. (1980) *Martingale limit theory and its applications*, Academic Press.

[7] Janson S. (2004) Functional limit theorems for multitype branching processes and generalized Polya urns, *Stoch. Proc. Appl.*, 110, 177-245.

[8] Janson S. (2005) Limit theorems for triangular urn schemes, *Probab. Theo. Rel. Fields*, 134, 417-452.

[9] May C. and Flournoy N. (2009) Asymptotics in response-adaptive designs generated by a two-color, randomly reinforced urn, *Ann. Statist.*, 37, 1058-1078.

[10] Muliere P., Paganoni A.M. and Secchi P. (2006) A randomly reinforced urn, *J. Statist. Plann. Inference*, 136, 1853-1874.

[11] Pemantle R. and Volkov S. (1999) Vertex-reinforced random walk on $\mathbb{Z}$ has finite range, *Ann. Probab.*, 27, 1368-1388.

[12] Pemantle R. (2007) A survey of random processes with reinforcement, *Probab. Surveys*, 4, 1-79.

Patrizia Berti, Dipartimento di Matematica Pura ed Applicata "G. Vitali", Universita' di Modena e Reggio-Emilia, via Campi 213/B, 41100 Modena, Italy
E-mail address: patrizia.berti@unimore.it

Irene Crimaldi, Dipartimento di Matematica, Universita' di Bologna, Piazza di Porta San Donato 5, 40126 Bologna, Italy
E-mail address: crimaldi@dm.unibo.it

Luca Pratelli, Accademia Navale, viale Italia 72, 57100 Livorno, Italy
E-mail address: pratel@mail.dm.unipi.it

Pietro Rigo (corresponding author), Dipartimento di Economia Politica e Metodi Quantitativi, Universita' di Pavia, via S. Felice 5, 27100 Pavia, Italy
E-mail address: prigo@eco.unipv.it