



HAL
open science

A central limit theorem for the number of descents and some urn models

Olivier Garet

► **To cite this version:**

Olivier Garet. A central limit theorem for the number of descents and some urn models. *Markov Processes And Related Fields*, In press, 27 (5), pp.789-801. 10.48550/arXiv.1906.02980 . hal-02149176v3

HAL Id: hal-02149176

<https://hal.science/hal-02149176v3>

Submitted on 3 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A CENTRAL LIMIT THEOREM FOR THE NUMBER OF DESCENTS AND SOME URN MODELS

OLIVIER GARET

ABSTRACT. The purpose of this work is to establish a Central Limit Theorem that can be applied to a particular form of Markov chains, including the number of descents in a random permutation of \mathfrak{S}_n , two-type generalized Pólya urns, and some other urn models.

The purpose of this work is to establish a Central Limit Theorem that can be applied to a particular form of Markov chains, a form that is encountered in various problems of a combinatorial nature. This class is large enough to cover a number of classical results, such as the Central Limit Theorem for the number of descents in a random permutation of \mathfrak{S}_n , two-type generalized Pólya urns, as well as some other urn models.

We can immediately state the theorem:

Theorem 0.1. *Let $(a_n)_{n \geq 1}$ be a sequence of integrable random variables. We note, for $n \geq 1$, $S_n = a_1 + \dots + a_n$. We suppose that there exists some constant M and, for $k \in \{1, 2, 3\}$, non-random sequences $(\alpha_k^n)_{n \geq 1}$, $(D_k^n)_{n \geq 1}$ and real numbers $\alpha_1, \alpha_2, \alpha_3, D_1, D_2, D_3$ such that*

- $\forall n \geq 1 \quad |a_n| \leq M.$
- $\forall n \geq 1 \quad \forall k \in \{1, 2, 3\} \quad \mathbb{E}[a_{n+1}^k | a_1, \dots, a_n] = D_k^n - \frac{\alpha_k^n}{n} S_n,$ with
 $\lim_{n \rightarrow +\infty} \alpha_k^n = \alpha_k$ and $\lim_{n \rightarrow +\infty} D_k^n = D_k.$
- $\alpha_1^n = \alpha_1 + o\left(\frac{1}{\sqrt{n}}\right)$ and $D_1^n = D_1 + o\left(\frac{1}{\sqrt{n}}\right).$

We also compute $\ell = \frac{D_1}{\alpha_1 + 1}$ and $D = D_2 - \ell(\ell + \alpha_2).$

If it is then verified that $\alpha_1 > -\frac{1}{2}$ and $D > 0$, then

$\frac{S_n - n\ell}{\sqrt{n}}$ converges in distribution to $\mathcal{N}(0, S)$ with $S = \frac{D}{2\alpha_1 + 1}.$

The article is organized as follows: in a first part, as an appetizer, we recall the definitions, the Central Limit Theorem on the number of descents and we see how to deduce it from our general theorem. The second part is devoted to the proof of the theorem, while the last two parts give applications to urn models, including the Pólya urn model, but not limited to it.

2000 *Mathematics Subject Classification.* 60F05, 60C05.

Key words and phrases. Central Limit Theorem, urn models.

1. THE NUMBER OF DESCENTS

When σ is a permutation of the set $\{1, \dots, n\}$, the descents of σ are the points

$$Desc(\sigma) = \{i \in \{1, \dots, n-1\}; \sigma(i) > \sigma(i+1)\}.$$

Similarly, the ascents are defined by

$$Asc(\sigma) = \{i \in \{1, \dots, n-1\}; \sigma(i) < \sigma(i+1)\}.$$

It is well known that the number of descents (or the number of ascents) of a uniformly chosen random permutation has an asymptotic normal behaviour: if σ_n is chosen uniformly among the permutations on $\{1, \dots, n\}$, we have

$$\frac{|Desc(\sigma_n)| - n/2}{\sqrt{n}} \implies \mathcal{N}\left(0, \frac{1}{12}\right).$$

In a recent article [2], Chatterjee and Diaconis counted no less than 6 proofs. Thus, we venture here to present a seventh – and perhaps even an eighth, if we count as such the transcription of the question into an urn problem, which can be studied using this article or the literature. Note also the recent proof by Özdemir using martingales [7]. Our proof of Theorem 0.1 is based on conditional expectation together with the method of moments. Among the six proofs for the CLT on the number of descents evoked by Chatterjee and Diaconis, the proof by David and Barton [3] (Chapter 10) also relies on the method of moments, however they use the generating functions whereas our approach is more related to conditioning.

An easy way to perform the uniform distribution on the permutations in $\{1, \dots, n\}$ is to apply the insertion algorithm: take σ_1 as the identity on $\{1\}$, then $\sigma_{n+1} = \sigma_n \circ (n+1 \ U_{n+1})$, where (U_n) is a sequence of independent random variables such that for each n , U_n follows the uniform distribution on $\{1, \dots, n\}$. Then, for each n , σ_n follows the uniform distribution on \mathfrak{S}_n .

Let $D_n = |Desc(\sigma_n)|$. We have, for $n \geq 1$:

$$D_{n+1} = D_n + \mathbb{1}_{E_{n+1}}, \text{ with } E_{n+1} = \{U_{n+1} - 1 \in Asc(\sigma_n) \cup \{0\}\}.$$

Let also $\mathcal{F}_n = \sigma(U_1, \dots, U_n)$, then

$$(1) \quad \mathbb{P}(E_{n+1} | \mathcal{F}_n) = \frac{|Asc(\sigma_n)| + 1}{n+1} = \frac{(n-1-D_n) + 1}{n+1} = \frac{n-D_n}{n+1}$$

Let $S_n = D_n - \frac{n-1}{2}$. Let us define $S_1 = 0$, $a_1 = S_1$, and for $n \geq 1$:

$$a_{n+1} = S_{n+1} - S_n = (D_{n+1} - D_n) - \frac{1}{2} = \mathbb{1}_{E_{n+1}} - \frac{1}{2},$$

then $S_n = a_1 + \dots + a_n$. The random variable a_n is $\{-1/2, 1/2\}$ valued.

Finally,

$$(2) \quad \mathbb{E}(a_{n+1} | \mathcal{F}_n) = \mathbb{P}(E_{n+1} | \mathcal{F}_n) - \frac{1}{2} = \frac{\frac{n-1}{2} - D_n}{n+1} = -\frac{S_n}{n+1}.$$

Theorem 0.1 applies particularly easily: we have $\mathcal{F}_n = \sigma(a_1, \dots, a_n)$, (a_n) is uniformly bounded by $1/2$, Equation (2) gives the second assumption for

$k = 1$, $k = 2$ and $k = 3$ easily follow because the sequence (a_n^2) is constant equal to $1/4$: we can apply the theorem with $M = 1/2$, $D_1 = 0$, $D_2 = 1/4$, $D_3 = 0$ and $\alpha_1 = 1$, $\alpha_2 = 0$, $\alpha_3 = 1/4$, which gives the announced result.

2. PROOF OF THE THEOREM

We already note $\mathcal{F}_n = \sigma(a_1, \dots, a_n)$. Let's start with the study of the first moment.

$$\begin{aligned} \mathbb{E}(S_{n+1}) &= \mathbb{E}(S_n) + \mathbb{E}(a_{n+1}) = \mathbb{E}(S_n) + D_1^n - \frac{\alpha_1^n}{n} \mathbb{E}(S_n) \\ &= \left(1 - \frac{\alpha_1^n}{n}\right) \mathbb{E}(S_n) + D_1^n \\ &= \left(1 - \frac{\alpha_1}{n}\right) \mathbb{E}(S_n) + D_1 + o\left(\frac{1}{\sqrt{n}}\right) \text{ since } |\mathbb{E}(S_n)| \leq nM \end{aligned}$$

To solve the recurrence, we rely on two analytic lemmas:

Lemma 2.1. *Let $C, k, \alpha, \beta \in \mathbb{R}$ with $\alpha - 1 < \beta \leq \alpha$, $\beta + k > -1$ and $(u_n)_{n \geq n_0}$, sequences such that*

$$u_{n+1} = \left(1 - \frac{k}{n+c}\right) u_n + Cn^\alpha + o(n^\beta),$$

Proof. Note that

$$1 - \frac{k}{n+c} = \frac{n+c-k}{n+c} = \frac{\Gamma(n+c-k+1)}{\Gamma(n+c-k)} \frac{\Gamma(n+c)}{\Gamma(n+c+1)} = \frac{P_{k,c}(n)}{P_{k,c}(n+1)},$$

with $P_{k,c}(x) = \frac{\Gamma(x+c)}{\Gamma(x+c-k)}$.

By the Stirling formula, $P_{k,c}(x) = x^k(1 + O(1/x))$ at infinity, the recurrence equation is then rewritten

$$\begin{aligned} P_{k,c}(n+1)u_{n+1} &= P_{k,c}(n)u_n + CP_{k,c}(n+1)n^\alpha + o(n^{\beta+k}) \\ &= P_{k,c}(n)u_n + Cn^{\alpha+k} + o(n^{\beta+k}), \end{aligned}$$

and then, with the telescopic sum, $u_n P_{k,c}(n) = C \frac{n^{\alpha+k+1}}{\alpha+k+1} + o(n^{\beta+k+1})$, and finally

$$u_n = \frac{C}{\alpha+k+1} n^{\alpha+1} + o(n^{\beta+1}).$$

□

Lemma 2.2. *Let $C, k, \alpha \in \mathbb{R}$ with $C > 0$, $\alpha + k > -1$ and $(u_n)_{n \geq n_0}$, $(k_n)_{n \geq n_0}$ sequences such that*

$$u_{n+1} = \left(1 - \frac{k_n}{n+c}\right) u_n + Cn^\alpha + o(n^\beta), \text{ with } \lim k_n = k.$$

We suppose moreover that $\forall n \geq n_0$ $u_n \geq 0$.

Then,

$$u_n = \frac{C}{\alpha+k+1} n^{\alpha+1} + o(n^{\alpha+1}).$$

Proof. One can find n_0 such that $n + c > 0$, $v_n > 0$, and $1 - \frac{k_n}{n+c} > 0$ for $n \geq n_0$, then for $\varepsilon > 0$ small enough, take $n_1 \geq n_0$ such that $1 - \frac{k-\varepsilon}{n+c} > 1 - \frac{k_n}{n+c} > 1 - \frac{k+\varepsilon}{n+c} > 0$ for $n \geq n_1$. Let $w_{n_1} = x_{n_1} = u_{n_1}$, then for $n \geq n_1$,

$$w_{n+1} = w_n \left(1 - \frac{k + \varepsilon}{n + c}\right) + v_n \text{ and } x_{n+1} = x_n \left(1 - \frac{k - \varepsilon}{n + c}\right) + v_n.$$

By natural induction, $x_n \geq u_n \geq w_n \geq 0$ for each $n \geq n_1$, then

$$\liminf_{n \rightarrow +\infty} u_n n^{-(\alpha+1)} \geq \liminf_{n \rightarrow +\infty} w_n n^{-(\alpha+1)} = \frac{C}{\alpha + k + \varepsilon + 1},$$

where the last equality follows from Lemma 2.1. Letting ε tend to 0, we get

$$\liminf_{n \rightarrow +\infty} u_n n^{-(\alpha+1)} \geq \frac{C}{\alpha + k + 1}. \text{ Similarly}$$

$$\limsup_{n \rightarrow +\infty} u_n n^{-(\alpha+1)} \leq \limsup_{n \rightarrow +\infty} x_n n^{-(\alpha+1)} = \frac{C}{\alpha + k - \varepsilon + 1},$$

and, letting ε tend to 0, $\limsup_{n \rightarrow +\infty} u_n n^{-(\alpha+1)} \leq \frac{C}{\alpha + k + 1}$, which completes the proof. \square

Applying Lemma 2.1 to the sequence $(\mathbb{E}(S_n))_{n \geq 1}$, we get $\mathbb{E}(S_n) = n\ell + o(\sqrt{n})$, with $\ell = \frac{D_1}{\alpha_1 + 1}$.

By the Slutsky Lemma, it is now sufficient to prove that $\frac{S_n - \mathbb{E}(S_n)}{\sqrt{n}} \implies \mathcal{N}(0, D/(2\alpha_1 + 1))$.

We will first reduce to the case where $D_1^n = 0$ for all n . Let $\ell_n = \mathbb{E}(a_n)$ and $a'_n = a_n - \ell_n$.

Since $\mathbb{E}[a_{n+1}|a_1, \dots, a_n] = D_1^n - \frac{\alpha_1^n}{n} S_n$, we have $\ell_n = D_1^n - \frac{\alpha_1^n}{n} \mathbb{E}(S_n)$, so $\ell_n \rightarrow D_1 - \alpha_1 \ell = (\alpha_1 + 1)\ell - \alpha_1 \ell = \ell$.

It is clear that (a'_n) checks the same sort of assumptions as (a_n) , only the limit values are changed. If we put $f_{k,n}(x) = D_k^n - \frac{\alpha_k^n}{n} x$, we have $\mathbb{E}(a_{n+1}|\mathcal{F}_n) = f_{1,n}(S_n)$, then $\ell_{n+1} = f_{1,n}(\mathbb{E}(S_n))$, and

$$\mathbb{E}(a'_{n+1}|\mathcal{F}_n) = f_{1,n}(S_n) - f_{1,n}(\mathbb{E}(S_n)) = -\frac{\alpha_1^n}{n}(S_n - \mathbb{E}(S_n))$$

which gives the constants associated to a'_n : $\alpha'_1 = \alpha_1$ and $D_1'^n = 0$ for all n . We also have

$$\begin{aligned} \mathbb{E}(a_{n+1}'^2|\mathcal{F}_n) &= \mathbb{E}(a_{n+1}^2 + \ell_n^2 - 2\ell_n a_{n+1}|\mathcal{F}_n) \\ &= f_{2,n}(S_n) + \ell_n^2 - 2\ell_n f_{1,n}(S_n) \\ &= f_{2,n}(S'_n) - \alpha_2^n \frac{\ell_1 + \dots + \ell_n}{n} + \ell_n^2 - 2\ell_n f_{1,n}(S'_n) + 2\alpha_1^n \ell_n \frac{\ell_1 + \dots + \ell_n}{n} \\ &= f'_{2,n}(S'_n), \end{aligned}$$

with $f'_{2,n}(x) = f_{2,n}(x) - \alpha_2^n \frac{\ell_1 + \dots + \ell_n}{n} + \ell_n^2 - 2\ell_n f_{1,n}(x) + 2\alpha_1^n \ell_n \frac{\ell_1 + \dots + \ell_n}{n}$, which leads to

$$\begin{aligned} D'_2 &= \lim f'_{2,n}(0) = D_2 - \alpha_2 \ell + \ell^2 - 2\ell D_1 + 2\alpha_1 \ell^2 \\ &= D_2 - \ell(\alpha_2 - \ell + 2D_1 - 2\alpha_1 \ell) \\ &= D_2 - \ell(\alpha_2 + \ell - 2(D_1 - \ell(1 + \alpha_1))) \\ &= D_2 - \ell(\ell + \alpha_2) \end{aligned}$$

The same kind of calculation applies for $k = 3$, it is not detailed because the explicit values of the corresponding constants are not necessary.

Then, we can assume without loss of generality that $D_1^n = 0$ and $\mathbb{E}(S_n) = 0$ for all n . We will now estimate by recurrence the sequences of the different moments, for the orders greater than one.

For each non-negative integer k , there exist polynomials R_k and T_k such that $R_k(x, y) = \frac{k(k-1)(k-2)}{6} x^{k-3} y^3 + T_k(x, y)$

$$(x + y)^k = x^k + kx^{k-1}y + \frac{k(k-1)}{2} x^{k-2}y^2 + R_k(x, y),$$

with

$$\forall x \in \mathbb{R} \quad \forall y \in [-M, M] \quad |R_k(x, y)| \leq (M+1)^k (1 \vee |x|^{k-3})$$

$$\forall x \in \mathbb{R} \quad \forall y \in [-M, M] \quad |T_k(x, y)| \leq (M+1)^k (1 \vee |x|^{k-4})$$

Particularly, we get

$$S_{n+1}^k = S_n^k + kS_n^{k-1}a_{n+1} + \frac{k(k-1)}{2} S_n^{k-2}a_{n+1}^2 + R_k(S_n, a_{n+1}).$$

Conditioning by \mathcal{F}_n , then reintegrating, we have

$$\begin{aligned} \mathbb{E}(S_{n+1}^k) &= (1 - \frac{k\alpha_1^n}{n})\mathbb{E}(S_n^k) + \frac{D_2^n k(k-1)}{2} \mathbb{E}(S_n^{k-2}) \\ &\quad + (kD_1^n - \frac{k(k-1)}{2} \frac{\alpha_2^n}{n})\mathbb{E}(S_n^{k-1}) + \mathbb{E}(R_k(S_n, a_{n+1})) \end{aligned}$$

so, since $D_1^n = 0$, we have

$$\begin{aligned} \mathbb{E}(S_{n+1}^k) &= (1 - \frac{k\alpha_1^n}{n})\mathbb{E}(S_n^k) + \frac{D_2^n k(k-1)}{2} \mathbb{E}(S_n^{k-2}) \\ (3) \quad &\quad + \mathbb{E}(R_k(S_n, a_{n+1})) + O(\frac{1}{n} \mathbb{E}(S_n^{k-1})). \end{aligned}$$

For $k = 2$, we have $R_2 = 0$ and $\mathbb{E}(S_n) = 0$, which simply gives

$$\mathbb{E}(S_{n+1}^2) = (1 - \frac{2\alpha_1^n}{n})\mathbb{E}(S_n^2) + D_2^n.$$

With Lemma 2.2, we get

$$\mathbb{E}(S_n^2) = D'n + o(n), \text{ where } D' = \frac{D}{2\alpha_1 + 1}.$$

Let's move on to the general case. With the help of the recursion (3), we will prove that for each $k \in \mathbb{N}$, we have

$$(4) \quad \mathbb{E}(S_n^k) = C_k n^{k/2} + o(n^{k/2}),$$

where $(C_k)_{k \geq 0}$ is the sequence defined by $C_0 = 1$, $C_1 = 0$ and $C_k = D' C_{k-2}(k-1)$: since $D' > 0$, this is the sequence for the moments of $\mathcal{N}(0, D')$. We note that terms with even index are positive. This will be useful to apply Lemma 2.2.

The proof is by induction on k . The asymptotic expansion is proved for $k = 0, 1, 2$. Assuming (4) holds for $\{0, 1, \dots, k-1\}$ and let us prove for k . We invoke the inductive assumption for $k-2$ and $k-1$; it comes

$$(5) \quad \begin{aligned} \mathbb{E}(S_{n+1}^k) &= \left(1 - \frac{k\alpha_1^n}{n}\right) \mathbb{E}(S_n^k) + C_{k-2} \frac{D_2 k(k-1)}{2} n^{k/2-1} \\ &\quad + \mathbb{E}(R_k(S_n, a_{n+1})) + o(n^{k/2-1}). \end{aligned}$$

If k is odd, then

$$|\mathbb{E}(R_k(S_n, a_{n+1}))| \leq (M+1)^k (1 + \mathbb{E}(S_n^{k-3})) = O(n^{\frac{k-3}{2}}).$$

If k is even, then

$$(6) \quad |\mathbb{E}(T_k(S_n, a_{n+1}))| \leq (M+1)^k (1 + \mathbb{E}(S_n^{k-4})) = O(n^{\frac{k-4}{2}}),$$

whereas

$$\mathbb{E}(S_n^{k-3} a_{n+1}^3) = \mathbb{E}(S_n^{k-3} (D_3^n - \frac{\alpha_3^n S_n}{n})) = O(n^{\frac{k-3}{2}}),$$

which gives $\mathbb{E}(R_k(S_n, a_{n+1})) = O(n^{\frac{k-3}{2}})$. In any case, we have thus

$$(7) \quad \mathbb{E}(S_{n+1}^k) = \left(1 - \frac{k\alpha_1^n}{n}\right) \mathbb{E}(S_n^k) + C_{k-2} \frac{D_2 k(k-1)}{2} n^{k/2-1} + o(n^{k/2-1}).$$

If k is odd, we have $|\mathbb{E}(S_n^k)| \leq \mathbb{E}(S_n^{k-1} | S_n|) \leq Mn \mathbb{E}(S_n^{k-1}) = O(n^{1+(k-1)/2})$, so

$$(8) \quad \mathbb{E}(S_{n+1}^k) = \left(1 - \frac{k\alpha_1}{n}\right) \mathbb{E}(S_n^k) + C_{k-2} \frac{D_2 k(k-1)}{2} n^{k/2-1} + o(n^{k/2-1}),$$

and we apply Lemma 2.1. Otherwise, the sequence $\mathbb{E}(S_n^k)$ is non-negative and we apply Lemma 2.2.

In both cases, it comes that

$$\mathbb{E}(S_n^k) = \frac{2}{(1+2\alpha_1)k} \frac{D_2 C_{k-2} k(k-1)}{2} n^{k/2} + o(n^{k/2}) = C_k n^{k/2-1} + o(n^{k/2-1}),$$

which completes the inductive step. We have proved that for each $k \geq 0$, we have

$$\lim_{n \rightarrow +\infty} \mathbb{E}\left(\left(\frac{S_n}{\sqrt{n}}\right)^k\right) = \mathbb{E}(X^k),$$

where $X \sim \mathcal{N}(0, D')$. By the method of moments, we conclude that

$$\frac{S_n}{\sqrt{n}} \Longrightarrow \mathcal{N}(0, D').$$

3. PÓLYA'S URN WITH RANDOM COEFFICIENTS

There is a wealth of literature on Pólya's urns and their various extensions. For the properties we will find here, the most natural references are Freedman [5], Bagchi and Pal [1], Smythe [8], and Janson [6]. We work here with ballot boxes with two types of balls. The model considered is a balanced model, i.e. at each step, the number of balls in the urn increases by N , where N is a natural non-zero fixed integer.

Let's describe the dynamics: the urn initially contains a_0 white balls and b_0 black balls. The composition of the urn varies over time as follows: at time n , a ball is shot into the urn, if it is white, it is put back into the urn as well as B_n^1 white and $N - B_n^1$ black, where, conditionally to the above, B_n^1 follows the law μ_1 . If it is black, it is put back in the box along with B_n^2 white and $N - B_n^2$ black, where, conditionally to the above, B_n^2 follows the law μ_2 . We only impose $-1 \leq B_n^1 \leq N$ and $-1 \leq N - B_n^2 \leq N$: it is therefore possible to remove a ball, but only if it has been fired.

At time n , the urn therefore contains $a_0 + b_0 + nN$ balls. The number of white balls in the urn at the time n is $a_0 + \dots + a_n$, with

$$\mathcal{L}(a_{n+1}|a_0, \dots, a_n) = \frac{a_0 + S_n}{a_0 + b_0 + nN} \mu_1 + \frac{b_0 + (nN - S_n)}{a_0 + b_0 + nN} \mu_2.$$

Noting $m_{i,k} = \int_{\mathbb{R}} x^k d\mu_i$, we have then

$$\mathbb{E}(a_{n+1}^k | a_0, \dots, a_n) = \frac{a_0 + S_n}{a_0 + b_0 + nN} m_{1,k} + \frac{b_0 + (nN - S_n)}{a_0 + b_0 + nN} m_{2,k}.$$

Assumptions about the form of conditional expectations are easily verified, with $D_k = m_{2,k}$ and $\alpha_k = \frac{m_{2,k} - m_{1,k}}{N}$.

The first condition of validity of the Central Limit Theorem is given by $\alpha_1 > -\frac{1}{2}$, or $\rho = \frac{m_{1,1} - m_{2,1}}{N} < \frac{1}{2}$. In the classic formalism of Pólya's ballot boxes, the mean replacement matrix is

$$\begin{pmatrix} m_{1,1} & N - m_{1,1} \\ m_{2,1} & N - m_{2,1} \end{pmatrix},$$

whose eigenvalues are N and $m_{1,1} - m_{2,1}$. The condition $\rho < \frac{1}{2}$ thus found is indeed the classic condition on the eigenvalue quotient, which is classically denoted as the "small urns" case. The second condition $D > 0$ requires a little more calculations. Letting $r_i = \frac{m_{i,1}}{N}$ and $v_i = \text{Var} \frac{B_i^i}{N} = \frac{m_{i,2}^2}{N^2} - r_i^2$, we find $\frac{D}{2\alpha_1 + 1} = \frac{N^2(R+S)}{(2\alpha_1+1)(\alpha_1+1)^2}$ with

$$R = r_2 \alpha_1^2 (1 - r_1) \text{ and } S = (\alpha_1 + 1)(v_2(1 - r_1) + v_1 r_2).$$

Since $r_2 \geq 0$ and $r_1 \leq 1$, R and S are non-negative. It is then easy to see that only a few degenerate cases lead to $D = 0$: the cases $\mu_1 = \delta_N$ and $\mu_2 = \delta_0$, as well as the case where there is k such as $\mu_1 = \mu_2 = \delta_k$. Thus, in the two-ball model, we recovers Smythe's CLT [8] on "small urns". It is

interesting to note that we have an explicit form for the covariance of the normal limit distribution.

3.1. Example: a removed ball and added balls. In this example, at each step, one ball is removed from the urn, then black and white balls are added, b in total (with $b \geq 2$), the number of white balls following a distribution μ on $\{0, \dots, b\}$, with $\mu \notin \{\delta_0, \delta_b\}$. With the above notations, we have $N = b - 1$, $\mu_1 = \mu * \delta_{-1}$ and $\mu_2 = \mu$. If we denote by m the expectation of μ and by σ^2 its variance, we have then $m_{1,1} = m - 1$, $m_{1,2} = \sigma^2 + (m - 1)^2$, $m_{2,1} = m$, $m_{2,2} = \sigma^2 + m^2$. $\alpha_1 = \frac{m_{1,1} - m_{2,1}}{N} = \frac{1}{b-1}$ and $\alpha_2 = \frac{m_{2,2} - m_{1,2}}{N} = \frac{2m-1}{b-1}$. The quotient $\rho = \frac{m_{1,1} - m_{2,1}}{N} = -\frac{1}{b-1}$ checks $\rho < \frac{1}{2}$. We have $D_1 = m_{2,1} = m$, hence

$$\ell = \frac{D_1}{1 + \alpha_1} = \frac{D_1}{1 - \rho} = \frac{m}{1 + \frac{1}{b-1}} = \frac{m(b-1)}{b},$$

$D_2 = m_{2,2} = \sigma^2 + m^2$. We finally have

$$D = D_2 - \ell(\ell + \alpha_2) = \sigma^2 + \frac{m}{b}\left(1 - \frac{m}{b}\right).$$

So if B_n denotes the number of white balls at time n

$$\frac{B_n - \frac{m(b-1)}{b}n}{\sqrt{n}} \Longrightarrow \mathcal{N}\left(0, \frac{b-1}{b+1}\left(\sigma^2 + \frac{m}{b}\left(1 - \frac{m}{b}\right)\right)\right).$$

3.2. Example: Bernard Friedman urn. In this example, we add α balls of the drawn color and β balls of the other color. We have thus $N = \alpha + \beta$, $\mu_1 = \delta_\alpha$ and $\mu_2 = \delta_\beta$, which gives us $D_1 = m_{2,1} = \beta$, $D_2 = m_{2,2} = \beta^2$, $\alpha_1 = \frac{m_{2,1} - m_{1,1}}{N} = \frac{\beta - \alpha}{\beta + \alpha}$ and $\alpha_2 = \frac{m_{2,2} - m_{1,2}}{N} = \frac{\beta^2 - \alpha^2}{\beta + \alpha} = \beta - \alpha$. So $\ell = \frac{D_1}{\alpha_1 + 1} = \frac{\beta}{\frac{\beta}{\alpha + \beta}} = \frac{\alpha + \beta}{2}$, $D = D_2 - \ell(\ell + \alpha_2) = \beta^2 - \frac{\alpha + \beta}{2} \frac{3\beta - \alpha}{2} = \frac{(\alpha - \beta)^2}{4}$. So if $3\beta > \alpha$ and S_n is the number of white balls in the urn after n steps, we have

$$\frac{S_n - \frac{\alpha + \beta}{2}n}{\sqrt{n}} \Longrightarrow \mathcal{N}\left(0, \frac{(\alpha - \beta)^2(\alpha + \beta)}{4(3\beta - \alpha)}\right).$$

In particular, if we add only one ball of the other color, we have $\alpha = 0$ and $\beta = 1$ and

$$\frac{S_n - \frac{n}{2}}{\sqrt{n}} \Longrightarrow \mathcal{N}\left(0, \frac{1}{12}\right).$$

3.3. Returning to the dynamics of descents. The coincidence of this last equation with the CLT for downhill runs may lead to a new look to Equation (1), which can be rewritten as

$$\mathcal{L}(\mathbb{1}_{E_{n+1}} | \mathcal{F}_n) = \frac{1 + D_n}{1 + n} \delta_0 + \frac{n - D_n}{1 + n} \delta_1.$$

We can then see that the dynamics of variation of the descents is exactly the dynamics of the variation of the number of white balls in a Bernard Friedman urn starting from a white ball, and where we add to each step a

ball of a different color from the one drawn. One can e.g. imagine that there is time n $D_n + 1$ white balls and $n - D_n$ black balls in the urn. It is then obvious that the Central Limit Theorems obtained correspond. This is the eighth proof that was announced.

Although it is easy, this identification of the process of the number of descents with a Bernard Friedman urn seems new to us. We can extend the analogy even further. After one step, our Bernard Friedman urn necessarily contains a white ball and a black ball: this is the initial state of the urn considered by Diaconis and Fulton [4] to study the number N_n of strictly negative points visited by the Internal Diffusion–Limited Aggregation model on \mathbb{Z} after he has received n points. Remember that the Internal Diffusion–Limited Aggregation model is defined as follows: start at time 1 with with one point at the origin and say it forms the first block. Then, at each turn of the clock, start a walker from the origin and stop it as soon as it leaves the block, then enlarge the block by adding the stopped walker to it. At times n , the block is $\{-N_n, \dots, -N_n + (n - 1)\}$, where N_n denotes the number of points in the negative part. The gambler’s ruin theory says that one will exit from the left with probability $\frac{-N_n + (n-1) + 1}{(-N_n + (n-1) + 1) + (N_n + 1)} = \frac{n - N_n}{n + 1}$. So, we get

$$\mathcal{L}(N_{n+1} - N_n | N_1, \dots, N_n) = \frac{1 + D_n}{1 + n} \delta_0 + \frac{n - D_n}{1 + n} \delta_1$$

and the processes $(N_n)_{n \geq 1}$ and $(D_n)_{n \geq 1}$ have the same distribution.

4. BALLS ON A CIRCLE

We conclude with an example that does not allow itself to be reduced to a Pólya urn. In this model, b_0 black balls and w_0 white balls are initially placed in a circle, with $1 \leq w_0 < b_0$. The balls are arranged along the circle to respect a placement rule which is described as follows:

First, white and black balls are placed alternately. At a given moment, when there are only black balls to be placed, they placed next to each other (see Figure 1).

Then, 3 neighbouring balls are randomly drawn and removed; then a white balls and b black balls are inserted, in order to respect the placement rule.

To ensure that the rule can always be met, assumptions must be made about a and b . The assumptions $a \geq 1$ and $a + b \geq 4$ are natural so that there is always at least one white ball in the circle and the number of balls grows. We also need to add assumptions so that the blacks always stay ahead. We can consider alternatively

- (1) $b - a \geq 3$.
- (2) $b - a = 1$ and $w_0 + b_0$ is odd.

At the n^{th} step, the urn contains $(a + b - 3)n + w_0 + b_0$ balls, including S_n white.

Let's clarify the equations of the dynamics. The three balls that are removed can be spotted by the central ball. There are 4 possible configurations for this group of balls:

- a white ball surrounded by two black balls. This happens S_n times, the gain for the whites is $a - 1$; The advance of the blacks increases by $(b - a) - 1 \geq 0$.
- a black ball surrounded by two white. This happens $S_n - 1$ times, the gain for the whites is $a - 2$; the advance of the blacks increases by $(b - a) + 1 \geq 2$.
- two adjacent black balls and one white ball. It happens 2 times if the black balls represent strictly more than half of the balls, which, as we will see, is ensured by the assumption that we have taken, the gain for the whites is $a - 1$; the advance of the blacks increases by $(b - a) - 1 \geq 0$.
- three black balls otherwise, the gain for the whites is a . The advance of the blacks increases by $(b - a) - 3$. Under Hypothesis (1), $(b - a) \geq 3$, so the blacks are still ahead. Under Hypothesis (2), $b - a = 1$, so the advance is reduced by two units.

However, the existence of this group of three balls means that Black's lead is at least 2; but the dynamics equations mean that the number of balls is always an odd number, so Black's lead is at least 3; losing two units means Black's lead is at least one.

If we write $a_{n+1} = S_{n+1} - S_n$, we have

$$(9) \quad \mathcal{L}(a_{n+1} | a_1, \dots, a_n) = \frac{S_n - 1}{(a + b - 3)n + w_0 + b_0} \delta_{a-2} + \frac{S_n + 2}{(a + b - 3)n + w_0 + b_0} \delta_{a-1} + \left(1 - \frac{2S_n + 1}{(a + b - 3)n + w_0 + b_0} \right) \delta_a$$

Thus for any positive integer k , we have

$$\begin{aligned} & \mathbb{E}(a_{n+1}^k | a_1, \dots, a_n) \\ &= \frac{(a - 2)^k (S_n - 1)}{(a + b - 3)n + w_0 + b_0} + \frac{(a - 1)^k (S_n + 2)}{(a + b - 3)n + w_0 + b_0} \\ & \quad + a^k - \frac{(2S_n + 1)a^k}{(a + b - 3)n + w_0 + b_0} \\ &= \frac{(a - 2)^k + (a - 1)^k - 2a^k}{(a + b - 3)n + w_0 + b_0} S_n + a^k + \frac{-(a - 2)^k + 2(a - 1)^k - 1}{(a + b - 3)n + w_0 + b_0}. \end{aligned}$$

Conditional expectations are indeed of the required form.

In particular $D_1 = a$, $D_2 = a^2$, $\alpha_1 = \frac{3}{a+b-3}$, $\alpha_2 = \frac{6a-5}{a+b-3}$. It is obvious that $\alpha_1 > -\frac{1}{2}$, but the fact that $D > 0$ needs more care.

For example, with $w_0 = a = 2, b_0 = b = 3$, we have

$$\frac{S_n - \frac{4}{5}n}{\sqrt{n}} \implies \mathcal{N}\left(0, \frac{7}{50}\right).$$

When $a + b \geq 4$, with $b = a + 2$ or ($b = a + 1$ and $w_0 + b_0$ is even), the white never dominates, but equality of the two colors is possible (which is why (9) fails). Simulations seem to show that eventually, blacks definitely prevail with $\frac{S_n}{n(a+b-3)} \rightarrow \frac{a}{a+b} < \frac{1}{2}$ and that the stated Central Limit Theorem is always valid.

When the Central Limit Theorem is established, the convergence in probability of $\frac{S_n}{n(a+b-3)}$ to $\frac{a}{a+b}$ follows. The question of almost sure convergence is natural, and, presumably, more delicate.

REFERENCES

- [1] A. Bagchi and A. K. Pal. Asymptotic normality in the generalized Pólya-Eggenberger urn model, with an application to computer data structures. *SIAM J. Algebraic Discrete Methods*, 6(3):394–405, 1985.
- [2] Sourav Chatterjee and Persi Diaconis. A central limit theorem for a new statistic on permutations. *Indian J. Pure Appl. Math.*, 48(4):561–573, 2017.
- [3] F. N. David and D. E. Barton. *Combinatorial chance*. Hafner Publishing Co., New York, 1962.
- [4] P. Diaconis and W. Fulton. A growth model, a game, an algebra, Lagrange inversion, and characteristic classes. *Rend. Sem. Mat. Univ. Politec. Torino*, 49(1):95–119 (1993), 1991. Commutative algebra and algebraic geometry, II (Italian) (Turin, 1990).
- [5] David A. Freedman. Bernard Friedman’s urn. *Ann. Math. Statist*, 36:956–970, 1965.
- [6] Svante Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Process. Appl.*, 110(2):177–245, 2004.
- [7] Alperen Y. Özdemir. Martingales and descent statistics. *ArXiv preprint 1901.01719*, 2019.
- [8] R. T. Smythe. Central limit theorems for urn models. *Stochastic Process. Appl.*, 65(1):115–137, 1996.

UNIVERSITÉ DE LORRAINE, CNRS, IECL, F-54000 NANCY, FRANCE
 Email address: Olivier.Garet@univ-lorraine.fr