A central limit theorem for a card shuffling problem

Shane Chern

Universität Wien

chenxiaohang92@gmail.com

Oberseminar Zahlentheorie
University of Cologne
Dec 02, 2024

(Joint work with Lin Jiu and Italo Simonelli)

Mark Kac's *Carus Monographs*

Simple statistical observations of number-theoretic or combinatorial objects "**are often the starting point of rich and fruitful theories.**"

✼

Erdős–Kac Theorem: *Let $\omega(n)$ be the number of distinct prime factors of n. If we randomly choose n from $1 \le n \le N$, then the probability distribution of*

$$\frac{\omega(n) - \log\log n}{\sqrt{\log\log n}}$$

*tends to the standard normal distribution as N becomes large enough.*

Card shuffling

- **(i).** Given a positive integer n, consider a random permutation τ of the set $[n] := \{1, 2, \ldots, n\}$;

  (ii). In τ, we look for sequences of consecutive integers that appear in adjacent positions, and a maximal such a sequence is called a <u>block</u>.

  Initial set: $(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$;

  Permuting: $(1, 7, 5, 6, 8, 10, 9, 2, 3, 4)$.

- **(i).** Each block in τ is merged into its first integer.

  (ii). After all the merges the elements of this new set are relabeled from 1 to the current number of elements.

  (iii). Permute this new set.

  Merging: $(1, 7, 5, 8, 10, 9, 2)$;

  Relabeling: $(1, 2, 3, 4, 5, 6, 7)$;

  Permuting: $(7, 5, 4, 6, 3, 2, 1)$ — NEW shuffling!

- We continue this merging and permuting until only one integer is left.

Card shuffling

- **The quantity of interest is X_n, the number of permutations needed for the process to end.**

- Rao et al. (2016):

Card shuffling
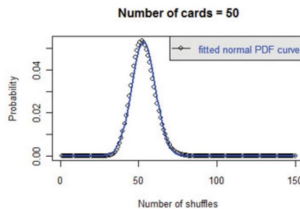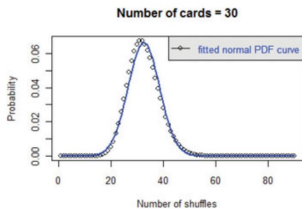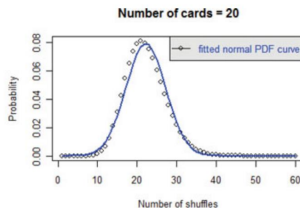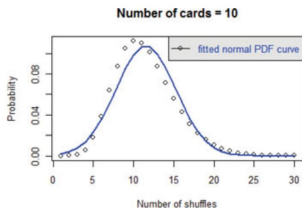
Theorem (C.–Jiu–Simonelli, 2023)

Let Z denote a standard normal random variable, i.e., $Z \sim \mathcal{N}(0,1)$. Then as $n \to \infty$,

$$\frac{X_n - n}{\sqrt{n}} \xrightarrow{w} Z.$$

Card shuffling

Chebyshev's method of moments (1887):

- Z is a standard normal random variable, so its moments are given by

$$\mathbf{E}\big[Z^m\big] = \begin{cases} (m-1)!!, & \text{if } m \text{ is even,} \\ 0, & \text{if } m \text{ is odd.} \end{cases}$$

- $Z_1, Z_2, \ldots, Z_n, \ldots$ are a series of random variables such that

$$\lim_{n \to \infty} \mathbf{E}\big[Z_n^m\big] = \mathbf{E}\big[Z^m\big].$$

Then Z_n converges to Z in distribution.

Card shuffling

Goal: Evaluate

$$\mathbf{E}\left[\left(\frac{X_n - n}{\sqrt{n}}\right)^m\right].$$

- $\mu_n = \mathbf{E}[X_n] \sim n$;
- $\mathbf{Var}[X_n] = \mathbf{E}\left[(X_n - \mu_n)^2\right] \sim n$;

Goal': Evaluate

$$\mathbf{E}\left[(X_n - \mu_n)^m\right].$$

Probability sequence

- Let Y_n be the number of blocks in a random permutation of $[n]$;

Then for $n \geq 2$,

$$X_n = \begin{cases} 1, & \text{with probability } \mathbf{P}(Y_n = 1), \\ 1 + X_k, & \text{with probability } \mathbf{P}(Y_n = k) \text{ for } 2 \leq k \leq n. \end{cases}$$

For $1 \leq k \leq n$,

$$\mathbf{P}(Y_n = k) = \frac{A(n, k)}{n!},$$

where $A(n, k)$ counts the number of permutations of n with k blocks.

Probability sequence

$$A(n, k) = \binom{n-1}{k-1} A(k, k).$$

- $[n]$ is split into k components $B_1, \ldots, B_k$:

$$\underbrace{1 \ 2}_{B_1} \ | \ \underbrace{3 \ 4 \ 5}_{B_2} \ | \ \underbrace{6 \ 7 \ 8}_{B_3} \qquad \Rightarrow \qquad \binom{n-1}{k-1}$$

- The components $B_1, \ldots, B_k$ **cannot** be merged:

$$B_1 \ B_3 \ B_2 \qquad \Rightarrow \qquad A(k, k)$$

$$(B_3 \ B_1 \ B_2 \qquad = \qquad \underline{678} \ \underline{12345} \qquad \textbf{✗})$$

Probability sequence

How to evaluate $A(k, k)$**?**

- Remove k to get a permutation of size $k - 1$;
- In this permutation of size $k - 1$, the number of blocks is either $k - 1$ or $k - 2$.

�֎

$$A(k, k) = (k - 1) \cdot A(k - 1, k - 1) + (k - 2) \cdot A(k - 2, k - 2).$$

- 53142 counted by $A(k - 1, k - 1) = A(5, 5)$:

$$☑ \; 5 \; ☒ \; 3 \; ☑ \; 1 \; ☑ \; 4 \; ☑ \; 2 \; ☑ \qquad \Rightarrow \qquad (k - 1) \cdot A(k - 1, k - 1)$$

- 52314 counted by $A(k - 1, k - 2) = A(5, 4)$:

$$☒ \; 5 \; ☒ \; 2 \; ☑ \; 3 \; ☒ \; 1 \; ☒ \; 4 \; ☒ \qquad \Rightarrow \qquad 1 \cdot A(k - 1, k - 2)$$

$$= 1 \cdot \binom{k - 2}{k - 3} A(k - 2, k - 2)$$

$$= (k - 2) \cdot A(k - 2, k - 2)$$

Probability sequence

- $A(1,1) = 1$;
- $A(2,2) = 1$ (i.e., the permutation 21);
- For $k \geq 3$,

$$A(k,k) = (k-1)A(k-1,k-1) + (k-2)A(k-2,k-2).$$

```
0 1 3 6 2 7
        13    THE ON–LINE ENCYCLOPEDIA
   OE  20
23 IS  12    OF INTEGER SEQUENCES ®
10 22 11 21
```

founded in 1964 by N. J. A. Sloane

| 1, 1, 3, 11, 53, 309, 2119, 16687, 148329, 1468457 | Search | Hints |

(Greetings from The On-Line Encyclopedia of Integer Sequences!)

Search: **seq:1,1,3,11,53,309,2119,16687,148329,1468457**

Displaying 1-1 of 1 result found.   page 1

Sort: relevance | references | number | modified | created   Format: long | short | data

A000255   a(n) = n*a(n-1) + (n-1)*a(n-2), a(0) = 1, a(1) = 1.   +30
               (Formerly M2905 N1166)   101

1, 1, 3, 11, 53, 309, 2119, 16687, 148329, 1468457, 16019531, 190899411, 2467007773,
34361893981, 513137616783, 8178130767479, 138547156531409, 2486151753313617, 47106033220679059,
939765362752547227, 19690321886243846661, 432292066866171724421 (list; graph; refs; listen; history; text;
internal format)

Probability sequence

$$\mathbf{P}(Y_n = k) = \frac{A(n, k)}{n!},$$

where the sequence $A(n, k)$ is defined for $1 \leq k \leq n$ by

$$A(n, k) := \binom{n-1}{k-1} A(k-1),$$

with the sequence $A(n)$ being recursively given by

$$A(n) = nA(n-1) + (n-1)A(n-2),$$

in combination with initial values $A(0) = A(1) = 1$.

Recurrences

Recurrence for $\mu_n = \mathbf{E}[X_n]$:

$$\mu_n = \mathbf{P}(Y_n = 1) + \sum_{k=2}^{n} \mathbf{P}(Y_n = k)\mathbf{E}\big[1 + X_k\big]$$

$$= \frac{A(n,1)}{n!} + \sum_{k=2}^{n} \frac{A(n,k)}{n!}(1 + \mu_k)$$

$$= \sum_{k=1}^{n} \frac{A(n,k)}{n!} + \sum_{k=2}^{n} \frac{A(n,k)}{n!}\mu_k$$

$$= 1 + \sum_{k=2}^{n} \frac{A(n,k)}{n!}\mu_k.$$

We adopt the convention that $X_1 = 0$, a definite constant, so that the mean value $\mu_1 = \mathbf{E}[X_1] = 0$, as there is no more shuffling needed. Then the above recurrence for μ_n becomes ($\mu_1 = 0$ and $\mu_2 = 2$):

$$\mu_n = 1 + \sum_{k=1}^{n} \frac{A(n,k)}{n!}\mu_k.$$

Recurrences

In general, if we assume that $p(x)$ is an arbitrary polynomial in x, then

$$\mathbf{E}\big[p(X_n)\big] = \sum_{k=1}^{n} \mathbf{P}(Y_n = k)\mathbf{E}\big[p(1 + X_k)\big] = \sum_{k=1}^{n} \frac{A(n,k)}{n!} \cdot \mathbf{E}\big[p(1 + X_k)\big].$$

Thus,

$$\begin{aligned}
\mathbf{E}\left[(X_n - \mu_n)^m\right] &= \sum_{k=1}^{n} \frac{A(n,k)}{n!} \cdot \mathbf{E}\left[\left(1 + X_k - \mu_n\right)^m\right] \\
&= \sum_{k=1}^{n} \frac{A(n,k)}{n!} \cdot \mathbf{E}\left[\left((X_k - \mu_k) + (1 + \mu_k - \mu_n)\right)^m\right],
\end{aligned}$$

so that

$$\left(1 - \frac{A(n,n)}{n!}\right)\mathbf{E}\big[(X_n - \mu_n)^m\big] = \sum_{k=1}^{n-1} \frac{A(n,k)}{n!}\,\mathbf{E}\big[(X_k - \mu_k)^m\big] + \sum_{\ell=1}^{m} \binom{m}{\ell} I_\ell^{(m)},$$

where for each ℓ with $1 \le \ell \le m$,

$$I_\ell^{(m)} := \sum_{k=1}^{n} \frac{A(n,k)}{n!} \cdot \mathbf{E}\big[(X_k - \mu_k)^{m-\ell}\big]\left(1 + \mu_k - \mu_n\right)^\ell.$$

Recurrences

Let $\{\lambda_n\}_{n \geq 1}$ be a complex sequence such that $\lambda_n \sim Mn^L$ as $n \to \infty$, wherein L is a fixed nonnegative integer, and M is a fixed complex number, which, in addition, is nonzero when L is nonzero. Write

$$\delta_n := \lambda_n - Mn^L.$$

We define a complex sequence $\{\xi_n\}_{n \geq 1}$ with given initial values $\xi_1, \ldots, \xi_{n_0}$ for a certain $n_0 \geq 2$ by the recurrence

$$\left(1 - \frac{A(n,n)}{n!}\right)\xi_n = \lambda_n + \sum_{k=1}^{n-1} \frac{A(n,k)}{n!}\xi_k,$$

for every $n > n_0$.

Recurrences

Theorem

As $n \to \infty$,

$$\xi_n \sim \frac{M}{L+1} \, n^{L+1},$$

where the asymptotic relation depends only on L and M. More precisely, letting

$$\eta_n := \xi_n - \frac{M}{L+1} \, n^{L+1},$$

there exists a positive constant C, depending only on L and M, such that for all $n \geq 1$,

$$|\eta_n| < C \sum_{j=1}^{n} \left(|\delta_j| + j^{L-1} \right).$$

Recurrences

Basic idea:

- Define the partial sums $S_n(t)$ with $1 \leq t \leq n$:

$$S_n(t) := \sum_{k=1}^{t} \frac{A(n,k)}{n!}.$$

- Rewrite the recurrence $\left(1 - \frac{A(n,n)}{n!}\right) \xi_n = \lambda_n + \sum_{k=1}^{n-1} \frac{A(n,k)}{n!} \xi_k$ as

$$S_n(n-1) \cdot \xi_n = \lambda_n + \Sigma_1 + \Sigma_2,$$

where

$$\Sigma_1 := \sum_{k=1}^{n-1} \frac{A(n,k)}{n!} \frac{M}{L+1} k^{L+1},$$

$$\Sigma_2 := \sum_{k=1}^{n-1} \frac{A(n,k)}{n!} \eta_k.$$

Recurrences

Abel summation formula: Let $\{u_n\}_{n \geq 1}$ and $\{v_n\}_{n \geq 1}$ be sequences of complex numbers. Then for any $N \geq 1$,

$$\sum_{n=1}^{N} u_n v_n = U(N) v_{N+1} + \sum_{n=1}^{N} U(n)\big(v_n - v_{n+1}\big),$$

where $U(n) := \sum_{k=1}^{n} u_k$.

Recurrences

- Apply the Abel summation formula to Σ_1:

$$
\begin{aligned}
\Sigma_1 &= \sum_{k=1}^{n-1} \frac{A(n,k)}{n!} \frac{M}{L+1} k^{L+1} \\
&= S_n(n-1) \cdot \frac{M}{L+1} n^{L+1} + \sum_{k=1}^{n-1} S_n(k) \cdot \frac{M}{L+1} \big(k^{L+1} - (k+1)^{L+1} \big) \\
&= S_n(n-1) \cdot \frac{M}{L+1} n^{L+1} - \sum_{k=1}^{n-1} S_n(k) \cdot \big(Mk^L + O(k^{L-1}) \big) \\
&= S_n(n-1) \cdot \frac{M}{L+1} n^{L+1} - Mn^L + O(n^{L-1}).
\end{aligned}
$$

- Substitute back to the recurrence for ξ_n:

$$
S_n(n-1) \cdot \eta_n = \delta_n + \sum_{k=1}^{n-1} \frac{A(n,k)}{n!} \eta_k + O(n^{L-1}).
$$

Recurrences

Further applying the Abel summation formula to Σ_2 gives an elaboration on the error term η_n when δ_n is efficiently bounded:

Theorem

We further have

1. *In the case where $L = 0$, if we further require that $\delta_n = O(n^{-1})$ as $n \to \infty$, then*

$$\left|\eta_n - \eta_{n-1}\right| = O(n^{-1}).$$

2. *In the case where $L \geq 1$, if we further require that $\delta_n = O(n^{L-1} \log n)$ as $n \to \infty$, then*

$$\left|\eta_n - \eta_{n-1}\right| = O(n^{L-1} \log n).$$

The above asymptotic relations depend only on L and M.

Recurrences

- Recall that for $\mu_n = \mathbf{E}[X_n]$,

$$\left(1 - \frac{A(n,n)}{n!}\right)\mu_n = 1 + \sum_{k=1}^{n-1} \frac{A(n,k)}{n!}\mu_k.$$

By choosing $\lambda_n = 1$ in the generic recurrence, it is immediate that

$$\mu_n \sim n.$$

- Let $\mathcal{H}_n := \sum_{k=1}^{n} \frac{1}{k}$ denote the n-th harmonic number.

$$S_n(n-1) \cdot \mu_n = 1 + \Sigma_1' + \Sigma_2' + \Sigma_3',$$

where

$$\Sigma_1' := \sum_{k=1}^{n-1} \frac{A(n,k)}{n!}k, \qquad \Sigma_2' := \sum_{k=1}^{n-1} \frac{A(n,k)}{n!}\mathcal{H}_{k-1}, \qquad \Sigma_3' := \sum_{k=1}^{n-1} \frac{A(n,k)}{n!}\varepsilon_k.$$

Recurrences

Theorem

We have

$$\mu_n = n + \mathcal{H}_{n-1} + \varepsilon_n,$$

where the limit $\lim_{n \to \infty} \varepsilon_n$ *exists. In particular, for* $n \geq 2$,

$$0 < \varepsilon_n - \varepsilon_{n+1} < \frac{1}{n^2}.$$

Recurrences

Why do we care about an explicit formula of μ_n?

- To understand the behavior of μ_n in a more accurate way.

Open Problem. Find an explicit formula for $\lim_{n\to\infty} \varepsilon_n$.

Expectation:

$$\lim_{n\to\infty} \varepsilon_n = \square \cdot \zeta(2) + \square \cdot \zeta(3) + \cdots.$$

Recurrences

Why do we care about an explicit formula of μ_n?

- We can effectively bound $\mu_n - \mu_k$ in terms of $n - k$:

$$n - k \leq \mu_n - \mu_k \leq (n-k)\left(1 + \tfrac{1}{k}\right).$$

- We can prove that

$$\sum_{k=1}^{n} \frac{A(n,k)}{n!}\left(\mu_n - \mu_k\right)^{\ell} = B_\ell + O(n^{-1}),$$

where B_ℓ is the ℓ-th Bell number defined by the exponential generating function

$$\sum_{\ell=0}^{\infty} B_\ell \frac{x^\ell}{\ell!} := e^{e^x - 1}.$$

Central moments

Theorem

For every $m \geq 2$, we have, as $n \to +\infty$,

$$\mathbf{E}\big[(X_n - \mu_n)^m\big] = \begin{cases} (2M-1)!! \cdot n^M + O(n^{M-1} \log n), & \text{if } m = 2M, \\ \frac{2}{3} M(2M+1)!! \cdot n^M + O(n^{M-1} \log n), & \text{if } m = 2M+1. \end{cases}$$

In particular, if we define

$$\varepsilon_n^{(m)} := \mathbf{E}\big[(X_n - \mu_n)^m\big] - \begin{cases} (2M-1)!! \cdot n^M, & \text{if } m = 2M, \\ \frac{2}{3} M(2M+1)!! \cdot n^M, & \text{if } m = 2M+1, \end{cases}$$

then

$$\big|\varepsilon_n^{(m)} - \varepsilon_{n-1}^{(m)}\big| = \begin{cases} O(n^{-1}), & \text{if } m = 2 \text{ or } 3, \\ O(n^{\lfloor \frac{m}{2} \rfloor - 2} \log n), & \text{if } m \geq 4. \end{cases}$$

Central moments

Recall that

$$\left(1 - \frac{A(n,n)}{n!}\right)\mathbf{E}\big[(X_n - \mu_n)^m\big] = \sum_{k=1}^{n-1}\frac{A(n,k)}{n!}\mathbf{E}\big[(X_k - \mu_k)^m\big] + \sum_{\ell=1}^{m}\binom{m}{\ell}I_\ell^{(m)},$$

where for each ℓ with $1 \le \ell \le m$,

$$I_\ell^{(m)} := \sum_{k=1}^{n}\frac{A(n,k)}{n!}\cdot\mathbf{E}\big[(X_k - \mu_k)^{m-\ell}\big]\big(1 + \mu_k - \mu_n\big)^\ell.$$

Basic idea:

- Induction on m;
- Split each $\mathbf{E}\big[(X_k - \mu_k)^{m-\ell}\big]$ as a main term $? \cdot k^?$ (by an inductive argument) and an error term $\varepsilon_k^{(m-\ell)}$.

Central moments

Variance $\mathbf{Var}[X_n] = \mathbf{E}\big[(X_n - \mu_n)^2\big]$:

$$\left(1 - \frac{A(n,n)}{n!}\right)\mathbf{Var}[X_n] = \sum_{k=1}^{n-1} \frac{A(n,k)}{n!}\,\mathbf{Var}[X_k] + I_1^{(2)} + I_2^{(2)},$$

where

$$I_1^{(2)} = \sum_{k=1}^{n} \frac{A(n,k)}{n!} \cdot \mathbf{E}\big[X_k - \mu_k\big]\big(1 + \mu_k - \mu_n\big) = 0,$$

$$I_2^{(2)} = \sum_{k=1}^{n} \frac{A(n,k)}{n!} \cdot \big(1 + \mu_k - \mu_n\big)^2 = 1 + O(n^{-1}).$$

Thus,

$$\left(1 - \frac{A(n,n)}{n!}\right)\mathbf{Var}[X_n] = \sum_{k=1}^{n-1} \frac{A(n,k)}{n!}\,\mathbf{Var}[X_k] + 1 + O(n^{-1}),$$

so that

$$\mathbf{Var}[X_n] \sim n.$$

Central moments

Third central moment $\mathrm{Var}[X_n] = \mathbf{E}\big[(X_n - \mu_n)^3\big]$:

$$\left(1 - \frac{A(n,n)}{n!}\right)\mathbf{E}\big[(X_n - \mu_n)^3\big] = \sum_{k=1}^{n-1}\frac{A(n,k)}{n!}\,\mathbf{E}\big[(X_k - \mu_k)^3\big] + 3I_1^{(3)} + 3I_2^{(3)} + I_3^{(3)},$$

where

$$I_1^{(3)} = \sum_{k=1}^{n}\frac{A(n,k)}{n!}\cdot\mathbf{E}\big[(X_k - \mu_k)^2\big]\big(1 + \mu_k - \mu_n\big),$$

$$I_2^{(3)} = \sum_{k=1}^{n}\frac{A(n,k)}{n!}\cdot\mathbf{E}\big[X_k - \mu_k\big]\big(1 + \mu_k - \mu_n\big)^2 = 0,$$

$$I_3^{(3)} = \sum_{k=1}^{n}\frac{A(n,k)}{n!}\cdot\big(1 + \mu_k - \mu_n\big)^3 = -1 + O(n^{-1}).$$

Central moments

Recalling that $\mathbf{E}\big[(X_k - \mu_k)^2\big] = k + O(\log k)$, we split $I_1^{(3)}$ as

$$I_1^{(3)} = I_{1,1}^{(3)} + I_{1,2}^{(3)},$$

where

$$I_{1,1}^{(3)} := \sum_{k=1}^{n} \frac{A(n,k)}{n!} \cdot k\big(1 + \mu_k - \mu_n\big),$$

$$I_{1,2}^{(3)} := \sum_{k=1}^{n} \frac{A(n,k)}{n!} \cdot \varepsilon_k^{(2)}\big(1 + \mu_k - \mu_n\big).$$

Then,

$$I_{1,1}^{(3)} = 1 + O(n^{-1}).$$

In general,

$$\sum_{k=1}^{n} \frac{A(n,k)}{n!} \cdot k^L\big(1 + \mu_k - \mu_n\big) = Ln^{L-1} + O(n^{L-2}).$$

Central moments

Since $\varepsilon_k^{(2)} = O(\log k)$, if we trivially bound $I_{1,2}^{(3)}$, we only have

$$I_{1,2}^{(3)} \ll \log n \sum_{k=1}^{n} \frac{A(n,k)}{n!} |1 + \mu_k - \mu_n| \ll \log n.$$

Hence, we shall invoke the bound $\left| \varepsilon_n^{(2)} - \varepsilon_{n-1}^{(2)} \right| = O(n^{-1})$ and apply the Abel summation formula (in a **very technical** way) to get

$$I_{1,2}^{(3)} = O(n^{-1}).$$

Thus,

$$\left(1 - \frac{A(n,n)}{n!} \right) \mathbf{E}\left[(X_n - \mu_n)^3 \right] = \sum_{k=1}^{n-1} \frac{A(n,k)}{n!} \mathbf{E}\left[(X_k - \mu_k)^3 \right] + 2 + O(n^{-1}),$$

so that

$$\mathbf{E}\left[(X_n - \mu_n)^3 \right] \sim 2n.$$

Central moments

Even-order central moment $\mathbf{Var}[X_n] = \mathbf{E}\big[(X_n - \mu_n)^{2M}\big]$:

$$\left(1 - \frac{A(n,n)}{n!}\right) \mathbf{E}\big[(X_n - \mu_n)^{2M}\big] = \sum_{k=1}^{n-1} \frac{A(n,k)}{n!} \mathbf{E}\big[(X_k - \mu_k)^{2M}\big] + \sum_{\ell=1}^{2M} \binom{2M}{\ell} I_\ell^{(2M)},$$

where

$$I_1^{(2M)} = \sum_{k=1}^{n} \frac{A(n,k)}{n!} \cdot \mathbf{E}\big[(X_k - \mu_k)^{2M-1}\big]\big(1 + \mu_k - \mu_n\big) = O(n^{M-2}\log n),$$

$$I_2^{(2M)} = \sum_{k=1}^{n} \frac{A(n,k)}{n!} \cdot \mathbf{E}\big[(X_k - \mu_k)^{2M-2}\big]\big(1 + \mu_k - \mu_n\big)^2 = (2M-3)!! \cdot n^{M-1} + O(n^{M-2}\log n),$$

$$\vdots$$

Thus,

$$\left(1 - \frac{A(n,n)}{n!}\right) \mathbf{E}\big[(X_n - \mu_n)^{2M}\big] = \sum_{k=1}^{n-1} \frac{A(n,k)}{n!} \mathbf{E}\big[(X_k - \mu_k)^{2M}\big]$$
$$+ \binom{2M}{2}(2M-3)!! \cdot n^{M-1} + O(n^{M-2}\log n),$$

so that

$$\mathbf{E}\big[(X_n - \mu_n)^{2M}\big] \sim (2M-1)!! \cdot n^M.$$

Central moments

Odd-order central moment $\mathrm{Var}[X_n] = \mathbf{E}\big[(X_n - \mu_n)^{2M+1}\big]$**:**

$$\left(1 - \frac{A(n,n)}{n!}\right) \mathbf{E}\big[(X_n - \mu_n)^{2M+1}\big] = \sum_{k=1}^{n-1} \frac{A(n,k)}{n!} \mathbf{E}\big[(X_k - \mu_k)^{2M+1}\big] + \sum_{\ell=1}^{2M+1} \binom{2M+1}{\ell} I_\ell^{(2M+1)},$$

where

$$I_1^{(2M+1)} = \sum_{k=1}^{n} \frac{A(n,k)}{n!} \cdot \mathbf{E}\big[(X_k - \mu_k)^{2M}\big]\big(1 + \mu_k - \mu_n\big)$$

$$= M(2M-1)!! \cdot n^{M-1} + O(n^{M-2}\log n), \quad \text{[DIFFICULT! Abel summation used.]}$$

$$I_2^{(2M+1)} = \sum_{k=1}^{n} \frac{A(n,k)}{n!} \cdot \mathbf{E}\big[(X_k - \mu_k)^{2M-1}\big]\big(1 + \mu_k - \mu_n\big)^2$$

$$= \tfrac{2}{3}(M-1)(2M-1)!! \cdot n^{M-1} + O(n^{M-2}\log n),$$

$$I_3^{(2M+1)} = \sum_{k=1}^{n} \frac{A(n,k)}{n!} \cdot \mathbf{E}\big[(X_k - \mu_k)^{2M-2}\big]\big(1 + \mu_k - \mu_n\big)^3,$$

$$= -(2M-3)!! \cdot n^{M-1} + O(n^{M-2}\log n),$$

$\vdots$

Central moments

Thus,

$$\left(1 - \frac{A(n,n)}{n!}\right) \mathbf{E}\big[(X_n - \mu_n)^{2M+1}\big]$$

$$= \sum_{k=1}^{n-1} \frac{A(n,k)}{n!} \mathbf{E}\big[(X_k - \mu_k)^{2M+1}\big]$$

$$+ \left(\binom{2M+1}{1} M(2M-1)!! + \binom{2M+1}{2} \tfrac{2}{3}(M-1)(2M-1)!! - \binom{2M+1}{3}(2M-3)!! \right) \cdot n^{M-1}$$

$$+ O(n^{M-2} \log n),$$

so that

$$\mathbf{E}\big[(X_n - \mu_n)^{2M+1}\big] \sim \tfrac{2}{3} M(2M+1)!! \cdot n^M.$$

Central moments

Let
$$Z_n := \frac{X_n - \mu_n}{\sqrt{\mathbf{Var}[X_n]}}.$$

Then $\mathbf{E}[Z_n] = 0$, $\mathbf{Var}[Z_n] = 1$, and more importantly, as $n \to \infty$,

$$\mathbf{E}\big[Z_n^m\big] = \frac{\mathbf{E}\big[(X_n - \mu_n)^m\big]}{\mathbf{Var}[X_n]^{m/2}} \to \begin{cases} (m-1)!!, & \text{if } m \text{ is even,} \\ 0, & \text{if } m \text{ is odd,} \end{cases}$$

matching with the moments of the standard normal distribution Z. So Chebyshev's method of moments asserts the weak convergence of $Z_n \Rightarrow Z$.

Further thoughts

(Ambitious) Goal: Find a criterion for the merging rules that result in a certain central limit theorem.

In principle, we need to look at recurrences of the form

$$\mathbf{E}\big[p(X_n)\big] = \sum_{k=1}^{n} \alpha(n, k) \cdot \mathbf{E}\big[p(1 + X_k)\big],$$

which holds for any polynomial $p(x)$. Here each $\alpha(n, k)$ is a probability evaluation so that $\sum_{k=1}^{n} \alpha(n, k) = 1$. We wish to find a way to characterize these $\alpha(n, k)$ so as to ensure that the moments $\mathbf{E}[Z_n^m]$, with Z_n the normalization of X_n, match with those of the standard normal distribution.

S. Chern, L. Jiu, and I. Simonelli, A central limit theorem for a card shuffling problem, arXiv:2309.08841.

Thank You!