Statistics and Probability Letters 197 (2023) 109804

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Berry-Essén theorem for random determinants

Charles-Philippe Diez, Ciprian A. Tudor *

CNRS, Université de Lille, Laboratoire Paul Painlevé UMR 8524, F-59655 Villeneuve d'Ascq, France

A R T I C L E I N F O

Article history:
Received 15 January 2022
Received in revised form 20 October 2022
Accepted 14 February 2023
Available online 21 February 2023

MSC:
60B20
60F05
60H07
60G22

Keywords:
Random matrix
Random determinant
Multiple stochastic integrals
Malliavin derivative

A B S T R A C T

Let $\mathcal{X}_n$ be a $n \times n$ random matrix with independent Gaussian entries. Then it is well-known that $\log(|\det(\mathcal{X}_n|))$ satisfies a Central Limit Theorem. Our purpose is to evaluate the 1-Wasserstein distance for this limit theorem via the techniques of the Stein–Malliavin calculus.

1. Introduction

Since several decades, the study of the distribution of the determinant of a random matrix constitutes a problem of interest for many researchers in probability theory. The first studies in this direction concerned the moments of a random determinant, see among others Forsythe and Tukey (1952), Nyquist et al. (1954), Prékopa (1967) or Turań (1955). Later, a further step in the analysis of the determinants of random matrices was to show that it satisfies a Central Limit Theorem (CLT in the sequel). Let $\mathcal{X}_n$ be a $n \times n$ random matrix and assume that its entries are standard Gaussian independent random variables. Then it has been shown in Goodman (1963) that, as $n \to \infty$,

$$\frac{\log(|\det \mathcal{X}_n|) - \frac{1}{2}\log(n-1)!}{\sqrt{\frac{1}{2}\log n}} \xrightarrow{(d)} N(0,1) \tag{1}$$

where $\to^{(d)}$ stands for the convergence in distribution and $N(0,1)$ denotes the standard normal distribution. The extensions of the above result to a more general situation (when the entries of the random matrix $\mathcal{X}_n$ are not Gaussian or not independent) are due, among others, to Bourgade and Mody (2019), Nguyen and Vu (2014), Rempala and Wesolowski (2005) or Bao et al. (2015). The rate of convergence for the limit theorem (1) has been studied in Nguyen and Vu (2014). It has been shown that the Kolmogorov distance between the left-hand side of (1) and the standard Gaussian law is of order less than $\log^{-\frac{1}{3}+o(1)} n$.

* Corresponding author.
E-mail addresses: diez@univ-lille.fr (C.-P. Diez), ciprian.tudor@univ-lille.fr (C.A. Tudor).

https://doi.org/10.1016/j.spl.2023.109804
0167-7152/

Our purpose is to estimate the 1-Wasserstein distance corresponding to the CLT (1) via the techniques of the Stein–Malliavin calculus. We will assume that the entries of the starting matrix $\mathcal{X}_n$ have Gaussian distribution but they are not necessarily mutually independent. As noticed in Goodman (1963) or Rempala and Wesolowski (2005), $\det(\mathcal{X}_n^2)$ can be expressed as a product of sums of independent chi-square random variables. We will then focus of the analysis of such a product of sums, by improving the results in Arnold and Villasenor (1998) or Rempala and Wesolowski (2005, 2002). We decompose it into a negligible part (whose $L^1(\Omega)$-norm can be relatively easily estimated) and a dominant part which can be decomposed in Wiener chaos and whose rate of convergence to the normal distribution can be analyzed by the tools of Malliavin calculus. We obtain a rate of convergence of order $\log^{-\frac{1}{2}} n$ which improves the result for the Kolmogorov distance obtained in Nguyen and Vu (2014).

We organized our work as follows. Section 2 is devoted to the notation and the description of the general context of this work. In Section 3 we estimate the negligible part of a product of sums of chi-square-distributed random variables. In Section 4 we apply the tools of the Malliavin calculus in order to evaluate the distance between the dominant part of such a product of sums and we deduce its rate of convergence under the 1-Wasserstein distance. Section 4 contains the application to random determinants while (the Appendix) contains some elements of the Malliavin calculus needed in this work.

2. Products of sums of chi-square random variables

This part is consecrated to the presentation of the problem and to some notation.

2.1. The settings

Let H be a real and separable Hilbert space and let $(W(h), h \in H)$ be an isonormal Gaussian process (see the Appendix). Let $\left(X_{k,l}, k \geq 1, l = 1, \ldots, k\right)$ be a family of independent real-valued random variables with standard normal distribution, i.e. $X_{k,l} \sim N(0,1)$ for $k \geq 1$ and $l = 1, \ldots, k$. We will assume

$$X_{k,l} = W(h_{k,l}) = I_1(h_{k,l}) \text{ for } k \geq 1, l = 1, \ldots, k$$

where $(h_{k,l}, k \geq 1, l = 1, \ldots, k)$ are orthogonal elements in H and I_q denotes the multiple integral of order $q \geq 1$ with respect to W. Denote, for $k \geq 1$,

$$S_k = X_{k,1}^2 + \cdots + X_{k,k}^2 \tag{2}$$

and

$$C_k = \frac{S_k}{k} \tag{3}$$

We will need the following lemma (see e.g. Lemma 1 in Rempala and Wesolowski (2005), which is borrowed from Lee (1990)).

Lemma 1. *For $k \geq 1$, let C_k be given by (3). Then for every $k \geq 1$ and $p \geq 2$*

$$\mathbf{E}\,|C_k - 1|^p \leq D_p k^{-\frac{p}{2}}$$

with $D_p > 0$ not depending on k.

For $N \geq 1$, we define

$$T_N = \frac{1}{\sqrt{2\log(N)}}\left(\sum_{k=1}^N \log(C_k) + \log(N)\right) = \frac{1}{\sqrt{2\log(N)}}\left(\log\left[\prod_{k=1}^N S_k\right) - \log(N-1)!\right]. \tag{4}$$

We know that (see e.g. Rempala and Wesolowski (2005))

$$T_N \overset{(d)}{\to} N(0,1) \text{ as } N \to \infty. \tag{5}$$

Our purpose is to find the rate of convergence in the above limit theorem. The idea is to write T_N as the sum of a negligible part whose $L^1(\Omega)$ norm can be evaluated by standard calculation and a dominant part which fits with the Stein–Malliavin calculus. More precisely, we will decompose T_N as follows

$$\begin{aligned} T_N &= \frac{1}{\sqrt{2\log(N)}}\left[\log(N) + \sum_{k=1}^N \log(C_k) 1_{|C_k-1|\leq\frac{1}{2}} + \sum_{k=1}^N \log(C_k) 1_{|C_k-1|>\frac{1}{2}}\right] \\ &= \frac{1}{\sqrt{2\log(N)}}\left[\log(N) + \sum_{k=1}^N \log(C_k) 1_{|C_k-1|\leq\frac{1}{2}}\right] \end{aligned}$$

$$\begin{aligned} &+\frac{1}{\sqrt{2\log(N)}}\left[\sum_{k=3}^{N}\log(C_k)1_{|C_k-1|>\frac{1}{2}}+\log(C_1)1_{|C_1-1|>\frac{1}{2}}+\log(C_2)1_{|C_2-1|>\frac{1}{2}}\right] \\ &= S_N+R_N \end{aligned} \tag{6}$$

where

$$S_N=\frac{1}{\sqrt{2\log(N)}}\left[\log(N)+\sum_{k=1}^{N}\log(C_k)1_{|C_k-1|\leq\frac{1}{2}}\right] \tag{7}$$

and

$$R_N=\frac{1}{\sqrt{2\log(N)}}\left[\log(C_1)1_{|C_1-1|>\frac{1}{2}}+\log(C_2)1_{|C_2-1|>\frac{1}{2}}+\sum_{k=3}^{N}\log(C_k)1_{|C_k-1|>\frac{1}{2}}\right]. \tag{8}$$

2.2. Chaos expansion

For every $k\geq 1, l=1,\dots,k$, we can express the random variable $X_{k,l}$ as $X_{k,l}=I_1(h_{k,l})$ where I_q denotes the multiple integral of order $q\geq 1$ with respect to the isonormal process W. Thus we can also write for every $k\geq 1$,

$$\begin{aligned} S_k &= I_1(h_{k,1})^2+I_1(h_{k,2})^2+\cdots+I_1(h_{k,k})^2 \\ &= I_2(h_{k,1}^{\otimes 2})+\cdots+I_2(h_{k,k}^{\otimes 2})+k=I_2(g_k)+k \end{aligned}$$

where for $k\geq 1$, we use the notation

$$g_k=h_{k,1}^{\otimes 2}+\cdots h_{k,k}^{\otimes 2}. \tag{9}$$

Similarly, the random variable C_k defined by (3) can be written as

$$C_k=\frac{I_2(g_k)}{k}+1. \tag{10}$$

3. The negligible part

In this section we estimate the negligible part of the sequence $(T_N)_{N\geq 1}$ given by (4). This negligible part will be composed by the remainder R_N given by (8) and from other summands derived from S_N which will appear when we use the Taylor expansion of the logarithm function.

Let us start with the following lemmas which will give and estimate of the summand (8).

Lemma 2. *For $i=1,2$, we have*

$$\mathbf{E}\left|\log(C_i)1_{|C_i-1|>\frac{1}{2}}\right|\leq C$$

with $C>0$ an universal constant.

Proof. Assume first $i=1$ and recall that $C_1\sim Z^2$ where $Z\sim N(0,1)$. Consequently, $C_1\sim\Gamma\left(\frac{1}{2},\frac{1}{2}\right)$ (we denoted by Γ the gamma function) and

$$\begin{aligned} \mathbf{E}\left|\log(C_1)1_{|C_1-1|>\frac{1}{2}}\right| &= C\int_0^\infty|\log(x)|x^{-\frac{1}{2}}e^{-\frac{1}{2}x}1_{|x-1|>\frac{1}{2}}dx \\ &= C\int_{\frac{3}{2}}^\infty\log(x)x^{-\frac{1}{2}}e^{-\frac{1}{2}x}dx \\ &\quad -C\int_0^{\frac{1}{2}}\log(x)x^{-\frac{1}{2}}e^{-\frac{1}{2}x}dx:=I_1+I_2 \end{aligned}$$

Since $0<\log(x)\leq x-1$ for $x>1$ we have

$$I_1\leq C\int_{\frac{3}{2}}^\infty x^{\frac{1}{2}}e^{-\frac{1}{2}x}dx\leq C.$$

For I_2, by bounding the exponential function by 1 and with the change of variables $-\log(x)=y$ we get

$$I_2\leq C\int_{\log(2)}^\infty ye^{-\frac{1}{2}y}dy\leq C.$$

Let us assume now $i = 2$ and recall that $C_2 \overset{(d)}{=} \frac{Z_1^2+Z_2^2}{2}$ with Z_1, Z_2 independent standard normal random variables. Therefore $C_2 \sim Exp(1)$ (the exponential distribution with parameter 1) and we have without difficulty

$$\begin{aligned}\mathbf{E}\left|\log(C_2)1_{|C_2-1|>\frac{1}{2}}\right| &= \int_0^\infty |\log(x)|e^{-x}1_{|x-1|>\frac{1}{2}}dx\\ &= \int_{\frac{3}{2}}^\infty \log(x)e^{-x}dx + \int_0^{\frac{1}{2}} -\log(x)e^{-x}dx \leq C. \quad \blacksquare\end{aligned}$$

Lemma 3. *For $N \geq 1$, we have*

$$\mathbf{E}\left|\sum_{k=3}^N \log(C_k)1_{|C_k-1|>\frac{1}{2}}\right| \leq C$$

with $C > 0$ not depending on N.

Proof. By definition

$$C_k \sim \frac{Z_1^2 + \cdots + Z_k^2}{k}$$

with $Z_1, \ldots, Z_k$ standard Gaussian independent random variables. We have, for every $k \geq 3$

$$\mathbf{E}\left|\log(C_k)1_{|C_k-1|>\frac{1}{2}}\right| = \mathbf{E}\log(C_k)1_{C_k>\frac{3}{2}} - \mathbf{E}\log(C_k)1_{0\leq C_k<\frac{1}{2}} = a_{1,k} + a_{2,k}.$$

To deal with $a_{2,k}$, we use the inequality $-\log(x) \leq \frac{1}{x} - 1$ for $x \in (0, 1)$. So

$$\begin{aligned}a_{2,k} &\leq \mathbf{E}\left(\frac{1}{C_k} - 1\right)1_{0\leq C_k<\frac{1}{2}} \leq \mathbf{E}\frac{1}{C_k}1_{0\leq C_k<\frac{1}{2}}\\ &\leq k\mathbf{E}\left(\frac{1}{Z_1^2+Z_2^2+Z_3^2}1_{0\leq C_k<\frac{1}{2}}\right)\end{aligned}$$

where we used $k \geq 3$. Take $p \in \left(1, \frac{3}{2}\right)$. Then

$$a_{2,k} \leq k\left(\mathbf{E}\left(\frac{1}{Z_1^2+Z_2^2+Z_3^2}\right)^p\right)^{\frac{1}{p}}\left(P(0 \leq C_k < \frac{1}{2})\right)^{\frac{p-1}{p}}$$

and, since the density of $Z_1^2 + \cdots + Z_k^2$ is $f(x) = 2^{-\frac{k}{2}}\frac{1}{\Gamma\left(\frac{k}{2}\right)}x^{\frac{k}{2}-1}e^{-\frac{1}{2}x}1_{x>0}$, we have

$$\mathbf{E}\left(\frac{1}{Z_1^2+Z_2^2+Z_3^2}\right)^p = C\int_0^\infty x^{-p+\frac{1}{2}}e^{-\frac{1}{2}x}dx < C$$

since $p < \frac{3}{2}$. Consequently, since $\left(0 \leq C_k < \frac{1}{2}\right) \subset \left(|C_k - 1| > \frac{1}{2}\right)$, we obtain, for every $a \geq 1$, via Lemma 1,

$$\begin{aligned}a_{2,k} &\leq Ck\left(P(0 \leq C_k < \frac{1}{2})\right)^{\frac{p-1}{p}} \leq Ck\left(P(|C_k - 1| > \frac{1}{2})\right)^{\frac{p-1}{p}}\\ &\leq Ck\left(\mathbf{E}|C_k - 1|^a\right)^{\frac{p-1}{p}} \leq Ck^{1-\frac{a}{2}\frac{p-1}{p}}\end{aligned}$$

and hence, by choosing a sufficiently large,

$$\sum_{k=3}^N a_{2,k} \leq C\sum_{k=3}^\infty k^{1-\frac{a}{2}\frac{p-1}{p}} \leq C. \tag{11}$$

To handle the summand $a_{1,k}$, we use the bound $\log(x) \leq x - 1 \leq x$ for $x \in (1, \infty)$. Then for every $p \geq 2$,

$$\begin{aligned}a_{1,k} &\leq \mathbf{E}\left(|C_k - 1|1_{C_k>\frac{3}{2}}\right) \leq \left(\mathbf{E}|C_k - 1|^2\right)^{\frac{1}{2}}\left(P\left(C_k \geq \frac{3}{2}\right)\right)^{\frac{1}{2}}\\ &\leq \left(\mathbf{E}|C_k - 1|^2\right)^{\frac{1}{2}}\left(P\left(|C_k - 1| \geq \frac{1}{2}\right)\right)^{\frac{1}{2}}\\ &\leq C_p\left(\mathbf{E}|C_k - 1|^2\right)^{\frac{1}{2}}\left(\mathbf{E}|C_k - 1|^p\right)^{\frac{1}{2}} \leq C_pk^{-\frac{1}{2}-\frac{p}{4}}.\end{aligned}$$

By choosing p large enough, we get

$$\sum_{k=3}^{N} a_{1,k} \leq C_p \sum_{k\geq 3} k^{-\frac{1}{2}-\frac{p}{4}} \leq C. \tag{12}$$

The conclusion follows by (11) and (12). ■

Let us conclude the estimation of the remainder R_N.

Corollary 1. *Let R_N be given by (8). Then for every $N \geq 1$,*

$$\mathbf{E}|R_N| \leq C\frac{1}{\sqrt{\log(N)}}.$$

Proof. The result is an immediate consequence of Lemmas 2 and 3. ■

Next, we analyze the sequence S_N defined by (7), i.e.

$$S_N = \frac{1}{\sqrt{2\log(N)}}\left[\log(N) + \sum_{k=1}^{N}\log(C_k)1_{|C_k-1|\leq\frac{1}{2}}\right]$$

for every $N \geq 1$. By using the expansion of the logarithm function

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3(1+\theta x)^3}$$

for $x \in (-1, 1)$ with $\theta \in (0, 1)$ (which depends on x), we can write

$$\begin{aligned}\log(C_k)1_{|C_k-1|\leq\frac{1}{2}} &= (C_k-1)1_{|C_k-1|\leq\frac{1}{2}} - \frac{1}{2}(C_k-1)^2 1_{|C_k-1|\leq\frac{1}{2}}\\ &\quad + \frac{(C_k-1)^3}{3(1+\theta(C_k-1))^3}1_{|C_k-1|\leq\frac{1}{2}}\end{aligned}$$

with $\theta \in (0, 1)$ a random point depending on C_k. Moreover,

$$\begin{aligned}\log(C_k)1_{|C_k-1|\leq\frac{1}{2}} &= (C_k-1) - (C_k-1)1_{|C_k-1|>\frac{1}{2}} - \frac{1}{2}(C_k-1)^2\\ &\quad + \frac{1}{2}(C_k-1)^2 1_{|C_k-1|>\frac{1}{2}}\\ &\quad + \frac{(C_k-1)^3}{3(1+\theta(C_k-1))^3}1_{|C_k-1|\leq\frac{1}{2}}.\end{aligned}$$

Consequently, for every $N \geq 1$,

$$S_N = S_{N,1} + S_{N,2} + S_{N,3} + S_{N,4} + S_{N,5}$$

where

$$\begin{aligned}S_{N,1} &:= \frac{1}{\sqrt{2\log(N)}}\sum_{k=1}^{N}(C_k-1)\\ S_{N,2} &:= -\frac{1}{\sqrt{2\log(N)}}\sum_{k=1}^{N}(C_k-1)1_{|C_k-1|>\frac{1}{2}}\\ S_{N,3} &:= \frac{1}{\sqrt{2\log(N)}}\sum_{k=1}^{N}\frac{1}{2}(C_k-1)^2 1_{|C_k-1|>\frac{1}{2}}\\ S_{N,4} &:= -\frac{1}{\sqrt{2\log(N)}}\left(\sum_{k=1}^{N}\frac{1}{2}(C_k-1)^2 - \log(N)\right)\\ S_{N,5} &:= \frac{1}{\sqrt{2\log(N)}}\sum_{k=1}^{N}\frac{(C_k-1)^3}{3(1+\theta(C_k-1))^3}1_{|C_k-1|\leq\frac{1}{2}}.\end{aligned} \tag{13}$$

We will analyze separately the five summands from above. Actually, we show that $S_{N,i}$, $i = 2, 3, 4, 5$ converge to zero in $L^1(\Omega)$ as $N \to \infty$ while the dominant term $S_{N,1}$ will be treated via Malliavin calculus.

Lemma 4. *Let $S_{N,j}, j = 1, \dots, 5$ be given by (13). Then, for $j = 2, 3, 5$,*

$$\mathbf{E}|S_{N,j}| \leq C \frac{1}{\sqrt{\log(N)}} \tag{14}$$

with $C > 0$ not depending on N.

Proof. Let us start with the summand $S_{N,5}$. We use the inequality

$$\frac{|x|^3}{|1+\theta x|^3} \leq 8|x|^3$$

for $\theta \in (0, 1)$ and $|x| \leq \frac{1}{2}$ and we get

$$\begin{aligned}\mathbf{E}|S_{N,5}| &\leq C \frac{1}{\sqrt{\log(N)}} \sum_{k=1}^{N} \mathbf{E}|C_k - 1|^3 \\ &\leq C \frac{1}{\sqrt{\log(N)}} \sum_{k=1}^{N} k^{-\frac{3}{2}} \leq C \frac{1}{\sqrt{\log(N)}}.\end{aligned}$$

For the term $S_{N,3}$, we write, by using Markov inequality and Lemma 1, with $p \geq 2$,

$$\begin{aligned}\mathbf{E}|S_{N,3}| &\leq C \frac{1}{\sqrt{\log(N)}} \sum_{k=1}^{N} \mathbf{E}\left(|C_k - 1|^2 1_{|C_k - 1| > \frac{1}{2}}\right) \\ &\leq C \frac{1}{\sqrt{\log(N)}} \sum_{k=1}^{N} \left(\mathbf{E}|C_k - 1|^4\right)^{\frac{1}{2}} \left(P(|C_k - 1| > \frac{1}{2})\right)^{\frac{1}{2}} \\ &\leq C \frac{1}{\sqrt{\log(N)}} \sum_{k=1}^{N} \left(\mathbf{E}|C_k - 1|^4\right)^{\frac{1}{2}} \left(\mathbf{E}|C_k - 1|^p\right)^{\frac{1}{2}} \\ &\leq C \frac{1}{\sqrt{\log(N)}} \sum_{k=1}^{N} k^{-1-\frac{p}{4}} \leq C \frac{1}{\sqrt{\log(N)}}\end{aligned}$$

by choosing p sufficiently large. The summand $S_{N,2}$ can be treated similarly. Indeed, by using the inequality $|C_{k-1}| 1_{|C_k - 1| > \frac{1}{2}} < 2|C_k - 1|^2 1_{|C_k - 1| > \frac{1}{2}}$, we get

$$\begin{aligned}\mathbf{E}|S_{N,2}| &\leq C \frac{1}{\sqrt{\log(N)}} \sum_{k=1}^{N} \mathbf{E}\left(|C_k - 1| 1_{|C_k - 1| > \frac{1}{2}}\right) \\ &\leq C \frac{1}{\sqrt{\log(N)}} \sum_{k=1}^{N} \mathbf{E}\left(|C_k - 1|^2 1_{|C_k - 1| > \frac{1}{2}}\right) \leq C \frac{1}{\sqrt{\log(N)}}\end{aligned}$$

by using the above estimate for $\mathbf{E}|S_{N,3}|$. ■

Let us analyze the term $S_{N,4}$ given by

$$S_{N,4} = -\frac{1}{\sqrt{2\log(N)}} \left(\sum_{k=1}^{N} \frac{1}{2}(C_k - 1)^2 - \log(N)\right).$$

Lemma 5. *Consider the sequence $S_{N,4}$ given by (13). Then, for every $N \geq 1$, with $C > 0$ an universal constant,*

$$\mathbf{E}|S_{N,4}|^2 \leq C \frac{1}{\log(N)}.$$

Proof. Recall (see (10)) that $C_k - 1 = \frac{1}{k} I_2(g_k)$ for every $k \geq 1$ where the term g_k is given by (9). By the product formula (28),

$$\begin{aligned}(C_k - 1)^2 &= \frac{1}{k^2} I_2(g_k)^2 \\ &= \frac{1}{k^2} \left[I_4(g_k \otimes g_k) + 4 I_2(g_k \otimes_1 g_k)\right] + \mathbf{E}(C_k - 1)^2 \\ &= \frac{1}{k^2} \left[I_4(g_k \otimes g_k) + 4 I_2(g_k \otimes_1 g_k)\right] + \frac{2}{k}\end{aligned}$$

where we used $\mathbf{E}(C_k-1)^2 = \frac{2}{k}$. Consequently,

$$S_{N,4} = \frac{-1}{\sqrt{2\log(N)}}\left[\frac{1}{2}\sum_{k=1}^{N}\frac{1}{k^2}\left(I_4(g_k\otimes g_k)+4I_2(g_k\otimes_1 g_k)\right)\right.$$
$$\left.+\sum_{k=1}^{N}\frac{1}{k}-\log(N)\right]. \tag{15}$$

Via the behavior of the partial sum of the divergent harmonic series

$$\sum_{k=1}^{N}\frac{1}{k} = \log(N)+\gamma+O\left(\frac{1}{N}\right) \tag{16}$$

with γ the Euler–Mascheroni constant, we get from (15) (in the sequel we denote by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ the scalar product and the norm in H, respectively)

$$\begin{aligned}\mathbf{E}|S_{N,4}|^2 &\leq C\frac{1}{\log(N)}\mathbf{E}\left(\sum_{k=1}^{N}\frac{1}{k^2}\left(I_4(g_k\otimes g_k)+4I_2(g_k\otimes_1 g_k)\right)\right)^2\\ &\quad+C\frac{1}{\log(N)}\\ &= C\frac{1}{\log(N)}\sum_{k,l=1}^{N}\frac{1}{k^2l^2}\left(4!\langle g_k\widetilde{\otimes}g_k, g_l\widetilde{\otimes}g_l\rangle+32\langle g_k\widetilde{\otimes}_1g_k, g_l\widetilde{\otimes}_1g_l\rangle\right)\\ &\quad+C\frac{1}{\log(N)}\end{aligned}$$

By the expression of the kernel g_k in (9), we notice the following facts:

1. For every $k\geq 1$,
$$g_k\widetilde{\otimes}_1g_k = g_k\otimes_1 g_k = g_k.$$
2.
$$\langle g_k, g_l\rangle = \begin{cases}k, & \text{if } k=l\\ 0, & \text{if } k\neq l.\end{cases} \tag{17}$$
3. For $k,l=1,\dots,N$,
$$\langle g_k\widetilde{\otimes}g_k, g_l\widetilde{\otimes}g_l\rangle = 0 \text{ for } k\neq l.$$

We arrive at the following estimate

$$\begin{aligned}\mathbf{E}|S_{N,4}|^2 &\leq C\frac{1}{\log(N)}\sum_{k=1}^{N}\frac{1}{k^4}\left[\|g_k\widetilde{\otimes}g_k\|^2+\|g_k\|^2\right]\\ &\quad+C\frac{1}{\log(N)}\end{aligned}$$

and since

$$\|g_k\widetilde{\otimes}g_k\|^2 \leq \|g_k\otimes g_k\|^2 = \|g_k\|^4 = k^2,$$

we will have

$$\mathbf{E}|S_{N,4}|^2 \leq C\frac{1}{\log(N)}\left[1+\sum_{k=1}^{N}\left(\frac{1}{k^2}+\frac{1}{k^3}\right)\right] \leq C\frac{1}{\log(N)}. \quad \blacksquare$$

4. The dominant part and the rate of convergence

Let us finally deal with the sequence $S_{N,1}$ from (13) i.e.

$$S_{N,1} = \frac{1}{\sqrt{2\log(N)}}\sum_{k=1}^{N}(C_k-1).$$

We know (see e.g. Rempala and Wesolowski (2005), Bourgade and Mody (2019) and Nguyen and Vu (2014)) that the sequence $(S_{N,1}, N \geq 1)$ converges in distribution to the standard normal law. Our purpose is to find its rate of convergence under the Kolmogorov, total variation, 1-Wasserstein or Fortet–Mourier metrics. Let us recall the definition of these metrics. If X, Y are two real-valued random variables, the usual way to define the distance between the law of X and the law of Y is to set

$$d(X,Y) = \sup_{h \in \mathcal{A}} |\mathbf{E}h(X) - \mathbf{E}h(Y)| \tag{18}$$

where $\mathcal{A}$ is a suitable class of functions. If $\mathcal{A} = \{1_{(-\infty,z]}, z \in \mathbb{R}\}$ then (18) gives the Kolmogorov distance, if $\mathcal{A} = \{1_A, A \in \mathcal{B}(\mathbb{R})\}$, then we have the total variation distance, when $\mathcal{A} = \{h : \mathbb{R} \to \mathbb{R}, \|h\|_{Lip} \leq 1\}$ we have the 1–Wasserstein distance and for $\mathcal{A} = \{h : \mathbb{R} \to \mathbb{R}, \|h\|_{Lip} + \|h\|_\infty \leq 1\}$ we obtain in (18) the Fortet–Mourier distance. By $\|\cdot\|_{Lip}$ we denoted the Lipschitz norm.

The recent Stein–Malliavin theory (see Nourdin and Peccati (2012)) represents a method to find an explicit bound for the distance between the law of a sequence $(G_N)_{N\geq 1}$ of random variables in a Wiener chaos of fixed order and the normal law. We refer, among others, to Nualart and Ortiz-Latorre (2009), Nourdin and Peccati (2009, 2012) and the references therein for more details. The notation in the below statement are those from Appendix.

Theorem 1. *Fix $q \geq 1$. Assume that $(G_N)_{N\geq 1} = (I_q(g_N))_{N\geq 1}$ with $g_N \in H^{\odot q}$, a sequence of random variables belonging to the qth Wiener chaos such that:*

$$\mathbf{E}(G_N^2) \underset{N\to\infty}{\longrightarrow} \sigma^2.$$

Then, G_N converges in law to $Z \sim \mathcal{N}(0,\sigma^2)$ if and only if

$$\|DG_N\|_H^2 \underset{N\to\infty}{\longrightarrow} q\sigma^2.$$

Furthermore,

$$d\left(G_N, \mathcal{N}(0,\sigma^2)\right) \leq C\left(\sqrt{\mathbf{Var}\left(\|DG_N\|_H^2\right)} + \sqrt{\mathbf{E}\left(\|DG_N\|_H^2\right) - q\sigma^2}\right). \tag{19}$$

The result in Theorem 1 can be applied to the sequence $S_{N,1}$ because this sequence belongs to the second Wiener chaos. Indeed, by (10),

$$S_{N,1} = \frac{1}{\sqrt{2\log(N)}} \sum_{k=1}^N \frac{1}{k} I_2(g_k) = I_2(f_N) \tag{20}$$

with

$$f_N = \frac{1}{\sqrt{2\log(N)}} \sum_{k=1}^N \frac{1}{k} g_k.$$

We will prove the following result. Below d is one of the distances defined before Theorem 1 (Kolmogorov, total variation, 1-Wasserstein or Fortet–Mourier distance).

Theorem 2. *Let $S_{N,1}$ be given by (13). Then for N large enough,*

$$d(S_{N,1}, N(0,1)) \leq C\frac{1}{\sqrt{\log(N)}}.$$

Proof. Let us notice that, via (17),

$$\begin{aligned}\mathbf{E}S_{N,1}^2 &= \frac{1}{2\log(N)} 2 \sum_{k,l=1}^N \frac{1}{kl}\langle g_k, g_l\rangle \\ &= \frac{1}{\log(N)} \sum_{k=1}^N \frac{1}{k^2}\|g_k\|^2 = \frac{1}{\log(N)} \sum_{k=1}^N \frac{1}{k} \to_{N\to\infty} 1\end{aligned}$$

and by using (16), for N sufficiently large,

$$\left|\mathbf{E}|S_{N,1}|^2 - 1\right| \leq C\frac{1}{\log(N)}. \tag{21}$$

The Malliavin derivative of $S_{N,1}$ reads

$$DS_{N,1} = \frac{2}{\sqrt{2\log(N)}} \sum_{k=1}^{N} \frac{1}{k} \left(h_{k,1} I_1(h_{k,1}) + \cdots + h_{k,k} I_1(h_{k,k}) \right)$$

and then, since $\langle h_{k,i}, h_{l,j} \rangle = 0$ if $k \neq l$ or $i \neq j$,

$$\begin{aligned} \|DS_{N,1}\|^2 &= \frac{4}{2\log(N)} \sum_{k,l=1}^{N} \frac{1}{kl} \langle \sum_{i=1}^{k} h_{k,i} I_1(h_{k,i}), \sum_{j=1}^{l} h_{l,j} I_1(h_{l,j}) \rangle \\ &= \frac{2}{\log(N)} \sum_{k=1}^{N} \frac{1}{k^2} \sum_{i=1}^{k} I_1(h_{k,i})^2 = \frac{2}{\log(N)} \sum_{k=1}^{N} \frac{1}{k^2} \sum_{i=1}^{k} \left(I_2(h_{k,i}^{\otimes 2}) + 1 \right) \\ &= \frac{2}{\log(N)} \sum_{k=1}^{N} \frac{1}{k^2} I_2(g_k) + \frac{2}{\log(N)} \sum_{k=1}^{N} \frac{1}{k}. \end{aligned}$$

We have

$$\mathbf{E}\|DS_{N,1}\|^2 = \frac{2}{\log(N)} \sum_{k=1}^{N} \frac{1}{k}$$

and by (17),

$$\begin{aligned} Var\left(\|DS_{N,1}\|^2 \right) &= \mathbf{E} \left(\frac{2}{\log(N)} \sum_{k=1}^{N} \frac{1}{k^2} I_2(g_k) \right)^2 = \frac{8}{\log(N)^2} \sum_{k,l=1}^{N} \frac{1}{k^2 l^2} \langle g_k, g_l \rangle \\ &= \frac{8}{\log(N)^2} \sum_{k=1}^{N} \frac{1}{k^4} \|g_k\|^2 = \frac{8}{\log(N)^2} \sum_{k=1}^{N} \frac{1}{k^3} \end{aligned}$$

and therefore

$$Var\left(\|DS_{N,1}\|^2 \right) \leq C \frac{1}{\log(N)^2}. \tag{22}$$

The conclusion is obtained by (21) and (22) via Theorem 1, by noticing that

$$\left| \mathbf{E}\|DS_{N,1}\|^2 - 2 \right| = 2 \left| \mathbf{E}S_{N,1}^2 - 1 \right| \leq C \frac{1}{\log(N)}. \quad \blacksquare$$

Below, d_W will stand for the 1-Wasserstein distance.

Theorem 3. *Consider the sequence $(T_N, N \geq 1)$ given by (4). Then $T_N \overset{(d)}{\to} N(0,1)$ and for N large*

$$d_W(T_N, N(0,1)) \leq C \frac{1}{\sqrt{\log(N)}}. \tag{23}$$

Proof. By the decomposition (6) and the triangle inequality,

$$d_W(T_N, N(0,1)) \leq d_W(S_{N,1}, N(0,1)) + d_W(T_N, S_{N,1}).$$

By Theorem 2, $d_W(S_{N,1}, N(0,1)) \leq C \frac{1}{\sqrt{\log(N)}}$ while by the definition of the 1-Wasserstein distance,

$$\begin{aligned} d_W(T_N, S_{N,1}) &\leq \mathbf{E}|T_N - S_{N,1}| \leq \mathbf{E} \left| R_N + \sum_{i=2}^{5} S_{N,i} \right| \\ &\leq \mathbf{E} \left| R_N + \sum_{i=2,3,5} S_{N,i} \right| + \left(\mathbf{E} \left| S_{N,4} \right|^2 \right)^{\frac{1}{2}} \\ &\leq C \frac{1}{\sqrt{\log(N)}} \end{aligned}$$

where we used Corollary 1, Lemmas 4 and 5. $\blacksquare$

5. Application to random determinants

Let us consider a $n \times n$ random matrix $\mathcal{X}_n = (x_{i,j})_{1\leq i,j\leq n}$ and denote by X_j its jth column vector, i.e. $X_j = (x_{1,j}, x_{2,j}, \ldots, x_{n,j})^T$ for $j = 1, \ldots, n$. Assume that the random vectors $(X_j, j = 1, \ldots, n)$ are independent and

$$X_j \sim N(0, \Sigma)$$

where Σ is an invertible matrix. We also denote by $\mathcal{W}_n$ the associated Wishart matrix,
i.e. $\mathcal{W}_n = \mathcal{X}_n\mathcal{X}_n^T$. Then we have the following result.

Theorem 4. *Let the above notation prevail. For every $N \geq 1$, let*

$$U_N = \frac{1}{\sqrt{2\log N}} \left[\log(\det \mathcal{W}_N) - \log(\det \Sigma) - \log(N-1)!\right]. \tag{24}$$

Then the sequence $(U_N)_{N\geq 1}$ convergence in law to the standard normal distribution and for N large enough

$$d_W\left(U_N, N(0,1)\right) \leq C\frac{1}{\sqrt{\log N}}.$$

Proof. It is well-known (see e.g. Theorem 4 in Rempala and Wesolowski (2005)) that ($\overset{(d)}{=}$ means the equality in distribution)

$$\frac{\det \mathcal{W}_n}{\det \Sigma} \overset{(d)}{=} \prod_{k=1}^{N}\left(\sum_{l=1}^{k} Y_{k,l}^2\right) \overset{(d)}{=} \prod_{k=1}^{N} S_k$$

where $(Y_{k,l}, k \geq 1, l = 1, \ldots, k)$ are i.i.d. standard normal random variables and S_k is given by (2). It then suffices to apply Theorem 3. ■

We can also write the above result as

$$\left(\frac{\det(\mathcal{W}_N)}{\det(\Sigma)(N-1)!}\right)^{\frac{1}{\sqrt{2\log(N)}}} \to^{(d)}_{=} e^Z$$

where $Z \sim N(0, 1)$. Concerning the determinant of the random matrix $\mathcal{X}_n$ described at the beginning of this section, we can write

$$\frac{1}{\sqrt{2\log N}}\left[2\log|\det \mathcal{X}_n| - \log(\det(\Sigma)) - \log(N-1)!\right] \overset{(d)}{=} N(0,1)$$

and

$$d_W\left(\frac{1}{\sqrt{2\log N}}\left[2\log|\det \mathcal{X}_n| - \log(det(\Sigma)) - \log(N-1)!\right], N(0,1)\right) \leq C\frac{1}{\sqrt{\log N}}.$$

Data availability

No data was used for the research described in the article.

Acknowledgments

C. Tudor acknowledges partial support from Labex CEMPI (ANR-11-LABX-007-01), the ANR project ANR-22-CE40-0015 (SDAIM), the MATHAMSUD (project 22- MATH-08), ECOS SUD (project C2107) and by the Ministry of Research, Innovation and Digitalization (Romania), CNCS-UEFISCDI (grant PN-III-P4-PCE-2021-0921, within PNCDI III).

Appendix

Here, we shall only recall some elementary facts; our main reference is Nualart (2006). Consider $\mathcal{H}$ a real separable infinite-dimensional Hilbert space with its associated inner product $\langle .,. \rangle_{\mathcal{H}}$, and $(B(\varphi), \varphi \in \mathcal{H})$ an isonormal Gaussian process on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, which is a centered Gaussian family of random variables such that $\mathbf{E}\left(B(\varphi)B(\psi)\right) = \langle \varphi, \psi \rangle_{\mathcal{H}}$, for every $\varphi, \psi \in \mathcal{H}$. Denote by I_q the qth multiple stochastic integral with respect to B. This I_q is actually an isometry between the Hilbert space $\mathcal{H}^{\odot q}$ (symmetric tensor product) equipped with the scaled norm $\frac{1}{\sqrt{q!}}\|\cdot\|_{\mathcal{H}^{\otimes q}}$ and the Wiener chaos of order q, which is defined as the closed linear span of the random variables $H_q(B(\varphi))$ where $\varphi \in \mathcal{H},\ \|\varphi\|_{\mathcal{H}} = 1$ and H_q is the Hermite polynomial of degree $q \geq 1$ defined by:

$$H_q(x) = (-1)^q \exp\left(\frac{x^2}{2}\right)\frac{\mathrm{d}^q}{\mathrm{d}x^q}\left(\exp\left(-\frac{x^2}{2}\right)\right), \quad x \in \mathbb{R}. \tag{25}$$

The isometry of multiple integrals can be written as: for $p,\ q \geq 1,\ f \in \mathcal{H}^{\otimes p}$ and $g \in \mathcal{H}^{\otimes q}$,

$$\mathbf{E}\Big(I_p(f)I_q(g)\Big) = \begin{cases} q!\langle \tilde{f}, \tilde{g}\rangle_{\mathcal{H}^{\otimes q}} & \text{if } p = q \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

where $\tilde{f}$ denotes the canonical symmetrization of f and it is defined by:

$$\tilde{f}(x_1, \dots, x_q) = \frac{1}{q!} \sum_{\sigma \in \mathcal{S}_q} f(x_{\sigma(1)}, \dots, x_{\sigma(q)}),$$

in which the sum runs over all permutations σ of $\{1, \dots, q\}$. It also holds that:

$$I_q(f) = I_q(\tilde{f}).$$

In the particular case when $\mathcal{H} = L^2(T, \mathcal{B}(T), \lambda)$ (λ being the Lebesgue measure), the rth contraction $f \otimes_r g$ is the element of $\mathcal{H}^{\otimes(p+q-2r)}$, which is defined by:

$$\begin{aligned} &(f \otimes_r g)(s_1, \dots, s_{p-r}, t_1, \dots, t_{q-r}) \\ &= \int_{T^r} \mathrm{d}u_1 \dots \mathrm{d}u_r f(s_1, \dots, s_{p-r}, u_1, \dots, u_r) g(t_1, \dots, t_{q-r}, u_1, \dots, u_r), \end{aligned} \tag{27}$$

for every $f \in L^2(T^p)$, $g \in L^2(T^q)$ and $r = 1, \dots, p \wedge q$. By $f \tilde{\otimes}_l g$ we denote the symmetrization of the contraction $f \otimes_l g$.

The product for two multiple integrals can be expanded into a sum of multiple integrals (see Nualart (2006)): if $f \in L^2(T^n)$ and $g \in L^2(T^m)$ are symmetric functions, then it holds that

$$I_n(f)I_m(g) = \sum_{l=0}^{m \wedge n} l! C_m^l C_n^l I_{m+n-2l}(f \otimes_l g). \tag{28}$$

Let $\mathcal{S}$ be the class of smooth functionals of the form

$$F = f(B_{t_1}, \dots, B_{t_n}), \qquad t_1, \dots, t_n \in T, \tag{29}$$

with $f \in C^\infty(\mathbb{R}^n)$ with at most polynomial growth (for f and its derivatives). For the random variable (29) we define its Malliavin derivative with respect to B by

$$D_t F = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(B_{t_1}, \dots, B_{t_n}) 1_{[0,t_i]}(t), \qquad t \in T.$$

The operator D is an unbounded closable operator and it can be extended to the closure of $\mathcal{S}$ with respect to the Malliavin–Sobolev norm

$$\|F\|_{k,p}^p = \mathbf{E}|F|^p + \sum_{j=1}^k \mathbf{E}\|D^{(j)}F\|_{L^2(T^j)}^p, \quad F \in \mathcal{S}, p \geq 2, k \geq 1. \tag{30}$$

where $D^{(j)}$ stands for the jth iterated Malliavin derivative. This closure will be denoted by $\mathbb{D}^{k,p}$.

The Malliavin derivative D acts on the Wiener chaos as an annihilation operator: if $F = I_n(f)$ with $f \in L^2(T^n)$ symmetric, then $D_t F = n I_{n-1}(f(\cdot, t))$ where "$\cdot$" stands for $n-1$ variables in T. The pseudo-inverse of the Ornstein–Uhlenbeck operator, denoted by L^{-1}, acts as follows: if $F = I_n(f)$ with $n \geq 1$, then $-L^{-1}F = \frac{1}{n}F$.

References

Arnold, B.C., Villasenor, J.A., 1998. The asymptotic distribution of sums of records. Extremes 1 (3), 351–363.
Bao, Z., Pan, G., Zhou, W., 2015. The logarithmic law of random determinant. Bernoulli 21 (3), 1600–1628.
Bourgade, P., Mody, K., 2019. Gaussian fluctuations of the determinant of Wigner matrices. Electron. J. Probab. 24, 28, Paper No. 96.
Forsythe, G.E., Tukey, J.W., 1952. The extent of n random unit vectors. Bull. Amer. Math. Sot. 58, 502.
Goodman, N.R., 1963. The distribution of the determinant of a complex Wishart distributed matrix. Ann. Math. Stat. 34, 178–180.
Lee, A.J., 1990. U-Statistics. In: Statistics: Textbooks and Monographs, Theory and Practice, vol. 110, Marcel Dekker, New York.
Nguyen, H.H., Vu, V., 2014. Random matrices: law of the determinant. Ann. Probab. 42 (1), 146–167.
Nourdin, I., Peccati, G., 2009. Stein's method on Wiener chaos. Probab. Theory Related Fields 145 (1–2), 75–118.
Nourdin, I., Peccati, G., 2012. Normal Approximations with Malliavin Calculus from Stein's Method to Universality. Cambridge University Press.
Nualart, D., 2006. The Malliavin calculus and related topics. In: Probability and its Applications, second ed. Springer-Verlag, Berlin.
Nualart, D., Ortiz-Latorre, S., 2009. Central limit theorems for multiple stochastic integrals and Malliavin calculus. Stochastic Process. Appl. 118 (4), 614–628.
Nyquist, H., Rice, S.O., Riordan, J., 1954. The distribution of random determinants. Quart. Appl. Math. 12, 97–104.
Prékopa, A., 1967. On random determinants. I. Studia Sci. Math. Hungar. 2, 125–132.
Rempala, G., Wesolowski, J., 2002. Asymptotics for products of sums and U-statistics. Electron. Commun. Probab. 7, 47–54.
Rempala, G., Wesolowski, J., 2005. Asymptotics for products of independent sums with an application to Wishart determinants. Statist. Probab. Lett. 77, 129–138.
Turań, P., 1955. On a problem in the theory of determinants. Acta Math. Sinica 5, 411–423.