Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 162 (2023) 423–455

stochastic processes and their applications

www.elsevier.com/locate/spa

Yule's "nonsense correlation" for Gaussian random walks

Philip A. Ernst[a,*], Dongzhou Huang[b], Frederi G. Viens[c]

[a] *Department of Mathematics, Imperial College London, London SW7 2AZ, UK*
[b] *Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA*
[c] *Department of Statistics, Rice University, Houston, TX 77005, USA*

Received 2 July 2022; received in revised form 7 March 2023; accepted 8 April 2023
Available online 26 April 2023

Abstract

This paper provides an exact formula for the second moment of the empirical correlation (also known as Yule's "nonsense correlation") for two independent standard Gaussian random walks, as well as implicit formulas for higher moments. We also establish rates of convergence of the empirical correlation of two independent standard Gaussian random walks to the empirical correlation of two independent Wiener processes.

MSC: primary 60G50; 60G15; secondary 60G05

Keywords: Nonsense correlation; Gaussian random walks; Wiener processes; Wasserstein distance

1. Introduction

The task of establishing an explicit formula for the moments of the empirical correlation (also known as Yule's "nonsense correlation") for two independent random walks has long been believed to be intractable (see, for example, [8, Remark 1.1], and references therein). In the present manuscript, we make significant progress towards closing this longstanding open question by providing an exact formula for the second moment of the empirical correlation of two independent random walks when the steps in the random walks are *standard Gaussian*. We also succeed in providing implicit formulas for higher moments. We then turn our focus towards

* Corresponding author.
E-mail addresses: p.ernst@imperial.ac.uk (P.A. Ernst), dongzhou.huang@colostate.edu (D. Huang), fv15@rice.edu (F.G. Viens).

https://doi.org/10.1016/j.spa.2023.04.007

establishing rates of convergence of the empirical correlation of two independent standard Gaussian random walks to the empirical correlation of two independent Wiener processes.

We proceed with some notation. Let $\{X_k\}_{k=1}^{\infty}$ and $\{Y_k\}_{k=1}^{\infty}$ be two independent sequences of independent identically distributed random variables with mean 0 and variance 1. Define the corresponding partial sums by

$$S_n = \sum_{j=1}^{n} X_j \quad \text{and} \quad T_n = \sum_{j=1}^{n} Y_j. \tag{1}$$

The empirical correlation of these two random walks is then defined in the usual way (see [21]) as

$$\theta_n := \frac{\frac{1}{n}\sum_{i=1}^{n} S_i T_i - \frac{1}{n^2}(\sum_{i=1}^{n} S_i)(\sum_{i=1}^{n} T_i)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} S_i^2 - \frac{1}{n^2}(\sum_{i=1}^{n} S_i)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n} T_i^2 - \frac{1}{n^2}(\sum_{i=1}^{n} T_i)^2}}. \tag{2}$$

Despite Udny Yule's warning in 1926 [21] that in the case of two independent random walks, the observed correlation coefficient has a very different distribution from that of the nominal t-distribution, it has been erroneously assumed that for large enough n, these empirical correlations should be small (see [8] and references therein).

The task of examining the distribution of θ_n for discrete-time processes is both interesting and relevant to practitioners because discrete stochastic process data (for example, time series data) occur most frequently and extensively in the real world. A test statistic for discrete processes is thus easier for statistical practitioners to apply than that for continuous stochastic processes. Studying the discrete-data test statistic directly also presents a means of minimizing the risk of using the continuous statistic when the discrete-data situation is not sufficiently well approximated by a continuous-data situation.

We now briefly survey the relevant literature. In [17], Phillips calculated an expression for the limit of the correlations θ_n (in the sense of weak convergence), which can be viewed as the empirical correlation of two independent Wiener processes. This work also provided a mathematical solution to the problem of spurious regression among integrated time series by demonstrating that statistical t-ratio and F-ratio tests diverge with the sample size, thereby explaining the observed 'statistical significance' in such regressions. In later work [18], the same author provided an explanation of such spurious regressions in terms of orthonormal representations of the Karhunen–Loève type (see e.g. [12, Section 5.3]). Let us denote θ to be the limit of the correlations θ_n. In 2017, Ernst et al. [8] investigated the distribution of the limit θ by explicitly calculating the standard deviation of the limit to be nearly 0.5, providing the first formal proof that these correlations θ_n are not small even for arbitrarily large n. In 2019, Ernst et al. [7] succeeded in calculating the moments of θ up to order 16 and provided the first approximation to the density of Yule's "nonsense correlation". A more recent manuscript concerning Yule's "nonsense correlation" is [5].

Despite the above work, the task of finding the exact distribution of θ_n for any n when calculated for two independent random walks has proven elusive. The most relevant work in this vein includes a series of papers by Andersen [1–3] which provided a combinatorial method based on the idea of cyclic permutations to investigate problems of discrete sequences of partial sums. However, Andersen's methods cannot be applied to evaluate the moments of θ_n since an event generated by θ_n is not invariant under cyclic permutations. The methods used in [18] to develop asymptotic theory for spurious regressions, namely, decomposing continuous stochastic process in terms of their orthonormal representations, cannot be employed to find the exact distribution of θ_n due to the lack of a continuous pattern in the partial sums S_k and T_k. In [6], Erdös and Kac investigated the asymptotic distributions of four statistics of partial sums of independent identically distributed random variables each having mean 0 and variance 1. However, their methods do not concern exact distributions, and therefore cannot be applied to the calculation of the exact distribution of θ_n. In [11], Magnus evaluated the moments of the ratio of a pair of quadratic forms in normal variables, i.e., $x'Ax/x'Bx$, where A is symmetric, B positive semidefinite and x is a Gaussian random vector. It is Magnus' work in particular which motivated this manuscript's focus on standard Gaussian random walks. As we shall see, this specific context enables us to derive an explicit formula to calculate the second moment of the empirical correlation θ_n for any n.

Our proof of this formula is based on a symbolically tractable integro-differential representation formula for the moments of any order in a class of empirical correlations, established by [7, Proposition 1] and investigated previously in [8] (see Proposition 1). The key step in applying this formula is the explicit computation of the joint moment generating function (mgf) ϕ_n of the three empirical sums of products and squares which appear in the empirical correlation θ_n. This is the topic of Section 3. One may also use this representation formula to compute moments of θ_n of any order numerically, using symbolic algebra software. Indeed, we provide these moments up to order 16 for any n. Thus the method for evaluating all moments relies on the joint mgf for the three bilinear and quadratic forms appearing in θ_n [7,8].

The key mathematical contribution of the present paper lies in the explicit computation of the joint trivariate mgf ϕ_n in Section 3. To express the second moment of θ_n via the aforementioned representation, it is necessary to compute the partial derivative of ϕ_n with respect to its middle variable (the variable representing the empirical covariance). This latter calculation, in Section 4, is only made possible by the explicitness of our formula for ϕ_n. The technical path followed in Section 3 to compute ϕ_n is to express in matrix form the bilinear form mapping the two i.i.d. data sequences X and Y up to the nth terms into the empirical covariance of their partial sums S_n and T_n, and to compute the matrix's alternative characteristic polynomial d_n. We derive an explicit expression for d_n in the Appendix, recursively for $n \geq 5$, by using standard operations to convert d_n into a linear recursion involving a new determinant in tri-diagonal form except for one line along which to expand the said new determinant. In doing so, we notice a slight break in the new determinant's recursive nature. When substituting a cell in the determinant's matrix which fixes this break, a second-order recursion emerges, which can be solved explicitly. Relating this back to the original d_n reveals a remarkably simple explicit relation, and thus, an explicit formula for d_n.

From a probabilistic viewpoint, it is the Gaussian property of (X, Y) which allows us to complete this calculation so explicitly. Specifically, we employ the following two properties: (i) the multivariate standard normal law is invariant under orthogonal transformations, and (ii) the Laplace transform of a quadratic form of a bivariate normal vector is a function of a quadratic function. From an analytical viewpoint, to compute d_n explicitly, we draw inspiration from the limiting case of S and T distributed as Brownian motions, where Hilbert's approach to Fredholm theory (see e.g. [9]) gives us a strong motivation to believe that d_n could be computed. Indeed, the limit of $d_n(\lambda)$ under the appropriate Brownian scaling is explicit, equal to $\sinh(i\sqrt{\lambda})/(i\sqrt{\lambda})$ which was a main ingredient in [8], and also equal, via Mercer's theorem, to $\prod_k(1-\lambda/(k\pi)^2)$. It is in this last expression where one recognizes the eigenvalues identified in [8].

A final contribution of this paper is our study of the rate of convergence of the empirical correlation θ_n of Gaussian random walks to the empirical correlation θ of Wiener processes in Wasserstein distance (see e.g. [20]). Inspired by Hilbert's approach to Fredholm theory, we first construct a ratio $A_n/\sqrt{B_nC_n}$ identically distributed with θ_n, where A_n, B_n and C_n are second-chaos variables up to constants. We also rewrite θ as $A/\sqrt{BC}$, where A, B and C are also second-chaos variables up to constants. A key element in the setup is to note that, not only can the empirical correlations be represented as ratios involving second-chaos variables, but they can also be coupled on the same Wiener space Ω by using their kernel representations as double integrals with respect to the same pair of independent Wiener processes. The details are contained in Section 5.1. Relying on techniques of Wiener chaos (for reference, see [13, Section 2.7]), we derive the convergences in $L^2(\Omega)$ of A_n, B_n and C_n to A, B and C respectively at the rate n^{-2}. We then note that the Wasserstein distance between θ_n and θ is bounded by the $L^1(\Omega)$-norm of $A_n/\sqrt{B_nC_n} - A/\sqrt{BC}$, which in turn is bounded by a function of the second moments of $A_n - A$, $B_n - B$ and $C_n - C$ and the negative moments of B_n, C_n, B and C. What then remains is to give upper bounds for the negative moments. Our idea is to represent these negative moments as a single integral of the product of a positive power function and their moment generating functions (mgfs) and then to give upper bounds for mgfs, hence, for negative moments. This idea only works when the mgfs are integrable at 0 and decay rapidly when approximating to ∞. Fortunately, these mgfs follow from the joint mgfs ϕ_n and ϕ and satisfy the above properties.

We wish to emphasize that the mgf of B_n or C_n is 1 over the square root of $d_n(-2s/n)$, which is a polynomial with strictly positive coefficients. Furthermore, the coefficients of $d_n(-2s/n)$ are eigenvalues of the positive definite matrix K_n (as defined in Section 2) after appropriate scaling. We anticipate that these eigenvalues converge to those of the positive definite operator T_M defined in [8]. This insight motivates us to establish the existence of a lower bound for $d_n(-2s/n)$ for $s \geq 0$ which is uniform for large enough n and hence a uniform upper bound for $E[B_n^{-1}]$ and $E[C_n^{-1}]$. All of these details are presented in Section 5.

The remainder of the paper is organized as follows. In Section 2, we introduce necessary notation. In Section 3, Theorem 3.2 provides the joint moment generating function needed for obtaining the distribution of θ_n (for all n). In Section 4, Theorem 4.1 provides an explicit formula for the second moment of θ_n for any n. Numerics for all moments of θ_n for all n are also given in Section 4. The latter motivates our investigation in Section 5 of the rate of convergence of θ_n to θ. We conclude with Section 6, which provides opportunities for future work which should be tractable given some known tools and techniques in the analysis on Wiener chaos, and could have potential applications to statistical testing based on paths of time series.

2. Notation

We use I_n to denote the $n \times n$ identity matrix. For $n \geq 2$ an integer, we define the $(n-1) \times (n-1)$ symmetric matrix K_n by

$$K_n = \left\{\min(j,k)/n - jk/n^2\right\}_{j,k=1}^{n-1},$$

and its "alternative characteristic polynomial" $d_n(\lambda)$ by

$$d_n(\lambda) = \det(I_{n-1} - \lambda K_n).$$

In the introduction, we explained that the matrix K_n is the discrete-time version of the operator T_M, the latter being critical to the success of the calculations in [8]. In this paper, it was shown that the numerator of the continuous-time Yule's "nonsense correlation" θ (see the definition of θ in (22) in Section 5) can be written as a member of the second Wiener chaos in its double-Wiener-integral representation, where the bivariate kernel M in that integral is $M(s,t) = \min(s,t) - st$. The expression above for K_n thus comes as no surprise, as the discrete version of M. However, as we will see in the next section, K_n also arises naturally when one attempts to express the numerator of θ_n using the increments X, Y of the random walks S, T. That natural phenomenon is exactly the discrete-time analog of what occurs when identifying the numerator of θ as a double Wiener integral.

Denoting the eigenvalues of K_n by $\lambda_2, \ldots, \lambda_n$ (where the numbering starting at 2 is used as a matter of convenience, whose utility will become apparent in the next section), the alternative characteristic polynomial can be written as

$$d_n(\lambda) = \prod_{j=2}^{n}(1 - \lambda_j \lambda). \tag{3}$$

We also define two $(n-1) \times 1$ column random vectors $\mathbf{X}_n$ and $\mathbf{Y}_n$ by

$$\mathbf{X}_n := (X_2, X_3, \ldots, X_n)^{\intercal} \quad \text{and} \quad \mathbf{Y}_n := (Y_2, Y_3, \ldots, Y_n)^{\intercal},$$

where $\{X_k\}_{k=1}^{\infty}$ and $\{Y_k\}_{k=1}^{\infty}$ are the two independent sequences of independent standard Gaussian random variables used to define the Gaussian random walks S and T. Let

$$Z_{11}^n := \frac{1}{n}\sum_{i=1}^{n} S_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} S_i\right)^2, \tag{4}$$

$$Z_{22}^n := \frac{1}{n}\sum_{i=1}^{n} T_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} T_i\right)^2, \tag{5}$$

$$Z_{12}^n := \frac{1}{n}\sum_{i=1}^{n} S_i T_i - \frac{1}{n^2}\left(\sum_{i=1}^{n} S_i\right)\left(\sum_{i=1}^{n} T_i\right), \tag{6}$$

where S_i and T_i are defined in (1). Together with (2), we may easily check that

$$\theta_n = \frac{Z_{12}^n}{\sqrt{Z_{11}^n Z_{22}^n}}.$$

Finally, let us define the joint moment generating function (joint mgf) of the random vector $\left(Z_{11}^n, Z_{12}^n, Z_{22}^n\right)$ by

$$\phi_n(s_{11}, s_{12}, s_{22}) := E\left[\exp\left\{-\frac{1}{2}\left(s_{11}Z_{11}^n + 2s_{12}Z_{12}^n + s_{22}Z_{22}^n\right)\right\}\right],$$

where s_{11}, s_{12} and s_{22} are such that $s_{11}, s_{22} \geq 0$ and $s_{12}^2 \leq s_{11}s_{22}$. These inequalities ensure that $\phi_n(s_{11}, s_{12}, s_{22})$ is well-defined, as we shall see in Section 3. The reader may also check, as a heuristic, that if the possibly ex-centered second-chaos variables $Z_{i,i}^n$ are thought of as independent squares of standard normals, and $Z_{1,2}^n$ is the product of the normals, that the condition $s_{12}^2 \leq s_{11}s_{22}$ becomes necessary.

3. Calculating the joint moment generating function

In this section, we provide an expression for the joint moment generating function $\phi_n(s_{11}, s_{12}, s_{22})$, enabling our computation of the moments of θ_n for all n. In the continuous-time setting of [8], being able to compute this mgf was as a critical step, and it relied on the fact that the kernel $M(s,t)=\min(s,t)-st$ of the operator T_M was immediately identified as the covariance of the pinned Brownian motion (a.k.a Brownian bridge) on [0, 1], for which the eigenvalues happen to be known. However, in the discrete case herein, there is no analogous shortcut.

First, by definition,

$$\sum_{i=1}^{n} S_i T_i = \sum_{i=1}^{n}\left(\sum_{j=1}^{i} X_j\right)\left(\sum_{k=1}^{i} Y_k\right)$$
$$= \sum_{j,k=1}^{n}\ \sum_{i=\max(j,k)}^{n} X_j Y_k = \sum_{j,k=1}^{n}(n-\max(j,k)+1)X_j Y_k. \tag{7}$$

Further,

$$\sum_{i=1}^{n} S_i = \sum_{i=1}^{n}\sum_{j=1}^{i} X_j = \sum_{j=1}^{n}\sum_{i=j}^{n} X_j = \sum_{j=1}^{n}(n-j+1)X_j.$$

Similarly,

$$\sum_{i=1}^{n} T_i = \sum_{k=1}^{n}(n-k+1)Y_k.$$

Hence,

$$\left(\sum_{i=1}^{n} S_i\right)\left(\sum_{i=1}^{n} T_i\right) = \sum_{j,k=1}^{n}(n-j+1)(n-k+1)X_j Y_k.$$

Together with (6) and (7), we have

$$\begin{aligned} Z_{12}^n &= \sum_{j,k=1}^{n}\left(\frac{1}{n}\Big(n-\max(j,k)+1\Big)-\frac{1}{n^2}(n-j+1)(n-k+1)\right)X_j Y_k \\ &= \sum_{j,k=1}^{n}\left(\frac{1}{n}\Big(\min(j,k)-1\Big)-\frac{1}{n^2}(j-1)(k-1)\right)X_j Y_k \\ &= \sum_{j,k=2}^{n}\left(\frac{1}{n}\Big(\min(j,k)-1\Big)-\frac{1}{n^2}(j-1)(k-1)\right)X_j Y_k \\ &= \sum_{j,k=1}^{n-1}\left(\frac{1}{n}\min(j,k)-\frac{1}{n^2}jk\right)X_{j+1}Y_{k+1} \\ &= \mathbf{X}_n^{\intercal} K_n \mathbf{Y}_n, \end{aligned} \tag{8}$$

where the third equality holds because $(\min(j,k)-1)/n-(j-1)(k-1)/n^2$ equals to 0 if either one of the indices j,k is 1 and the fourth equality holds by making the change of variables $j := j-1$ and $k := k-1$. As mentioned previously, we recognize $K_n(j,k)$ defined there and identified here in the last displayed line above, as the discrete version of $M(s,t)=\min(s,t)-st$. Similarly to the expression for Z_{12}^n, we have

$$Z_{11}^n = \mathbf{X}_n^{\intercal} K_n \mathbf{X}_n \quad \text{and} \quad Z_{22}^n = \mathbf{Y}_n^{\intercal} K_n \mathbf{Y}_n.$$

Since K_n is a $(n-1)\times(n-1)$ symmetric matrix, there exists a $(n-1)\times(n-1)$ orthogonal matrix P_n such that

$$K_n = P_n^{\top}\mathrm{diag}(\lambda_2, \lambda_3, \dots, \lambda_n)P_n,$$

where $\lambda_2, \lambda_3, \dots, \lambda_n$ are eigenvalues of K_n and $\mathrm{diag}(\lambda_2, \lambda_3, \dots, \lambda_n)$ is a diagonal matrix whose entry in the jth row and the jth column is λ_{j+1}. Let

$$\widetilde{\mathbf{X}}_n = \left(\widetilde{X}_2, \widetilde{X}_3, \dots, \widetilde{X}_n\right)^{\top} := P_n\mathbf{X}_n,$$
$$\widetilde{\mathbf{Y}}_n = \left(\widetilde{Y}_2, \widetilde{Y}_3, \dots, \widetilde{Y}_n\right)^{\top} := P_n\mathbf{Y}_n,$$

be two $(n-1)\times 1$ column random vectors. Since $\mathbf{X}_n$ and $\mathbf{Y}_n$ are two independent Gaussian random vectors with distribution $\mathcal{N}(\mathbf{0}, I_{n-1})$ and because P_n is an orthogonal matrix, then $\widetilde{\mathbf{X}}_n$ and $\widetilde{\mathbf{Y}}_n$ are also two independent Gaussian random vectors with distribution $\mathcal{N}(\mathbf{0}, I_{n-1})$. This implies that $\widetilde{X}_2, \widetilde{X}_3, \dots, \widetilde{X}_n, \widetilde{Y}_2, \widetilde{Y}_3, \dots, \widetilde{Y}_n$ are independent standard Gaussian random variables.

Before presenting our formula for the trivariate mgf ϕ_n in Theorem 3.2, we reveal an explicit calculation of the alternative characteristic polynomial $d_n(\lambda)$. The proof is relegated to the Appendix.

Lemma 3.1. *The alternative characteristic polynomial $d_n(\lambda)$ may be written as*

$$\begin{aligned} d_n(\lambda) &= \frac{1}{n\sqrt{\left(\frac{\lambda}{n}-2\right)^2-4}}\left(-\frac{\left(\frac{\lambda}{n}-2\right)-\sqrt{\left(\frac{\lambda}{n}-2\right)^2-4}}{2}\right)^n \\ &\quad - \frac{1}{n\sqrt{\left(\frac{\lambda}{n}-2\right)^2-4}}\left(-\frac{\left(\frac{\lambda}{n}-2\right)+\sqrt{\left(\frac{\lambda}{n}-2\right)^2-4}}{2}\right)^n \\ &= \frac{(-1)^{n-1}}{n\cdot 2^{n-1}}\sum_{k=1}^{\lceil n/2\rceil}\binom{n}{2k-1}\left(\frac{\lambda}{n}-2\right)^{n-(2k-1)}\left(\left(\frac{\lambda}{n}-2\right)^2-4\right)^{k-1}, \end{aligned} \tag{9}$$

where $\lceil x\rceil$ is the least integer greater than or equal to x.

Proof. See Appendix. $\square$

We now proceed to calculate the joint mgf ϕ_n.

Theorem 3.2. *The joint moment generating function ϕ_n for the triple $(Z_{11}^n, Z_{12}^n, Z_{22}^n)$ of random variables defined in (4), (6), (5) is given, for $s_{11}, s_{22}\geq 0$, and for $s_{12}^2\leq s_{11}s_{22}$, by*

$$\phi_n(s_{11}, s_{12}, s_{22}) = \left(d_n(\alpha)\, d_n(\beta)\right)^{-1/2}$$

where α and β are given as:

$$\alpha := \alpha\,(s_{11}, s_{12}, s_{22}) = -\frac{s_{11}+s_{22}+\sqrt{(s_{11}-s_{22})^2+4s_{12}^2}}{2}, \tag{10}$$

$$\beta := \beta\,(s_{11}, s_{12}, s_{22}) = -\frac{s_{11}+s_{22}-\sqrt{(s_{11}-s_{22})^2+4s_{12}^2}}{2}. \tag{11}$$

Proof. We first calculate

$$\begin{aligned}
&s_{11} Z_{11}^n + 2 s_{12} Z_{12}^n + s_{22} Z_{22}^n \\
&= s_{11} \mathbf{X}_n^\intercal K_n \mathbf{X}_n + 2 s_{12} \mathbf{X}_n^\intercal K_n \mathbf{Y}_n + s_{22} \mathbf{Y}_n^\intercal K_n \mathbf{Y}_n \\
&= s_{11}\, \widetilde{\mathbf{X}}_n^\intercal \operatorname{diag}(\lambda_2, \lambda_3, \ldots, \lambda_n)\, \widetilde{\mathbf{X}}_n + 2 s_{12}\, \widetilde{\mathbf{X}}_n^\intercal \operatorname{diag}(\lambda_2, \lambda_3, \ldots, \lambda_n)\, \widetilde{\mathbf{Y}}_n \\
&\quad + s_{22}\, \widetilde{\mathbf{Y}}_n^\intercal \operatorname{diag}(\lambda_2, \lambda_3, \ldots, \lambda_n)\, \widetilde{\mathbf{Y}}_n \\
&= s_{11} \sum_{j=2}^n \lambda_j \widetilde{X}_j^2 + 2 s_{12} \sum_{j=2}^n \lambda_j \widetilde{X}_j \widetilde{Y}_j + s_{22} \sum_{j=2}^n \lambda_j \widetilde{Y}_j^2 \\
&= \sum_{j=2}^n \lambda_j \left(s_{11} \widetilde{X}_j^2 + 2 s_{12} \widetilde{X}_j \widetilde{Y}_j + s_{22} \widetilde{Y}_j^2 \right).
\end{aligned}$$

By independence of $\widetilde{X}_j$ and $\widetilde{Y}_k$ for $j, k \in \{2, 3, \ldots, n\}$,

$$\begin{aligned}
\phi_n(s_{11}, s_{12}, s_{22}) &= E\left[\exp\left\{-\frac{1}{2}\left(s_{11} Z_{11}^n + 2 s_{12} Z_{12}^n + s_{22} Z_{22}^n\right)\right\}\right] \\
&= \prod_{j=2}^n E\left[\exp\left\{-\frac{1}{2}\lambda_j \left(s_{11} \widetilde{X}_j^2 + 2 s_{12} \widetilde{X}_j \widetilde{Y}_j + s_{22} \widetilde{Y}_j^2\right)\right\}\right] \\
&= \prod_{j=2}^n \left(1 + (s_{11} + s_{22})\lambda_j + (s_{11} s_{22} - s_{12}^2)\lambda_j^2\right)^{-1/2} \qquad (12)
\end{aligned}$$

Note that in line (12) a standard expression for the mgf of a linear-quadratic functional of a normal variable has been used (iteratively twice), and, further, the independence of $\widetilde{X}_j$ and $\widetilde{Y}_j$ has been employed. Further note that in line (12) the conditions $s_{11}, s_{22} \geq 0$ and $s_{12}^2 \leq s_{11} s_{22}$ ensure the applicability of the standard expression for the mgf of a linear-quadratic functional of a bivariate random vector. Factorizing the quadratic polynomial in line (12) yields

$$\begin{aligned}
\phi_n(s_{11}, s_{12}, s_{22}) &= \prod_{j=2}^n \left((1 - \alpha\lambda_j)(1 - \beta\lambda_j)\right)^{-1/2} \\
&= \left(\prod_{j=2}^n (1 - \alpha\lambda_j) \prod_{j=2}^n (1 - \beta\lambda_j)\right)^{-1/2} \\
&= (d_n(\alpha)\, d_n(\beta))^{-1/2}, \qquad (13)
\end{aligned}$$

The last equality holds by the representation of the alternative characteristic polynomial of K_n by the eigenvalues of K_n, see (3). Combining (9), (10), and (11) allows us to represent the joint mgf $\phi_n(s_{11}, s_{12}, s_{22})$ explicitly in terms of $d_n(\lambda)$, $\alpha(s_{11}, s_{12}, s_{22})$ and $\beta(s_{11}, s_{12}, s_{22})$, as given in (13). This completes the proof. □

4. Moments of θ_n

In the previous section, we provided an exact representation for the joint trivariate mgf ϕ_n. In this section, we use it to calculate the moments of θ_n by a method provided by Ernst et al. (see Proposition 1 in [7]), which we cite as follows:

Proposition 1 (*Ernst et al. (2019)*). *For $m = 0, 1, 2, \ldots$, we have*

$$E\left(\theta_n^m\right) = \frac{(-1)^m}{2^m \Gamma(m/2)^2} \int_0^\infty \int_0^\infty s_{11}^{m/2-1} s_{22}^{m/2-1} \frac{\partial^m \phi_n}{\partial s_{12}^m}(s_{11}, 0, s_{22})\, ds_{11} ds_{22}. \tag{14}$$

An immediate application of this proposition yields that the second moment of θ_n is given by the following double Riemann integral:

$$E\left(\theta_n^2\right) = \frac{1}{4} \int_0^\infty \int_0^\infty \frac{\partial^2 \phi_n}{\partial s_{12}^2}(s_{11}, 0, s_{22})\, ds_{11} ds_{22}. \tag{15}$$

4.1. Explicit formula for the second moments of θ_n

We now calculate the integrand in the previous integral representation explicitly, yielding the next theorem. This theorem gives a closed-form expression for the second moment of θ_n for any n.

Theorem 4.1. *The second moment of θ_n is*

$$\begin{aligned}
&E\left(\theta_n^2\right)\\
&= -\frac{1}{4} \int_0^\infty \int_0^\infty \frac{(s_{11}+2n)(s_{22}+2n)+4n^2}{[s_{11}s_{22}(s_{11}+4n)(s_{22}+4n)]^{3/4}} \left[f(s_{11}/n)^n - f(s_{11}/n)^{-n}\right]^{-1/2}\\
&\qquad \times \left[f(s_{22}/n)^n - f(s_{22}/n)^{-n}\right]^{-1/2} ds_{11} ds_{22}\\
&\quad + \frac{1}{4} \int_0^\infty \int_0^\infty \left\{ \frac{n(s_{11}+s_{22}+4n)}{\sqrt{s_{11}^2+4ns_{11}} + \sqrt{s_{22}^2+4ns_{22}}} \cdot \left[f(s_{11}/n)^n f(s_{22}/n)^n \right.\right.\\
&\qquad \left. - f(s_{11}/n)^{-n} f(s_{22}/n)^{-n}\right] + \frac{1}{2}\left(\sqrt{s_{11}^2+4ns_{11}} + \sqrt{s_{22}^2+4ns_{22}} + s_{11} + s_{22} + 4n\right)\\
&\qquad \left. \cdot \frac{f(s_{11}/n)^n f(s_{22}/n)^{-n} - f(s_{22}/n)^n f(s_{11}/n)^{-n}}{f(s_{11}/n) - f(s_{22}/n)} \right\} \cdot \left[f(s_{11}/n)^n - f(s_{11}/n)^{-n}\right]^{-3/2}\\
&\qquad \cdot \left[f(s_{22}/n)^n - f(s_{22}/n)^{-n}\right]^{-3/2} \cdot [s_{11}s_{22}(s_{11}+4n)(s_{22}+4n)]^{-1/4}\, ds_{11} ds_{22},
\end{aligned} \tag{16}$$

where

$$f(\lambda) := \frac{(\lambda+2) + \sqrt{(\lambda+2)^2 - 4}}{2}. \tag{17}$$

Proof. It is sufficient to provide the announced closed form expression for $\frac{\partial^2 \phi_n}{\partial s_{12}^2}(s_{11}, 0, s_{22})$. Recalling the definition of $f(\lambda)$ in (17), straightforward calculation yields

$$d_n(-\lambda) = \frac{1}{n\sqrt{(\lambda/n+2)^2 - 4}} \left[f(\lambda/n)^n - f(\lambda/n)^{-n}\right], \tag{18}$$

$$\begin{aligned}
d_n'(-\lambda) = &\frac{1}{n^2} \frac{\lambda/n + 2}{\left[(\lambda/n+2)^2 - 4\right]^{3/2}} \left[f(\lambda/n)^n - f(\lambda/n)^{-n}\right]\\
&- \frac{1}{n} \frac{1}{(\lambda/n+2)^2 - 4} \left[f(\lambda/n)^n + f(\lambda/n)^{-n}\right],
\end{aligned} \tag{19}$$

Table 1
Numerical results for the second moment of θ_n for varying values of n.

n	2	5	10	20	50	100
$E\left(\theta_n^2\right)$	1.000000	0.341109	0.265140	0.246645	0.241501	0.240767
n	200	500	1000	2000	5000	∞
$E\left(\theta_n^2\right)$	0.240584	0.240532	0.240525	0.240523	0.240523	0.240523

and

$$\begin{aligned}\frac{\partial^2\phi_n}{\partial s_{12}^2} &= \frac{3}{4}\big(d_n(\alpha)d_n(\beta)\big)^{-5/2}\left(d_n'(\alpha)\,d_n(\beta)\,\frac{\partial\alpha}{\partial s_{12}}+d_n(\alpha)\,d_n'(\beta)\,\frac{\partial\beta}{\partial s_{12}}\right)^2\\ &\quad-\frac{1}{2}\big(d_n(\alpha)d_n(\beta)\big)^{-3/2}\sum_{j=0}^{2}\binom{2}{j}d_n^{(j)}(\alpha)\,d_n^{(2-j)}(\beta)\left(\frac{\partial\alpha}{\partial s_{12}}\right)^j\left(\frac{\partial\beta}{\partial s_{12}}\right)^{2-j}\\ &\quad-\frac{1}{2}\big(d_n(\alpha)d_n(\beta)\big)^{-3/2}\left(d_n'(\alpha)\,d_n(\beta)\,\frac{\partial^2\alpha}{\partial s_{12}^2}+d_n(\alpha)\,d_n'(\beta)\,\frac{\partial^2\beta}{\partial s_{12}^2}\right),\end{aligned}\tag{20}$$

with α and β as defined in (10) and (11). Note that

$$\begin{aligned}-\frac{\partial\alpha}{\partial s_{12}} &= \frac{\partial\beta}{\partial s_{12}} = \frac{2s_{12}}{\sqrt{(s_{11}-s_{22})^2+4s_{12}^2}}\\ -\frac{\partial^2\alpha}{\partial s_{12}^2} &= \frac{\partial^2\beta}{\partial s_{12}^2} = \frac{2(s_{11}-s_{22})^2}{\left[(s_{11}-s_{22})^2+4s_{12}^2\right]^{3/2}}.\end{aligned}$$

It follows easily that $\alpha(s_{11},0,s_{22}) = -\max(s_{11},s_{22})$, $\beta(s_{11},0,s_{22}) = -\min(s_{11},s_{22})$, and

$$\begin{aligned}-\frac{\partial\alpha}{\partial s_{12}}(s_{11},0,s_{22}) &= \frac{\partial\beta}{\partial s_{12}}(s_{11},0,s_{22}) = 0\\ -\frac{\partial^2\alpha}{\partial s_{12}^2}(s_{11},0,s_{22}) &= \frac{\partial^2\beta}{\partial s_{12}^2}(s_{11},0,s_{22}) = 2|s_{11}-s_{22}|^{-1}.\end{aligned}$$

Plugging the above results into (20) yields

$$\frac{\partial^2\phi_n}{\partial s_{12}^2}(s_{11},0,s_{22}) = \frac{\frac{d_n'(-\max(s_{11},s_{22}))}{d_n(-\max(s_{11},s_{22}))}-\frac{d_n'(-\min(s_{11},s_{22}))}{d_n(-\min(s_{11},s_{22}))}}{\left[d_n(-\max(s_{11},s_{22}))d_n(-\min(s_{11},s_{22}))\right]^{1/2}|s_{11}-s_{22}|}.\tag{21}$$

Combining (15), (18), (19) and (21) and performing straightforward calculations, we arrive at an explicit formula for $\frac{\partial^2\phi_n}{\partial s_{12}^2}(s_{11},0,s_{22})$ as well as the double integral expression for the second moment of θ_n for all n given in the statement of the theorem. □

4.2. Numerics

We now turn to numerics. Mathematica allows us to calculate the second moment of Yule's "nonsense correlation" θ_n for any given n. The numerical results are summarized in Table 1.

For higher-order moments as represented in (14), we can use Mathematica to perform symbolic high-order differentiation and then the two dimensional integration, thereby implicitly calculating higher moments of θ_n for all n. The numerical results of some higher-order moments of θ_{50} are summarized in Table 2.

Table 2
Numerical results for higher-order moments of θ_{50}.

k	2	4	6	8
$E\left(\theta_{50}^k\right)$	0.241501	0.109961	0.061465	0.038257
k	10	12	14	16
$E\left(\theta_{50}^k\right)$	0.025485	0.017803	0.012885	0.009586

5. Convergence in wasserstein distance

Tables 1 and 2 in the previous section give us insight into the behavior of the distribution of θ_n for large n, as it approximates the distribution of its limit θ defined below in (22). In Table 1, we note the rather rapid convergence of $E\left(\theta_n^2\right)$ as $n \to \infty$. We observe that this convergence occurs at a rate which appears to be faster than n^{-1}. This encouraged us to investigate the rate of convergence of θ_n to θ. In this section, we establish an upper bound for the Wasserstein distance between θ_n and θ, which comes from a coupling of θ_n and θ on the same probability space Ω, in which the convergence occurs in $L^1(\Omega)$.

Let W_1 and W_2 be two independent Wiener processes. Then Yule's "nonsense correlation" is given by (see [8])

$$\theta = \frac{\int_0^1 W_1(t)W_2(t)dt - \int_0^1 W_1(t)dt \int_0^1 W_2(t)dt}{\sqrt{\int_0^1 W_1^2(t)dt - \left(\int_0^1 W_1(t)dt\right)^2}\sqrt{\int_0^1 W_2^2(t)dt - \left(\int_0^1 W_2(t)dt\right)^2}}. \tag{22}$$

If X, Y are two real-valued random variables, recall that the Wasserstein distance between the law of X and the law of Y is given by

$$d_W(X,Y) := \sup_{f \in \mathrm{Lip}(1)} |Ef(X) - Ef(Y)|,$$

where Lip(1) is the set of all Lipschitz functions with Lipschitz constant ≤ 1. Our key result (Theorem 5.5) regarding the convergence of θ_n to θ is as follows:

$$d_W(\theta_n, \theta) = \mathcal{O}\left(\frac{1}{n}\right). \tag{23}$$

We shall prove the claimed result by showing that $E\left[|\theta_n - \theta|\right] = \mathcal{O}\left(\frac{1}{n}\right)$ under a natural coupling of θ_n and θ.

The reader will find some heuristic comments regarding how this result arises, and what more could be expected for other processes, at the end of the next subsection, which provides the preparatory setup needed to prove Theorem 5.5.

5.1. Notation, coupling, extensions and implications

Define $M(s,t) := \min(s,t) - st$. For every $n \in \mathbb{N}_+$, define

$$M_n(s,t) := \sum_{1 \leq j,k \leq n} M\left(\frac{j-1}{n}, \frac{k-1}{n}\right) \mathbb{1}_{\{(j-1)/n < s \leq j/n\}} \mathbb{1}_{\{(k-1)/n < t \leq k/n\}}. \tag{24}$$

For every $n \in \mathbb{N}_+$, define

$$A_n = \int_0^1 \int_0^t M_n(s,t)\, dW_1(s)\, dW_2(t) + \int_0^1 \int_0^s M_n(s,t)\, dW_2(t)\, dW_1(s),$$
$$B_n = 2\int_0^1 \int_0^t M_n(s,t)\, dW_1(s)\, dW_1(t) + \frac{1}{n}\sum_{j=1}^n M\left(\frac{j-1}{n}, \frac{j-1}{n}\right),$$
$$C_n = 2\int_0^1 \int_0^t M_n(s,t)\, dW_2(s)\, dW_2(t) + \frac{1}{n}\sum_{j=1}^n M\left(\frac{j-1}{n}, \frac{j-1}{n}\right).$$

Further, let

$$A = \int_0^1 \int_0^t M(s,t)\, dW_1(s)\, dW_2(t) + \int_0^1 \int_0^s M(s,t)\, dW_2(t)\, dW_1(s),$$
$$B = 2\int_0^1 \int_0^t M(s,t)\, dW_1(s)\, dW_1(t) + \int_0^1 M(t,t)\, dt,$$
$$C = 2\int_0^1 \int_0^t M(s,t)\, dW_2(s)\, dW_2(t) + \int_0^1 M(t,t)\, dt.$$

A key point here, as emphasized in the introduction, is that we choose to use the same pair (W_1, W_2) of independent Wiener processes to represent all six of these variables. This is a natural coupling on the common Wiener space Ω defined by this pair, which allows us to relate the two empirical correlations to each other in a way that easily yields the Wasserstein distance between their distributions. In particular, in Section 5.2, we will show $\theta_n \overset{D}{=} A_n/\sqrt{B_n C_n}$ and $\theta = A/\sqrt{BC}$, while in the first step in the proof of Theorem 5.5 in Section 5.3, we establish that $d_W(\mathcal{L}(X), \mathcal{L}(Y)) \le E[|X - Y|]$ for any pair of integrables rv's (X, Y) on the same probability space, from which we conclude

$$d_W(\theta_n, \theta) = d_W\left(\mathcal{L}\left(\frac{A_n}{\sqrt{B_n C_n}}\right), \mathcal{L}\left(\frac{A}{\sqrt{BC}}\right)\right) \le E\left[\left|\frac{A_n}{\sqrt{B_n C_n}} - \frac{A}{\sqrt{BC}}\right|\right].$$

In the sequel, we shall restrict our attention to A_n, B_n, C_n, A, B and C. One key reason for defining A_n, B_n, C_n, A, B and C is that, being defined as second-chaos variable plus a constant (which may be 0), the upper bounds for the second moments of $A_n - A$, $B_n - B$ and $C_n - C$ can be estimated as $\mathcal{O}(n^{-2})$. Hence, A_n, B_n and C_n converge in $L^2(\Omega)$ to A, B and C respectively at rate $\sqrt{n^{-2}}$. In fact, the second moments of $A_n - A$, $B_n - B$ and $C_n - C$ can be calculated explicitly. These details will be stated and proved in Section 5.2.

This convergence rate $\mathcal{O}(n^{-2})$ converts into the rate $\mathcal{O}(n^{-1})$ in Theorem 5.5 because of the need to separate numerator from denominator. One might view this as the cost to pay for this conversion. However, we see it as more fruitful to view the rate of convergence at the level of norms, which preserve scales: the $\mathcal{O}(n^{-1})$ is the rate of convergence of the three elements constituting θ_n in $L^2(\Omega)$-norm. This leads us to presume this Wasserstein-distance rate of convergence is sharp, although it is beyond the scope of this paper to establish this rigorously. From the so-called property of hyper-contractivity on fixed Wiener chaos (see [13] Chapter 2), for all $p > 1$, all $L^p(\Omega)$-norms of the three differences $A_n - A$, $B_n - B$, $C_n - C$ are equivalent, making it unnecessary to speculate whether computing the convergence rates of any specific higher moment might provide additional insight. Expanding the ratios defining $\theta_n - \theta$ into tri-variate Taylor series did not lead us to any further insight based on those explicit norm-equivalence universal constants.

We now turn to the following question: is the rate $\mathcal{O}(n^{-1})$ for $d_W(\theta_n, \theta)$, which is inherited from the rate of $\mathcal{O}(n^{-2})$ for $Var(A_n - A)$, $Var(B_n - B)$, and $Var(C_n - C)$, generic, or is it specific to random walks? Resolving this question rigorously is also beyond the scope of this paper, but our preliminary calculations indicate that the aforementioned $\mathcal{O}(n^{-2})$ only holds because of the property of independence of increments of a random walk defined as a partial sum of a sequence of independent terms.

We believe that for other Markov chains which might converge in law, and specifically for any reasonable discretization of processes which are far from having independent increments, such as long-memory processes or mean-reverting processes, the rate of convergence to 0 of $Var(A_n - A)$, $Var(B_n - B)$, and $Var(C_n - C)$ is $\mathcal{O}(n^{-1})$. Using a simple polarization argument, the rates of convergence to 0 of these three differences should be essentially equivalent, so that looking at merely one of them would give the order for all of them. It should also be noted that in some mean-reverting and/or stationary cases, like the AR(1) process or the discretely observed Ornstein–Uhlenbeck process (see, e.g., [19]), ρ_n converges to 0, and that the empirical correlation for pairs of independent stationary processes does converge to 0, not to a "diffuse" limit ρ, where "diffuse" refers to the distribution being widely dispersed and frequently large in absolute value. In that sense, for these processes, ρ_n and ρ are not "nonsense" correlations, since they correctly converge to 0 under the assumption of independence of the two paths. As mentioned elsewhere (e.g. Section 6), the method of proof below to establish the rate $\mathcal{O}(n^{-2})$ involves direct calculation, but the same rate can also be established using a more generic, less precise calculation where one compares Riemann integrals to their approximations using step functions. When attempting that calculation, the property of independence of increments comes plainly into view, implying a number of cancellations much like what one observes when computing the quadratic variation of a martingale. This same methodology seems to indicate that no such cancellations occur for non-independent-increment cases, but that our conjectured rate $\mathcal{O}(n^{-1})$ is straightforward to establish for other Gaussian processes, using the same type of coupling as for Gaussian random walks. Extending the conjecture to non-Gaussian processes would require more work, and would indeed fall outside the scope of the present paper.

The distinction which we conjecture above between Gaussian random walks and other Gaussian processes may indeed be of important to statistical practitioners. It means that the use of the properties of the continuous-time limit of Yule's "nonsense correlation" θ, which are straightforward to establish using simulations, to infer statistical properties of discrete-time random walks, is legitimate for moderate sample sizes, but not so if the data does not behave like the path of a random walk with independent increments. For instance, a statistic on θ that relates to the construction of the Wasserstein metric (e.g. a mean value or a moment) can be presumed generically accurate at a 1% level for a Gaussian random walk with several hundred data points, while tens of thousands of data points would be needed, according to our conjecture, when working with a mean-reverting stationary time series, to exploit the "non-nonsense" zero limit of θ_n. In the environmental sciences, where such time series are ubiquitous, and where many have yearly frequencies, no such reliance on θ directly can be assumed on a historical scale. In other application domains, such as in quantitative finance, high-frequency studies over several years, such as when studying the long-term distribution and movements of interest rates or of market volatility, can routinely draw on enough data points, however. In financial markets, shorter-term studies of other objects, such as stock returns, relate more readily to Gaussian random walks, where our results herein indicate that only hundreds of measurements over time would allow the use of θ's law instead of needing to rely on θ_n. For random-walk time series which are shorter yet, our explicit results on θ_n from Section 4 are available.

5.2. Properties of A_n, B_n, C_n, A, B, and C

In this section, we derive several properties of A_n, B_n, C_n, A, B, and C, including justifying the coupling, some convenient a.s. constraints, explicit formulae for univariate moment-generating functions, and most importantly from the standpoint of our analysis, the last two propositions in this section provide the aforementioned convergences to zero of the variances of the differences between the approximating and limiting three elements constituting θ_n and θ.

Proposition 2. *The following statements hold, where the equality in (a) and the first equality in (c) are in distribution:*

(a) $(Z_{11}^n/n, Z_{12}^n/n, Z_{22}^n/n) \stackrel{D}{=} (B_n, A_n, C_n)$ *for every* $n \in \mathbb{N}_+$*;*
(b)

$$A = \int_0^1 W_1(t)W_2(t)dt - \int_0^1 W_1(t)dt \int_0^1 W_2(t)dt, \tag{25}$$

$$B = \int_0^1 W_1^2(t)dt - \left(\int_0^1 W_1(t)dt\right)^2, \tag{26}$$

$$C = \int_0^1 W_2^2(t)dt - \left(\int_0^1 W_2(t)dt\right)^2. \tag{27}$$

(c) $\theta_n \stackrel{D}{=} A_n/\sqrt{B_n C_n}$ *and* $\theta = A/\sqrt{BC}$.

Proof. See Appendix. □

A helpful corollary of Proposition 2 is as follows.

Corollary 1. *(a)* $|A_n/\sqrt{B_n C_n}| \le 1$ *a.s.; (b)* $|A/\sqrt{BC}| \le 1$ *a.s.; (c)* $B_n, C_n > 0$ *a.s. for* $n \ge 2$*; (d)* $B, C > 0$ *a.s.*

Proof. By Cauchy–Schwarz, $|Z_{12}^n/\sqrt{Z_{11}^n Z_{22}^n}| \le 1$. By Proposition 2, $(Z_{11}^n/n, Z_{12}^n/n, Z_{22}^n/n) \stackrel{D}{=} (B_n, A_n, C_n)$, and so $|Z_{12}^n/\sqrt{Z_{11}^n Z_{22}^n}| \stackrel{D}{=} A_n/\sqrt{B_n C_n}$. Statement (a) thus follows. Similarly, statement (b) follows from Cauchy–Schwarz and from Proposition 2.

The non-negativity of the terms in statements (c) and (d) comes from Proposition 2 and Jensen's inequality. For $n \ge 2$, $Z_{11}^n/n = 0$ implies $X_1 = X_2 = \cdots = X_n$, where $X_1, X_2, \ldots, X_n$ are independent standard Gaussian random variables as defined in Section 1. We immediately note that the probability of the event $\{X_1 = X_2 = \cdots = X_n\}$ is 0. Thus, $Z_{11}^n/n > 0$ a.s., and hence $B_n > 0$ a.s. Similarly, $C_n > 0$ a.s. This proves statement (c). Finally, by Proposition 2, $B = 0$ implies that $W_1(t)$ is a constant on the interval [0, 1], the probability of which is 0. Hence $B > 0$ a.s. and similarly, $C > 0$ a.s. This proves statement (d). □

Let $\phi_B(s) := E[e^{-sB}]$ and $\phi_C(s) := E[e^{-sC}]$. $\phi_B(s)$ and $\phi_C(s)$ are Laplace–Stieltjes transforms of B and C, respectively. However, to be consistent with our definition of joint mgf ϕ_n, we will call $\phi_B(s)$ and $\phi_C(s)$ moment generating functions (mgfs) in the remainder of this paper. The only difference between the Laplace–Stieltjes transform and the mgf is the sign before s. Similarly, let $\phi_{B_n}(s)$ and $\phi_{C_n}(s)$ be the mgfs of B_n and C_n respectively. These functions can be computed explicitly, as the following lemma shows.

Lemma 5.1. *We have*

$$\phi_{B_n}(s) = \phi_{C_n}(s) = (d_n(-2s/n))^{-1/2}, \quad \text{for every } n \in \mathbb{N}_+,$$
$$\phi_B(s) = \phi_C(s) = \left(\frac{\sinh\sqrt{2s}}{\sqrt{2s}}\right)^{-1/2}.$$

Proof. By Theorem 3.2, the joint mgf of $(Z_{11}^n, Z_{12}^n, Z_{22}^n)$ is

$$E\left[\exp\left\{-\frac{1}{2}(s_{11}Z_{11}^n + 2s_{12}Z_{12}^n + s_{22}Z_{22}^n)\right\}\right] = (d_n(\alpha(s_{11}, s_{12}, s_{22}))d_n(\beta(s_{11}, s_{12}, s_{22})))^{-1/2}.$$

Plugging $s_{11} = 2s/n$, $s_{12} = 0$ and $s_{22} = 0$ into the last display, it follows that

$$E[e^{-s Z_{11}^n/n}] = (d_n(-2s/n)d_n(0))^{-1/2} = (d_n(-2s/n))^{-1/2},$$

where the last equality comes from the fact that $d_n(0) = 1$. Note that since (by Proposition 2) $Z_{11}^n/n \overset{D}{=} B_n$,

$$\phi_{B_n}(s) = E[e^{-sB_n}] = E[e^{-s Z_{11}^n/n}] = (d_n(-2s/n))^{-1/2}.$$

Symmetrically, $\phi_{C_n}(s) = (d_n(-2s/n))^{-1/2}$.

Combining the results of Section 4.1 of [7] with Proposition 2, the joint mgf of (A, B, C) is

$$E\left[\exp\left\{-\frac{1}{2}(s_{11}B + 2s_{12}A + s_{22}C)\right\}\right] = \left(\frac{\sinh\sqrt{-\alpha}\sinh\sqrt{-\beta}}{\sqrt{-\alpha}\sqrt{-\beta}}\right)^{-1/2},$$

where α and β are defined in (10) and (11). Plugging $s_{11} = 2s$, $s_{12} = 0$ and $s_{22} = 0$ into the last display, and recalling that $\sinh x/x$ equals 1 at $x = 0$, it follows that the mgf ϕ_B of B is

$$\phi_B(s) = E[e^{-sB}] = \left(\frac{\sinh\sqrt{2s}}{\sqrt{2s}}\right)^{-1/2}.$$

Symmetrically, $\phi_C(s) = (\sinh\sqrt{2s}/\sqrt{2s})^{-1/2}$. □

We now proceed to give upper bounds for the second moments of $A_n - A$, $B_n - B$ and $C_n - C$.

Proposition 3. *For $n > 2$, we have $E[(A_n - A)^2] = \frac{5}{72}n^{-2} - \frac{7}{120}n^{-4}$. Hence, $E[(A_n - A)^2] \le \frac{5}{72}n^{-2}$ for $n > 2$.*

Proof. From the definition (24), a routine calculation shows that M is a sublinear function in both its variables, with Lipschitz constant 1: $|M(s_2, t_2) - M(s_1, t_1)| \le \max(s_2 - s_1, t_2 - t_1)$, for $0 \le s_1 \le s_2 \le 1$ and $0 \le t_1 \le t_2 \le 1$. It follows immediately that

$$|M_n(s, t) - M(s, t)| \le \frac{1}{n}.$$

By the definitions of A_n and A,

$$A_n - A = \int_0^1 \int_0^t (M_n(s,t) - M(s,t))\, dW_1(s)\, dW_2(t) + \int_0^1 \int_0^s (M_n(s,t) - M(s,t))\, dW_2(t)\, dW_1(s).$$

By Jensen's inequality,

$$(A_n - A)^2 \le 2\left(\int_0^1 \int_0^t (M_n(s,t) - M(s,t))\, dW_1(s)\, dW_2(t)\right)^2 + 2\left(\int_0^1 \int_0^s (M_n(s,t) - M(s,t))\, dW_2(t)\, dW_1(s)\right)^2.$$

Taking expectations on both sides yields

$$\begin{aligned} & E[(A_n - A)^2] \\ \le\ & 2E\left[\left(\int_0^1 \int_0^t (M_n(s,t) - M(s,t))\, dW_1(s)\, dW_2(t)\right)^2\right] \\ & + 2E\left[\left(\int_0^1 \int_0^s (M_n(s,t) - M(s,t))\, dW_2(t)\, dW_1(s)\right)^2\right] \\ =\ & 2\int_0^1 \int_0^t E\left[(M_n(s,t) - M(s,t))^2\right] ds\, dt + 2\int_0^1 \int_0^s E\left[(M_n(s,t) - M(s,t))^2\right] dt\, ds \\ \le\ & 2\int_0^1 \int_0^t \frac{1}{n^2}\, ds\, dt + 2\int_0^1 \int_0^s \frac{1}{n^2}\, dt\, ds = \frac{2}{n^2}, \end{aligned}$$

where in the first equality the Itô isometry has been applied. We refer to [4, Theorem 2.3] or [15, Proposition 2.1.5] for references on the Itô isometry. In the last display, letting $n \to \infty$ yields $\lim_{n\to\infty} E[(A_n - A)^2] = 0$. Hence, $\lim_{n\to\infty} E[A_n^2] = E[A^2]$.

By (68) in the Appendix, we have

$$A_n = \sum_{j,k=1}^{n} M\left(\frac{j-1}{n}, \frac{k-1}{n}\right)\left(W_1\left(\frac{j}{n}\right) - W_1\left(\frac{j-1}{n}\right)\right)\left(W_2\left(\frac{k}{n}\right) - W_2\left(\frac{k-1}{n}\right)\right).$$

Then, by the independence of the Wiener processes W_1 and W_2,

$$\begin{aligned} & E[A_n^2] \\ =\ & E\Bigg[\sum_{j,k,i,l=1}^{n} M\left(\frac{j-1}{n}, \frac{k-1}{n}\right) M\left(\frac{i-1}{n}, \frac{l-1}{n}\right)\left(W_1\left(\frac{j}{n}\right) - W_1\left(\frac{j-1}{n}\right)\right) \\ & \left(W_2\left(\frac{k}{n}\right) - W_2\left(\frac{k-1}{n}\right)\right)\left(W_1\left(\frac{i}{n}\right) - W_1\left(\frac{i-1}{n}\right)\right)\left(W_2\left(\frac{l}{n}\right) - W_2\left(\frac{l-1}{n}\right)\right)\Bigg] \\ =\ & \sum_{j,k,i,l=1}^{n} \Bigg\{ M\left(\frac{j-1}{n}, \frac{k-1}{n}\right) M\left(\frac{i-1}{n}, \frac{l-1}{n}\right) \\ & \times E\left[\left(W_1\left(\frac{j}{n}\right) - W_1\left(\frac{j-1}{n}\right)\right)\left(W_1\left(\frac{i}{n}\right) - W_1\left(\frac{i-1}{n}\right)\right)\right] \end{aligned}$$

$$\left.\times E\left[\left(W_2\left(\frac{k}{n}\right)-W_2\left(\frac{k-1}{n}\right)\right)\left(W_2\left(\frac{l}{n}\right)-W_2\left(\frac{l-1}{n}\right)\right)\right]\right\}$$
$$=\sum_{j,k,i,l=1}^{n} M\left(\frac{j-1}{n},\frac{k-1}{n}\right)M\left(\frac{i-1}{n},\frac{l-1}{n}\right)\cdot\frac{1}{n}\mathbb{1}_{\{i=j\}}\cdot\frac{1}{n}\mathbb{1}_{\{l=k\}}$$
$$=\sum_{j,k=1}^{n} M\left(\frac{j-1}{n},\frac{k-1}{n}\right)^2\cdot\frac{1}{n^2}$$
$$=2\sum_{j=1}^{n}\sum_{k=1}^{j-1}\frac{1}{n^2}\left(\frac{k-1}{n}-\frac{(j-1)(k-1)}{n^2}\right)^2+\sum_{j=1}^{n}\frac{1}{n^2}\left(\frac{j-1}{n}-\frac{(j-1)^2}{n^2}\right)^2, \tag{28}$$

Note that the first term on the right-hand side of (28) is twice a double summation of a polynomial of j and k, calculating the summation with respect to k by Faulhaber's formula (see [16]) yields twice a single summation of a polynomial of j over $j=1,\ldots,n$. Applying Faulhaber's formula again to this summation, we have that the first term on the right-hand side of (28) is equal to

$$\frac{1}{90}-\frac{1}{30}n^{-1}+\frac{1}{36}n^{-2}-\frac{7}{180}n^{-4}+\frac{1}{30}n^{-5}. \tag{29}$$

Note that the second term on the right-hand side of (28) is a single summation of a polynomial of j. Applying Faulhaber's formula again gives

$$\frac{1}{30}n^{-1}-\frac{1}{30}n^{-5}. \tag{30}$$

Combining (28), (29) and (30), we have

$$E[A_n^2]=\frac{1}{90}+\frac{1}{36}n^{-2}-\frac{7}{180}n^{-4}. \tag{31}$$

In the last display, letting $n\to\infty$ yields

$$E[A^2]=\lim_{n\to\infty}E[A_n^2]=\frac{1}{90}. \tag{32}$$

In what follows, we proceed to calculate the expectation of A_nA, which, of course, is handy to compute the variance of A_n-A. By Fubini's theorem and the independence of W_1 and W_2,

$$E\left[\left(W_1\left(\frac{j}{n}\right)-W_1\left(\frac{j-1}{n}\right)\right)\left(W_2\left(\frac{k}{n}\right)-W_2\left(\frac{k-1}{n}\right)\right)\int_0^1 W_1(t)W_2(t)\,dt\right]$$
$$=\int_0^1 E\left[\left(W_1\left(\frac{j}{n}\right)-W_1\left(\frac{j-1}{n}\right)\right)\left(W_2\left(\frac{k}{n}\right)-W_2\left(\frac{k-1}{n}\right)\right)W_1(t)W_2(t)\right]dt$$
$$=\int_0^1 E\left[\left(W_1\left(\frac{j}{n}\right)-W_1\left(\frac{j-1}{n}\right)\right)W_1(t)\right]E\left[\left(W_2\left(\frac{k}{n}\right)-W_2\left(\frac{k-1}{n}\right)\right)W_2(t)\right]dt$$
$$=\int_0^1\left(t\wedge\left(\frac{j}{n}\right)-t\wedge\left(\frac{j-1}{n}\right)\right)\cdot\left(t\wedge\left(\frac{k}{n}\right)-t\wedge\left(\frac{k-1}{n}\right)\right)dt$$
$$=\frac{1}{n^2}+\frac{1}{2}\frac{1}{n^3}-\frac{j\vee k}{n^3}-\frac{1}{6}\frac{1}{n^3}\mathbb{1}_{\{j=k\}}.$$

Similarly,

$$
\begin{aligned}
&E\left[\left(W_1\left(\frac{j}{n}\right)-W_1\left(\frac{j-1}{n}\right)\right)\left(W_2\left(\frac{k}{n}\right)-W_2\left(\frac{k-1}{n}\right)\right)\int_0^1 W_1(t)\,dt\int_0^1 W_1(t)\,dt\right]\\
&=\int_0^1\int_0^1 E\left[\left(W_1\left(\frac{j}{n}\right)-W_1\left(\frac{j-1}{n}\right)\right)\left(W_2\left(\frac{k}{n}\right)-W_2\left(\frac{k-1}{n}\right)\right)W_1(t)W_2(s)\right]ds\,dt\\
&=\int_0^1\int_0^1 E\left[\left(W_1\left(\frac{j}{n}\right)-W_1\left(\frac{j-1}{n}\right)\right)W_1(t)\right]E\left[\left(W_2\left(\frac{k}{n}\right)-W_2\left(\frac{k-1}{n}\right)\right)W_2(s)\right]ds\,dt\\
&=\int_0^1\int_0^1\left(t\wedge\left(\frac{j}{n}\right)-t\wedge\left(\frac{j-1}{n}\right)\right)\cdot\left(s\wedge\left(\frac{k}{n}\right)-s\wedge\left(\frac{k-1}{n}\right)\right)ds\,dt\\
&=\left(\frac{1}{n}+\frac{1}{2}\frac{1}{n^2}-\frac{j}{n^2}\right)\cdot\left(\frac{1}{n}+\frac{1}{2}\frac{1}{n^2}-\frac{k}{n^2}\right).
\end{aligned}
$$

Combining the last two displays and by linearity of expectation, we have

$$
\begin{aligned}
&E[A_nA]\\
&=\sum_{j,k=1}^{n}M\left(\frac{j-1}{n},\frac{k-1}{n}\right)\left[\frac{1}{n^2}+\frac{1}{2}\frac{1}{n^3}-\frac{j\vee k}{n^3}-\frac{1}{6}\frac{1}{n^3}\mathbb{1}_{\{j=k\}}\right.\\
&\qquad\left.-\left(\frac{1}{n}+\frac{1}{2}\frac{1}{n^2}-\frac{j}{n^2}\right)\cdot\left(\frac{1}{n}+\frac{1}{2}\frac{1}{n^2}-\frac{k}{n^2}\right)\right]\\
&=2\sum_{j=1}^{n}\sum_{k=1}^{j-1}\left(\frac{k-1}{n}-\frac{j-1}{n}\cdot\frac{k-1}{n}\right)\left[\frac{1}{n^2}+\frac{1}{2}\frac{1}{n^3}-\frac{j}{n^3}\right.\\
&\qquad\left.-\left(\frac{1}{n}+\frac{1}{2}\frac{1}{n^2}-\frac{j}{n^2}\right)\cdot\left(\frac{1}{n}+\frac{1}{2}\frac{1}{n^2}-\frac{k}{n^2}\right)\right]\\
&\quad+\sum_{j=1}^{n}\left(\frac{j-1}{n}-\frac{j-1}{n}\cdot\frac{j-1}{n}\right)\left[\frac{1}{n^2}+\frac{1}{2}\frac{1}{n^3}-\frac{j}{n^3}-\frac{1}{6}\frac{1}{n^3}\right.\\
&\qquad\left.-\left(\frac{1}{n}+\frac{1}{2}\frac{1}{n^2}-\frac{j}{n^2}\right)\cdot\left(\frac{1}{n}+\frac{1}{2}\frac{1}{n^2}-\frac{j}{n^2}\right)\right].
\end{aligned}
\tag{33}
$$

By a similar argument to that of the calculation of the right-hand side of (28), the right-hand side of (33) is

$$\frac{1}{90}-\frac{1}{48}n^{-2}+\frac{7}{720}n^{-4}.$$

Hence,

$$E[A_nA]=\frac{1}{90}-\frac{1}{48}n^{-2}+\frac{7}{720}n^{-4}.$$

Together with (31) and (32), we have

$$E[(A_n-A)^2]=E[A_n^2]+E[A^2]-2E[A_nA]=\frac{5}{72}n^{-2}-\frac{7}{120}n^{-4}.\quad\square$$

Proposition 4. *We have* $E[(B_n-B)^2]=E[(C_n-C)^2]=\frac{5}{36}n^{-2}-\frac{4}{45}n^{-4}$. *Hence,* $E[(B_n-B)^2]=E[(C_n-C)^2]\le\frac{5}{36}n^{-2}$.

Proof. The second assertion is a direct result of the first, which we proceed to establish. By the definitions of B_n and B,

$$B_n - B = 2\int_0^1 \int_0^t (M_n(s,t) - M(s,t))\, dW_1(s)\, dW_1(t) + \frac{1}{n}\sum_{j=1}^{n} M\left(\frac{j-1}{n}, \frac{j-1}{n}\right) - \int_0^1 M(t,t)\, dt. \tag{34}$$

By a standard property of the double Wiener–Itô integral (see [10, Chapter 9]),

$$E\left[\int_0^1 \int_0^t (M_n(s,t) - M(s,t))\, dW_1(s)\, dW_1(t)\right] = 0.$$

Taking squares and then expectation on both sides of (34), we have (after rearrangement of terms) that

$$E[(B_n - B)^2] = 4E\left[\left(\int_0^1 \int_0^t (M_n(s,t) - M(s,t))\, dW_1(s)\, dW_1(t)\right)^2\right] + \left(\frac{1}{n}\sum_{j=1}^{n} M\left(\frac{j-1}{n}, \frac{j-1}{n}\right) - \int_0^1 M(t,t)\, dt\right)^2. \tag{35}$$

By the Itô isometry, the first term on the right-hand side of (35) is

$$\begin{aligned}
&4\int_0^1 \int_0^t E\left[(M_n(s,t) - M(s,t))^2\right] ds\, dt \\
&= 4\int_0^1 \int_0^t (M_n(s,t) - M(s,t))^2\, ds\, dt = 4\sum_{j=1}^{n} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_0^t (M_n(s,t) - M(s,t))^2\, ds\, dt \\
&= 4\sum_{j=1}^{n} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \left(\sum_{k=1}^{j-1} \int_{\frac{k-1}{n}}^{\frac{k}{n}} (M_n(s,t) - M(s,t))^2\, ds + \int_{\frac{j-1}{n}}^{t} (M_n(s,t) - M(s,t))^2\, ds\right) dt \\
&= 4\sum_{j=1}^{n}\sum_{k=1}^{j-1} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{k-1}{n}}^{\frac{k}{n}} (M_n(s,t) - M(s,t))^2\, ds\, dt \\
&\quad + 4\sum_{j=1}^{n} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{j-1}{n}}^{t} (M_n(s,t) - M(s,t))^2\, ds\, dt \\
&= 4\sum_{j=1}^{n}\sum_{k=1}^{j-1} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{k-1}{n}}^{\frac{k}{n}} \left(\frac{k-1}{n} - \frac{k-1}{n}\cdot\frac{j-1}{n} - s + st\right)^2 ds\, dt \\
&\quad + 4\sum_{j=1}^{n} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{j-1}{n}}^{t} \left(\frac{j-1}{n} - \frac{j-1}{n}\cdot\frac{j-1}{n} - s + st\right)^2 ds\, dt \\
&= 4\sum_{j=1}^{n}\sum_{k=1}^{j-1} \left(\frac{2j^2 + 2k^2 + 3jk}{6n^6} - \frac{(4n+5)j}{6n^6} - \frac{(3n+5)k}{6n^6} + \frac{6n^2 + 15n + 11}{18n^6}\right) \\
&\quad + 4\sum_{j=1}^{n} \left(\frac{7j^2}{12n^6} - \frac{5(n+2)j}{12n^6} + \frac{15n^2 + 51n + 55}{180n^6}\right),
\end{aligned} \tag{36}$$

where in the last equality we have explicitly calculated the two double integrals. Note that the first term on the right-hand side of (36) is 4 times a double summation of a polynomial of j and k, calculating the summation with respect to k by Faulhaber's formula yields four times a single summation of a polynomial of j over $j = 1, \ldots, n$. Applying Faulhaber's formula again to this summation, we have that the first term on the right-hand side of (36) is

$$\frac{5}{36}n^{-2} - \frac{5}{18}n^{-3} + \frac{1}{12}n^{-4} + \frac{1}{18}n^{-5}. \tag{37}$$

Again, by Faulhaber's formula, the second term on the right-hand side of (36) is

$$\frac{5}{18}n^{-3} - \frac{1}{5}n^{-4} - \frac{1}{18}n^{-5}. \tag{38}$$

Combining (37) and (38), the first term on the right-hand side of (35) is

$$\frac{5}{36}n^{-2} - \frac{7}{60}n^{-4}. \tag{39}$$

The second term on the right-hand side of (35) is

$$\begin{aligned} &\left(\frac{1}{n}\sum_{j=1}^{n}\left(\frac{j-1}{n} - \left(\frac{j-1}{n}\right)^2\right) - \int_0^1 (t-t^2)\,dt\right)^2 \\ &= \left(\frac{n^2-1}{6n^2} - \frac{1}{6}\right)^2 = \frac{1}{36}n^{-4}. \end{aligned} \tag{40}$$

Combining (35), (39), (40) gives $E[(B_n - B)^2] = \frac{5}{36}n^{-2} - \frac{4}{45}n^{-4}$. Symmetrically, $E[(C_n - C)^2] = \frac{5}{36}n^{-2} - \frac{4}{45}n^{-4}$ too. This completes the proof. □

5.3. An upper bound for the Wasserstein distance

In this section, we will derive an upper bound for $d_W(A_n/\sqrt{B_n C_n}, A/\sqrt{BC})$. The result relies on three preparatory lemmas. The first, Lemma 5.2, is a special case of Proposition 1 in [7]. When used in conjunction with Lemma 5.1, it shows that we must have a good lower-bound handle on the behavior of d_n, which is the topic of the Lemma 5.3. These then culminate in showing (Lemma 5.4) that B and B_n have inverse moments, with the latter being uniformly bounded in n. This fact may seem surprising, since, as second chaos variables, negative moments can explode, but this does not apply because B, B_n are non-centered, and a.s. positive. The uniformity over n in Lemma 5.4 is a consequence of the convergence of the moment-generating functions of the B_n's to a limit which decays rapidly at $+\infty$ (at the rate $\sqrt{2s}e^{-2s}$), ensuring control of the tails.

We exploit the explicit nature of these expressions to prove Lemma 5.4 and the results that precede it, but our strategy could also work for other processes, for instance by invoking dominated convergence and by controlling d_n via its constituent eigenvalues. This means that our methodology could handle other processes, or other quadratic forms than B_n, if one could still control d_n, via the properties of the matrix K_n, whose positive-definite character is very general. This is an important point in understanding the ingredients in the proof of Lemma 5.3. We obtain lower bounds for d_n by estimating selected terms in its sum representation, ignoring others because none of them are negative, and the positive-definite property of K_n is the reason all terms in the sum are non-negative. This last justification is not entirely trivial, and though it is not used in our proofs because all our formulas are explicit, it is worth mentioning the reason here. We are interested in lower bounds on the moment-generating function of B_n, which equals $d_n(-2s/n) = \prod_{k=2}^{n}(1+2\lambda_k s/n)$. Since all λ_k are positive, this expression is thus a polynomial in s with positive coefficients. That positivity translates into the one used in the proof of Lemma 5.3.

Lemma 5.2. *Let X be a random variable satisfying $X > 0$ a.s. and $\phi_X(s) = E[e^{-sX}]$ be its moment generating function. Then for every $m \in \mathbb{N}_+$,*

$$E\left[X^{-m}\right] = \frac{1}{(m-1)!}\int_0^\infty s^{m-1}\phi_X(s)\,ds.$$

Proof. See Proposition 1 in [7]. □

We now turn to Lemma 5.3.

Lemma 5.3. *For $n \geq 11$, we have*

(a) $d_n(-2s/n) \geq 1$ *for* $s \geq 0$*;*
(b) $d_n(-2s/n) \geq 2^5\binom{n}{11}n^{-11}s^5$ *for* $s \geq 0$*;*
(c) $d_n(-2s/n) \geq \left(e^{\sqrt{s/2}} - e^{-\sqrt{s/2}}\right)/\sqrt{10s}$ *for* $0 \leq s \leq n^2/2$.

Proof. It follows from (9) that

$$\begin{aligned} d_n\left(-\frac{2s}{n}\right) &= \frac{(-1)^{n-1}}{n\cdot 2^{n-1}}\sum_{k=1}^{\lceil n/2\rceil}\binom{n}{2k-1}\left(-\frac{2s}{n^2}-2\right)^{n-(2k-1)}\left(\left(-\frac{2s}{n^2}-2\right)^2-4\right)^{k-1} \\ &= \sum_{k=1}^{\lceil n/2\rceil}\frac{1}{n}\binom{n}{2k-1}\left(\frac{s}{n^2}+1\right)^{n-(2k-1)}\left(\frac{2s}{n^2}+\frac{s^2}{n^4}\right)^{k-1}. \end{aligned} \tag{41}$$

Note the first term of the summation on the right-hand side of (41) is $(s/n^2+1)^{n-1}$. Then,

$$d_n\left(-\frac{2s}{n}\right) \geq \left(\frac{s}{n^2}+1\right)^{n-1} \geq 1.$$

This proves statement (a). We now note that $n \geq 11$, $\lceil n/2\rceil \geq 6$. Let us consider the sixth term of the summation on the right-hand side of (41), i.e.

$$\frac{1}{n}\binom{n}{11}\left(\frac{s}{n^2}+1\right)^{n-11}\left(\frac{2s}{n^2}+\frac{s^2}{n^4}\right)^5 \geq \frac{1}{n}\binom{n}{11}\left(\frac{2s}{n^2}\right)^5 = 2^5\binom{n}{11}n^{-11}s^5.$$

Then $d_n(-2s/n) \geq 2^5\binom{n}{11}n^{-11}s^5$. This completes the proof of statement (b). Finally, it follows from (9) that

$$d_n\left(-\frac{2s}{n}\right) = \frac{1}{2\sqrt{2s+s^2/n^2}}\left[\left(\frac{s}{n^2}+1+\sqrt{\frac{2s}{n^2}+\frac{s^2}{n^4}}\right)^n - \left(\frac{s}{n^2}+1-\sqrt{\frac{2s}{n^2}+\frac{s^2}{n^4}}\right)^n\right]. \tag{42}$$

Recall the useful fact that $\log(1+x) \geq x/2$ for $0 \leq x \leq 1$. Then for $0 \leq s \leq n^2/2$,

$$\left(\frac{s}{n^2}+1+\sqrt{\frac{2s}{n^2}+\frac{s^2}{n^4}}\right)^n \geq \left(1+\frac{\sqrt{2s}}{n}\right)^n = e^{n\log(1+\sqrt{2s}/n)} \geq e^{\sqrt{s/2}}, \tag{43}$$

Further,

$$\left(\frac{s}{n^2}+1-\sqrt{\frac{2s}{n^2}+\frac{s^2}{n^4}}\right)^n=\left(\frac{s}{n^2}+1+\sqrt{\frac{2s}{n^2}+\frac{s^2}{n^4}}\right)^{-n}\leq e^{-\sqrt{s/2}}. \tag{44}$$

Note that

$$2s+\frac{s^2}{n^2}\leq 2s+\frac{n^2}{2}\cdot\frac{s}{n^2}=\frac{5s}{2},$$

Together with (42), (43) and (44), statement (c) follows. □

We now turn to Lemma 5.4.

Lemma 5.4. *We have*

(a) $E[B^{-m}]=E[C^{-m}]<\infty$ *for every* $m\in\mathbb{N}_+$*;*
(b) $\sup_{n\geq 11}E[B_n^{-1}]=\sup_{n\geq 11}E[C_n^{-1}]<\infty$.

Proof. We first consider statement (a). Since B and C are identically distributed, it suffices to prove that $E[B^{-m}]<\infty$ for every $m\in\mathbb{N}_+$. Applying Lemma 5.2 gives

$$E[B^{-m}]=\frac{1}{(m-1)!}\int_0^\infty s^{m-1}\phi_B(s)\,ds.$$

By Lemma 5.1, we have $\phi_B(0)=1$ and $\phi_B(s)\sim 2^{3/2}s^{1/4}e^{-\sqrt{s/2}}$ as $s\to\infty$. Then the boundedness of $E[B^{-m}]$ follows immediately. For statement (b), similarly, we need only prove that $\sup_{n\geq 11}E[B_n^{-1}]<\infty$. By Lemmas 5.1, 5.2 and 5.3, we have

$$\begin{aligned}
E[B_n^{-1}]&=\int_0^\infty\phi_{B_n}(s)\,ds=\int_0^\infty d_n\left(-\frac{2s}{n}\right)^{-1/2}ds\\
&=\int_0^1 d_n\left(-\frac{2s}{n}\right)^{-1/2}ds+\int_1^{n^2/2}d_n\left(-\frac{2s}{n}\right)^{-1/2}ds+\int_{n^2/2}^\infty d_n\left(-\frac{2s}{n}\right)^{-1/2}ds\\
&\leq\int_0^1 1\,ds+\int_1^{n^2/2}\left(\left(e^{\sqrt{s/2}}-e^{-\sqrt{s/2}}\right)/\sqrt{10s}\right)^{-1/2}ds+\int_{n^2/2}^\infty 2^{-\frac{5}{2}}\binom{n}{11}^{-\frac{1}{2}}n^{\frac{11}{2}}s^{-\frac{5}{2}}\,ds\\
&=1+\int_1^{n^2/2}\left(\left(e^{\sqrt{s/2}}-e^{-\sqrt{s/2}}\right)/\sqrt{10s}\right)^{-1/2}ds+\frac{1}{3}\binom{n}{11}^{-\frac{1}{2}}n^{\frac{5}{2}},
\end{aligned}$$

where in the first inequality we have applied Lemma 5.3. Taking the supremum over $n\geq 11$ on both sides of the last display yields

$$\sup_{n\geq 11}E[B_n^{-1}]\leq 1+\int_1^\infty\left(\left(e^{\sqrt{s/2}}-e^{-\sqrt{s/2}}\right)/\sqrt{10s}\right)^{-1/2}ds+\sup_{n\geq 11}\frac{1}{3}\binom{n}{11}^{-\frac{1}{2}}n^{\frac{5}{2}}. \tag{45}$$

The boundedness of the second term on the right-hand side of (45) follows by the fact that, as $n\to\infty$,

$$\left(\left(e^{\sqrt{s/2}}-e^{-\sqrt{s/2}}\right)/\sqrt{10s}\right)^{-1/2}\sim(10s)^{1/4}e^{-\sqrt{s/8}}.$$

The boundedness of the third term on the right-hand side of (45) follows by the fact $\binom{n}{11}^{-\frac{1}{2}}n^{\frac{5}{2}}\to 0$ as $n\to\infty$. This completes the proof of statement (b). □

Let us define $L_m := E[B^{-m}] = E[C^{-m}]$ for $m = 1, 2, 3$ and $L_4 := \sup_{n\geq 11} E[B_n^{-1}] = \sup_{n\geq 11} E[C_n^{-1}]$. With the above preparations in hand, we are ready to reveal the section's main result.

Theorem 5.5. *For $n \geq 11$, with A, B, C, A_n, B_n, C_n defined in Section 5.2, we have*

$$d_W(\theta_n, \theta) \leq E\left|\frac{A_n}{\sqrt{B_n C_n}} - \frac{A}{\sqrt{BC}}\right| \leq \frac{L_5}{n},$$

where, via the constants L_1, L_3, L_4 defined above, the constant L_5 is defined as

$$L_5 := \frac{1}{12}\left\{\frac{1}{132}\left[\frac{5}{2}(L_3 + L_4)\right]^{\frac{1}{2}} + 2\right\}\left[\frac{5}{2}(L_3 + L_4)\right]^{\frac{1}{2}} + \frac{1}{6}\left(\frac{5}{2}\right)^{\frac{1}{2}} L_1.$$

Proof. By Proposition 2, we have

$$d_W(\theta_n, \theta) = d_W\left(\frac{A_n}{\sqrt{B_n C_n}}, \frac{A}{\sqrt{BC}}\right) = \sup_{f\in \mathrm{Lip}(1)}\left|Ef\left(\frac{A_n}{\sqrt{B_n C_n}}\right) - Ef\left(\frac{A}{\sqrt{BC}}\right)\right|.$$

For every $f \in \mathrm{Lip}(1)$, and every pair of integrable random variables (X, Y) on the same probability space,

$$|Ef(X) - Ef(Y)| \leq E|f(X) - f(Y)| \leq E|X - Y|.$$

Taking the supremum over $f \in \mathrm{Lip}(1)$ on both sides of the above displays with X, Y replaced by $A_n/\sqrt{B_n C_n}$, $A/\sqrt{BC}$ yields

$$d_W(\theta_n, \theta) \leq E\left|\frac{A_n}{\sqrt{B_n C_n}} - \frac{A}{\sqrt{BC}}\right|.$$

Thus, we need only bound the expectation of $|A_n/\sqrt{B_n C_n} - A/\sqrt{BC}|$. Note that

$$\begin{aligned}
&\frac{A_n}{\sqrt{B_n C_n}} - \frac{A}{\sqrt{BC}} = \frac{A_n\sqrt{BC} - A\sqrt{B_n C_n}}{\sqrt{B_n C_n}\sqrt{BC}}\\
&= \frac{A_n(\sqrt{BC} - \sqrt{B_n C_n}) + (A_n - A)\sqrt{B_n C_n}}{\sqrt{B_n C_n}\sqrt{BC}}\\
&= \frac{A_n}{\sqrt{B_n C_n}} \cdot \frac{\sqrt{BC} - \sqrt{B_n C_n}}{\sqrt{BC}} + \frac{A_n - A}{\sqrt{BC}}.
\end{aligned}$$

Then,

$$\begin{aligned}
&\left|\frac{A_n}{\sqrt{B_n C_n}} - \frac{A}{\sqrt{BC}}\right| \leq \left|\frac{A_n}{\sqrt{B_n C_n}}\right| \cdot \frac{|\sqrt{BC} - \sqrt{B_n C_n}|}{\sqrt{BC}} + \frac{|A_n - A|}{\sqrt{BC}}\\
&\leq \frac{|\sqrt{BC} - \sqrt{B_n C_n}|}{\sqrt{BC}} + \frac{|A_n - A|}{\sqrt{BC}}\\
&= \frac{|(\sqrt{B} - \sqrt{B_n})\sqrt{C} + \sqrt{B_n}(\sqrt{C} - \sqrt{C_n})|}{\sqrt{BC}} + \frac{|A_n - A|}{\sqrt{BC}}\\
&\leq \frac{|\sqrt{B} - \sqrt{B_n}|}{\sqrt{B}} + \frac{\sqrt{B_n}}{\sqrt{B}}\frac{|\sqrt{C} - \sqrt{C_n}|}{\sqrt{C}} + \frac{|A_n - A|}{\sqrt{BC}}, \qquad (46)
\end{aligned}$$

where in the second inequality we have invoked Corollary 1. By the inequality of arithmetic and geometric means,

$$\frac{|\sqrt{B}-\sqrt{B_n}|}{\sqrt{B}}=\frac{|B_n-B|}{\sqrt{B}(\sqrt{B}+\sqrt{B_n})}\leq\frac{|B_n-B|}{2\sqrt{B}\sqrt{\sqrt{B}\sqrt{B_n}}}=\frac{|B_n-B|}{2B^{\frac{3}{4}}\,B_n^{\frac{1}{4}}}.$$

Then by Hölder's inequality,

$$\begin{aligned}
&E\left[\frac{|\sqrt{B}-\sqrt{B_n}|}{\sqrt{B}}\right]\leq E\left[\frac{|B_n-B|}{2B^{\frac{3}{4}}\,B_n^{\frac{1}{4}}}\right]\leq\left\{E\left[(B_n-B)^2\right]\right\}^{\frac{1}{2}}\left\{E\left[\frac{1}{4}B^{-\frac{3}{2}}\,B_n^{-\frac{1}{2}}\right]\right\}^{\frac{1}{2}}\\
&\leq\left\{E\left[(B_n-B)^2\right]\right\}^{\frac{1}{2}}\left\{E\left[\frac{1}{8}\left(B^{-3}+B_n^{-1}\right)\right]\right\}^{\frac{1}{2}}\leq\left(\frac{5}{36n^2}\right)^{\frac{1}{2}}\left[\frac{1}{8}(L_3+L_4)\right]^{\frac{1}{2}}\\
&=\frac{1}{12}\left[\frac{5}{2}(L_3+L_4)\right]^{\frac{1}{2}}\frac{1}{n},
\end{aligned}\tag{47}$$

where the third inequality follows from the inequality of arithmetic and geometric means, and the fourth inequality is due to Proposition 4. Similarly, we have

$$E\left[\frac{|\sqrt{C}-\sqrt{C_n}|}{\sqrt{C}}\right]\leq\frac{1}{12}\left[\frac{5}{2}(L_3+L_4)\right]^{\frac{1}{2}}\frac{1}{n},$$

and

$$\begin{aligned}
&E\left[\frac{\sqrt{B_n}}{\sqrt{B}}\right]=E\left[\frac{\sqrt{B_n}-\sqrt{B}}{\sqrt{B}}+1\right]\leq E\left[\frac{|\sqrt{B_n}-\sqrt{B}|}{\sqrt{B}}+1\right]\\
&\leq\frac{1}{12}\left[\frac{5}{2}(L_3+L_4)\right]^{\frac{1}{2}}\frac{1}{n}+1\leq\frac{1}{132}\left[\frac{5}{2}(L_3+L_4)\right]^{\frac{1}{2}}+1.
\end{aligned}$$

Since (B_n, B) and (C_n, C) are independent, we have

$$\begin{aligned}
&E\left[\frac{\sqrt{B_n}}{\sqrt{B}}\frac{|\sqrt{C}-\sqrt{C_n}|}{\sqrt{C}}\right]=E\left[\frac{\sqrt{B_n}}{\sqrt{B}}\right]E\left[\frac{|\sqrt{C}-\sqrt{C_n}|}{\sqrt{C}}\right]\\
&\leq\frac{1}{12}\left\{\frac{1}{132}\left[\frac{5}{2}(L_3+L_4)\right]^{\frac{1}{2}}+1\right\}\left[\frac{5}{2}(L_3+L_4)\right]^{\frac{1}{2}}\frac{1}{n}.
\end{aligned}\tag{48}$$

By Hölder's inequality and Proposition 3,

$$\begin{aligned}
&E\left[\frac{|A_n-A|}{\sqrt{BC}}\right]\leq\left\{E\left[(A_n-A)^2\right]\right\}^{\frac{1}{2}}\left\{E\left[B^{-1}C^{-1}\right]\right\}^{\frac{1}{2}}\\
&=\left\{E\left[(A_n-A)^2\right]\right\}^{\frac{1}{2}}\left\{E[B^{-1}]E[C^{-1}]\right\}^{\frac{1}{2}}\leq\left(\frac{5}{72n^2}\right)^{\frac{1}{2}}\cdot L_1=\frac{1}{6}\left(\frac{5}{2}\right)^{\frac{1}{2}}L_1\frac{1}{n}.
\end{aligned}\tag{49}$$

Combining (46), (47), (48) and (49) yields

$$E\left[\left|\frac{A_n}{\sqrt{B_nC_n}}-\frac{A}{\sqrt{BC}}\right|\right]\leq\frac{L_5}{n}.$$

This completes the proof. □

6. Future work

In this section, we present conjectures which, while beyond the scope of the present paper, should constitute opportunities for future work which should be tractable given some known tools and techniques in the analysis on Wiener chaos, and could have interesting applications to statistical testing based on paths of time series.

The reader can refer to Section 5.1 for conjectures on convergence rates, and their implications, regarding the distinction between random walks and other types of time series. Those conjectures would apply to statistics which can be related to the Wasserstein distance.

Going beyond them, we conjecture that, for practical purposes, the convergence of θ_n to θ also occurs in total variation at the same rate as in Wasserstein distance, in the sense that the probability law of θ_n converges at the rate $r(n) := cn^{-1}$ for some constant c though this may be harder to establish except empirically or via simulations. The practical conjecture, that extends from the Wasserstein to the total variation distance, would be significant for several reasons, including because the total variation distance is an upper bound on the Kolmogorov distance (see [15, Section 8.1] for further details on Kolmogorov distance), but only the square root of the Wasserstein distance bounds the Kolmogorov distance. As the latter is the supremum norm for the distance between cumulative distribution functions (CDFs), an application of the practical conjecture, using specifically the implication for the Kolmogorov distance, would be as follows. An upper bound of order of magnitude 10^{-2}, say, could legitimately imply that any estimate on the αth percentile of θ could result in the same estimate on the $(\alpha - 10^{-2})$th percentile of θ_n. One could thus build a test of independence of two (Gaussian) random walks of length n where the rejection region at the confidence level α could be equated to the rejection region using the CDF of θ at the confidence level $\alpha + 10^{-2}$ as long as $r(n) < 10^{-2}$. This argument could take into account the multiplicative constant c in the speed of convergence $r(n)$, which could also be determined from simulations. Without our conjecture on total variation rate of convergence, using instead our Theorem 5.5, this strategy for rejection regions at level $\alpha+10^{-2}$ would follow from $r(n)^{1/2} < 10^{-2}$, because, as we mentioned, the Wasserstein distance only bounds the square root of the Kolmogorov distance.

Other options for conjectures for statistical testing could include studying the speed of convergence of moment ratios of paths, such as a kurtosis-type statistic, and their fluctuations. Though this is also beyond the scope of this paper, we conjecture that, unlike the limit of the law of θ_n itself, whose numerator and denominator converge in the second chaos, the polarization of an empirical kurtosis for two Gaussian random walks has normal fluctuations. Such a study could use either the technique presented in Section 5 via bounding the negative moments of the denominator from its moment-generating function, or the so-called optimal fourth moment theorem [14], where the speed of convergence of normal fluctuation for chaos sequences is known sharply in total variation.

We also suspect that the convergence phenomena we uncover here in the previous section are not restricted to Gaussian random walks, but hold for a wide range of random walks and other processes, including walks with other step distributions. Because of the heavy reliance on the Gaussian property in our work, particularly to be able to work in the second Wiener chaos, using non-Gaussian step distributions would require different tools. However, going from Gaussian random walks and Wiener processes to other Gaussian time series and their continuous limits could preserve a number of the tools we present here. For instance, we rely on the extraordinary convenience of Lemma 5.2 and the explicit nature of the corresponding moment-generating function, to estimate negative moments, but this can be done by other means for other Gaussian processes and their discrete-time observations. For example, relying on Karhunen–Loève expansions, we can derive lower bounds for discrete-time observations in terms of a product of i.i.d. Gaussian random variables, which in turn give upper bounds on the negative moments. Similarly, as mentioned in Section 5.1, we use the convenience of being able to calculate the exact value of the $L^2(\Omega)$ distance between the constituent elements of θ and θ_n (e.g. by employing Faulhaber's formula for the partial sum of the powers of integers). But these expressions can be estimated nearly as precisely, using the kernel representations, by invoking comparisons between series and Riemann integrals, with error terms of the same order as the second-order terms in Propositions 3 and 4.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We wish to thank an anonymous referee, whose report greatly helped improve the quality of this manuscript. The first-named author acknowledges, with gratitude, the support of the U.S. Office of Naval Research (grant N00014-21-1-2672). The first-named and third-named authors acknowledge, with gratitude, the support of NSF DMS-1811936.

Appendix

This appendix proves Lemma 3.1 and Proposition 2.

A.1. Proof of Lemma 3.1

For simplicity, we will start with $d_n(n^2\lambda)$, for $n \geq 5$. From the definition of $d_n(\lambda)$, we have

$$d_n(n^2\lambda) = \det\left(I_{n-1} - n^2\lambda K_n\right) = \det\left(I_{n-1} - \lambda\{n\min(j,k) - jk\}_{j,k=1}^{n-1}\right)$$

$$= \begin{vmatrix} 1-(n-1)\lambda, & -(n-2)\lambda, & -(n-3)\lambda, & -(n-4)\lambda, & \cdots, & -\lambda \\ -(n-2)\lambda, & 1-2(n-2)\lambda, & -2(n-3)\lambda, & -2(n-4)\lambda, & \cdots, & -2\lambda \\ -(n-3)\lambda, & -2(n-3)\lambda, & 1-3(n-3)\lambda, & -3(n-4)\lambda, & \cdots, & -3\lambda \\ -(n-4)\lambda, & -2(n-4)\lambda, & -3(n-4)\lambda, & 1-4(n-4)\lambda, & \cdots, & -4\lambda \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\lambda, & -2\lambda, & -3\lambda, & -4\lambda, & \cdots, & 1-(n-1)\lambda \end{vmatrix} \tag{50}$$

$$= \begin{vmatrix} 1-(n-1)\lambda, & -(n-2)\lambda, & -(n-3)\lambda, & -(n-4)\lambda, & \cdots, & -\lambda \\ n\lambda-2, & 1, & 0, & 0, & \cdots, & 0 \\ 2n\lambda-3, & n\lambda, & 1, & 0, & \cdots, & 0 \\ 3n\lambda-4, & 2n\lambda, & n\lambda, & 1, & \cdots, & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (n-2)n\lambda-(n-1), & (n-3)n\lambda, & (n-4)n\lambda, & (n-5)n\lambda, & \cdots, & 1 \end{vmatrix} \tag{51}$$

$$= \begin{vmatrix} 1-(n-1)\lambda, & 0, & 0, & \cdots, & -\lambda \\ n\lambda-2, & 1, & 0, & \cdots, & 0 \\ 2n\lambda-3, & n\lambda, & 1, & \cdots, & 0 \\ 3n\lambda-4, & 2n\lambda, & n\lambda, & \cdots, & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (n-2)n\lambda-(n-1), & (n-3)n\lambda-(n-2), & (n-4)n\lambda-(n-3), & \cdots, & 1 \end{vmatrix}, \tag{52}$$

where from (50) to (51), we add $(-j)\times$ the first row to the jth row, for $j = 2, 3, \dots, n-1$. From (51) to (52), we add $-(n-j)\times$ the last column to the jth column, for $j = 2, 3 \cdots, n-2$. We expand the determinant in (52) by its first row and obtain

$$d_n(n^2\lambda) = 1-(n-1)\lambda + (-1)^{n+1}\lambda \cdot \begin{vmatrix} n\lambda-2, & 1, & 0, & \cdots, & 0 \\ 2n\lambda-3, & n\lambda, & 1, & \cdots, & 0 \\ 3n\lambda-4, & 2n\lambda, & n\lambda, & \cdots, & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (n-3)n\lambda-(n-2), & (n-4)n\lambda, & (n-5)n\lambda, & \cdots, & 1 \\ (n-2)n\lambda-(n-1), & (n-3)n\lambda-(n-2), & (n-4)n\lambda-(n-3), & \cdots, & n\lambda-2 \end{vmatrix}. \tag{53}$$

Further, the determinant in (53) is equal to

$$= \begin{vmatrix} n\lambda-2, & 1, & 0, & \cdots, & 0, & 0, & 0 \\ n\lambda-1, & n\lambda-1, & 1, & \cdots, & 0, & 0, & 0 \\ n\lambda-1, & n\lambda, & n\lambda-1, & \cdots, & 0, & 0, & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ n\lambda-1, & n\lambda, & n\lambda, & \cdots, & n\lambda-1, & 1, & 0 \\ n\lambda-1, & n\lambda, & n\lambda, & \cdots, & n\lambda, & n\lambda-1, & 1 \\ n\lambda-1, & n\lambda-(n-2), & n\lambda-(n-3), & \cdots, & n\lambda-4, & n\lambda-3, & n\lambda-3 \end{vmatrix} \tag{54}$$

$$= \begin{vmatrix} n\lambda-2, & 1, & 0, & \cdots, & 0, & 0, & 0 \\ 1, & n\lambda-2, & 1, & \cdots, & 0, & 0, & 0 \\ 0, & 1, & n\lambda-2, & \cdots, & 0, & 0, & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0, & 0, & 0, & \cdots, & n\lambda-2, & 1, & 0 \\ 0, & 0, & 0, & \cdots, & 1, & n\lambda-2, & 1 \\ 0, & -(n-2), & -(n-3), & \cdots, & -4, & -2, & n\lambda-4 \end{vmatrix}. \tag{55}$$

From (53) to (54), we add $(-1)\times(j-1)$th row to jth row for $j = n-2, n-3, \dots, 2$. From (54) to (55), similarly, we add $(-1)\times(j-1)$th row to jth row for $j = n-2, n-3, \dots, 2$.

Before proceeding to calculate $d_n(n^2\lambda)$, we pause here to introduce a new determinant, closely related to $d_n(n^2\lambda)$. Let us denote by $p_n(\lambda)$ the following $(n-2)\times(n-2)$ determinant:

$$\begin{vmatrix} n\lambda-2, & 1, & 0, & \cdots, & 0, & 0, & 0 \\ 1, & n\lambda-2, & 1, & \cdots, & 0, & 0, & 0 \\ 0, & 1, & n\lambda-2, & \cdots, & 0, & 0, & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0, & 0, & 0, & \cdots, & n\lambda-2, & 1, & 0 \\ 0, & 0, & 0, & \cdots, & 1, & n\lambda-2, & 1 \\ -(n-1), & -(n-2), & -(n-3), & \cdots, & -4, & -2, & n\lambda-4 \end{vmatrix}. \tag{56}$$

As mentioned, we introduce this determinant to compensate for the break in symmetry in (55) because of the zero in the lower-left-hand corner there. We may easily verify that $p_n(\lambda) = (55)+(-1)^n(n-1)$. Note that, from the expression of $d_n(n^2\lambda)$ in (53) and (55), we have

$$\begin{aligned} d_n(n^2\lambda) &= 1-(n-1)\lambda + (-1)^{n+1}\lambda \times (55) \\ &= 1-(n-1)\lambda + (-1)^{n+1}\lambda \left(p_n(\lambda) - (-1)^n(n-1)\right) \\ &= 1 + (-1)^{n+1}\lambda\, p_n(\lambda). \end{aligned} \tag{57}$$

By (57), the problem of calculating $d_n(n^2\lambda)$ is converted to the problem of calculating $p_n(\lambda)$. To calculate $p_n(\lambda)$, our strategy is to derive a second-order recursion formula for $p_n(\lambda/n)$, see (59). In what follows, we derive an explicit expression for $p_n(\lambda)$.

For $n \geq 7$, we expand the determinant (56) by its first column and obtain

$$p_n(\lambda) = (n\lambda - 2)\, p_{n-1}\left(\frac{n}{n-1}\lambda\right) + (-1)^n(n-1)$$
$$-1 \cdot \begin{vmatrix} 1, & 0, & 0, & \cdots, & 0, & 0, & 0 \\ 1, & n\lambda - 2, & 1, & \cdots, & 0, & 0, & 0 \\ 0, & 1, & n\lambda - 2, & \cdots, & 0, & 0, & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0, & 0, & 0, & \cdots, & n\lambda - 2, & 1, & 0 \\ 0, & 0, & 0, & \cdots, & 1, & n\lambda - 2, & 1 \\ -(n-2), & -(n-3), & -(n-4), & \cdots, & -4, & -2, & n\lambda - 4 \end{vmatrix}. \tag{58}$$

If we expand the determinant in (58) by its first row, then it is exactly $p_{n-2}\left(\frac{n}{n-2}\lambda\right)$. Hence,

$$p_n(\lambda) = (n\lambda - 2)\, p_{n-1}\left(\frac{n}{n-1}\lambda\right) - p_{n-2}\left(\frac{n}{n-2}\lambda\right) + (-1)^n(n-1).$$

In the above equation, we make a change of variables from λ to λ/n and obtain

$$p_n\left(\frac{\lambda}{n}\right) = (\lambda - 2)p_{n-1}\left(\frac{\lambda}{n-1}\right) - p_{n-2}\left(\frac{\lambda}{n-2}\right) + (-1)^n(n-1).$$

For $\lambda \neq 0$, rearranging the above equation yields

$$(-1)^n\, p_n\left(\frac{\lambda}{n}\right) - \frac{1}{\lambda}n = -(\lambda - 2)\left[(-1)^{n-1} p_{n-1}\left(\frac{\lambda}{n-1}\right) - \frac{1}{\lambda}(n-1)\right]$$
$$- \left[(-1)^{n-2} p_{n-2}\left(\frac{\lambda}{n-2}\right) - \frac{1}{\lambda}(n-2)\right]. \tag{59}$$

The above iterative formula of $(-1)^n p_n(\lambda/n) - n/\lambda$ tells us that for $\lambda \neq 0$ or 4, it must have the following form:

$$C_1 \cdot \left(-\frac{(\lambda - 2) + \sqrt{(\lambda - 2)^2 - 4}}{2}\right)^n + C_2 \cdot \left(-\frac{(\lambda - 2) - \sqrt{(\lambda - 2)^2 - 4}}{2}\right)^n,$$

where C_1 and C_2 are two constants. Direct calculation gives

$$(-1)^5\, p_5\left(\frac{\lambda}{5}\right) - \frac{5}{\lambda} = -(\lambda - 4)\left((\lambda - 2)^2 + 1\right) - \frac{5}{\lambda},$$
$$(-1)^6\, p_6\left(\frac{\lambda}{6}\right) - \frac{6}{\lambda} = (\lambda - 4)(\lambda - 2)^3 + 3 - \frac{6}{\lambda}.$$

Then the constants C_1 and C_2 can be determined as

$$C_1 = -C_2 = \frac{1}{\lambda\sqrt{(\lambda - 2)^2 - 4}}.$$

Hence, for $n \geq 5$ and $\lambda \neq 0$ or 4,

$$
\begin{aligned}
&(-1)^n\, p_n\left(\frac{\lambda}{n}\right) - \frac{1}{\lambda}\, n \\
&= \frac{1}{\lambda\sqrt{(\lambda-2)^2-4}}\left[\left(-\frac{(\lambda-2)+\sqrt{(\lambda-2)^2-4}}{2}\right)^n \right.\\
&\qquad \left. - \left(-\frac{(\lambda-2)-\sqrt{(\lambda-2)^2-4}}{2}\right)^n\right] \\
&= \frac{(-1)^n}{2^{n-1}\,\lambda}\sum_{k=1}^{\lceil n/2\rceil}\binom{n}{2k-1}(\lambda-2)^{n-(2k-1)}\left((\lambda-2)^2-4\right)^{k-1}.
\end{aligned} \tag{60}
$$

Combining (57) and (60) yields that, for $n \geq 5$ and $\lambda \neq 0$ or $4n$, (9) holds. Since both sides of (9) are continuous function of λ, it also holds for every $\lambda \in \mathbb{R}$ and $n \geq 5$. It is straightforward to verify that (9) also holds for $n = 2, 3$ and 4, hence, (9) holds for all $n \geq 2$.

A.2. Proof of Proposition 2

We first note that statement (c) follows immediately from statements (a) and (b). We first prove statement (b). By integration by parts for the Wiener integral, we have

$$
\int_0^t M(s,t)\,dW_1(s) = \int_0^t (s-st)\,dW_1(s) = (t-t^2)W_1(t) - (1-t)\int_0^t W_1(s)\,ds.
$$

Then

$$
\begin{aligned}
&\int_0^1\int_0^t M(s,t)\,dW_1(s)\,dW_2(t) \\
&= \int_0^1 (t-t^2)W_1(t)\,dW_2(t) - \int_0^1 (1-t)\left(\int_0^t W_1(s)\,ds\right)dW_2(t).
\end{aligned} \tag{61}
$$

Applying Itô's lemma to $(1-t)W_2(t)\int_0^t W_1(s)\,ds$ yields

$$
\begin{aligned}
0 = &-\int_0^1\left(\int_0^t W_1(s)\,ds\right)W_2(t)\,dt + \int_0^1 (1-t)\left(\int_0^t W_1(s)\,ds\right)dW_2(t) \\
&+ \int_0^1 (1-t)W_1(t)W_2(t)\,dt.
\end{aligned}
$$

Together with (61), we have

$$
\begin{aligned}
&\int_0^1\int_0^t M(s,t)\,dW_1(s)\,dW_2(t) \\
&= \int_0^1 (t-t^2)W_1(t)\,dW_2(t) + \int_0^1 (1-t)W_1(t)W_2(t)\,dt - \int_0^1\left(\int_0^t W_1(s)\,ds\right)W_2(t)\,dt.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\int_0^1\int_0^s M(s,t)\,dW_2(t)\,dW_1(s) \\
&= \int_0^1 (t-t^2)W_2(t)\,dW_1(t) + \int_0^1 (1-t)W_1(t)W_2(t)\,dt - \int_0^1\left(\int_0^t W_2(s)\,ds\right)W_1(t)\,dt.
\end{aligned}
$$

Then

$$
\begin{aligned}
A &= \int_0^1 \int_0^t M(s,t)\, dW_1(s)\, dW_2(t) + \int_0^1 \int_0^s M(s,t)\, dW_2(t)\, dW_1(s) \\
&= \int_0^1 (t-t^2) W_1(t)\, dW_2(t) + \int_0^1 (t-t^2) W_2(t)\, dW_1(t) + \int_0^1 (2-2t) W_1(t) W_2(t)\, dt \\
&\quad - \int_0^1 \left(\int_0^t W_1(s)\, ds \right) W_2(t)\, dt - \int_0^1 \left(\int_0^t W_2(s)\, ds \right) W_1(t)\, dt.
\end{aligned} \tag{62}
$$

Applying Itô's lemma to $(t-t^2)\, W_1(t)\, W_2(t)$ yields

$$
0 = \int_0^1 (1-2t) W_1(t) W_2(t)\, dt + \int_0^1 (t-t^2) W_2(t)\, dW_1(t) + \int_0^1 (t-t^2) W_1(t)\, dW_2(t). \tag{63}
$$

Note that

$$
\begin{aligned}
\int_0^1 \left(\int_0^t W_1(s)\, ds \right) W_2(t)\, dt &= \int_0^1 \left(\int_s^1 W_2(t)\, dt \right) W_1(s)\, ds \\
&= \int_0^1 \left(\int_t^1 W_2(s)\, ds \right) W_1(t)\, dt,
\end{aligned}
$$

where the first equality follows by interchanging the order of integrals and the second equality follows by substituting (s,t) for (t,s). We then calculate

$$
\begin{aligned}
&\int_0^1 \left(\int_0^t W_1(s)\, ds \right) W_2(t)\, dt + \int_0^1 \left(\int_0^t W_2(s)\, ds \right) W_1(t)\, dt \\
&= \int_0^1 \left(\int_t^1 W_2(s)\, ds \right) W_1(t)\, dt + \int_0^1 \left(\int_0^t W_2(s)\, ds \right) W_1(t)\, dt \\
&= \int_0^1 \left(\int_0^1 W_2(s)\, ds \right) W_1(t)\, dt = \int_0^1 W_1(t)\, dt \int_0^1 W_2(s)\, ds.
\end{aligned} \tag{64}
$$

Combining (62), (63) and (64), (25) follows. Since (26) and (27) are symmetric, we need only prove (26), and then (27) will follow similarly. By a similar argument to that in the derivation of (62), we have

$$
\begin{aligned}
B &= 2 \int_0^1 (t-t^2) W_1(t)\, dW_1(t) + \int_0^1 (2-2t)\, W_1^2(t)\, dt \\
&\quad - 2 \int_0^1 \left(\int_0^t W_1(s)\, ds \right) W_1(t)\, dt + \int_0^1 M(t,t)\, dt.
\end{aligned} \tag{65}
$$

Applying Itô's lemma to $(t-t^2)\, W_1^2(t)$ yields

$$
0 = \int_0^1 (1-2t)\, W_1^2(t)\, dt + 2 \int_0^1 (t-t^2) W_1(t)\, dW_1(t) + \int_0^1 (t-t^2)\, dt. \tag{66}
$$

Further,

$$
\begin{aligned}
\int_0^1 \left(\int_0^t W_1(s)\, ds \right) W_1(t)\, dt &= \int_0^1 \left(\int_0^t W_1(s)\, ds \right) d \left(\int_0^t W_1(s)\, ds \right) \\
&= \frac{1}{2} \left(\int_0^1 W_1(s)\, ds \right)^2 .
\end{aligned} \tag{67}
$$

Combining (65), (66) and (67), (26) follows.

In the remainder of the proof, we will show that

$$A_n = \sum_{j,k=1}^{n} M\left(\frac{j-1}{n}, \frac{k-1}{n}\right)\left(W_1\left(\frac{j}{n}\right) - W_1\left(\frac{j-1}{n}\right)\right) \times \left(W_2\left(\frac{k}{n}\right) - W_2\left(\frac{k-1}{n}\right)\right), \tag{68}$$

$$B_n = \sum_{j,k=1}^{n} M\left(\frac{j-1}{n}, \frac{k-1}{n}\right)\left(W_1\left(\frac{j}{n}\right) - W_1\left(\frac{j-1}{n}\right)\right) \times \left(W_1\left(\frac{k}{n}\right) - W_1\left(\frac{k-1}{n}\right)\right), \tag{69}$$

$$C_n = \sum_{j,k=1}^{n} M\left(\frac{j-1}{n}, \frac{k-1}{n}\right)\left(W_2\left(\frac{j}{n}\right) - W_2\left(\frac{j-1}{n}\right)\right) \times \left(W_2\left(\frac{k}{n}\right) - W_2\left(\frac{k-1}{n}\right)\right). \tag{70}$$

Note that $W_1\left(\frac{j}{n}\right) - W_1\left(\frac{j-1}{n}\right)$, $W_2\left(\frac{k}{n}\right) - W_2\left(\frac{k-1}{n}\right)$, $j, k = 1, 2 \ldots, n$ are mutually independent Gaussian random variables with distribution $\mathcal{N}(0, 1/n)$. By Section 3 (see line (8)), it follows easily that Z_{12}^n/n, Z_{11}^n/n and Z_{22}^n/n are quadratic forms of the random variables $(X_1/\sqrt{n}, X_2/\sqrt{n}, \ldots, X_n/\sqrt{n})$ and $(Y_1/\sqrt{n}, Y_2/\sqrt{n}, \ldots, Y_n/\sqrt{n})$ with same coefficients as A_n, B_n and C_n respectively. Thus, statement (a) of Proposition 2 follows immediately.

For simplicity, let u_j denote j/n for $j = 0, 1, 2, \ldots, n$. We proceed to calculate

$$\begin{aligned}
&\int_0^1 \int_0^t M_n(s,t)\, dW_1(s)\, dW_2(t) \\
&= \int_0^1 \int_0^t \sum_{j,k=1}^{n} M\left(u_{j-1}, u_{k-1}\right) \mathbb{1}_{\{u_{j-1} < s \le u_j\}} \mathbb{1}_{\{u_{k-1} < t \le u_k\}}\, dW_1(s)\, dW_2(t) \\
&= \sum_{j<k} \int_0^1 \int_0^t M\left(u_{j-1}, u_{k-1}\right) \mathbb{1}_{\{u_{j-1} < s \le u_j\}} \mathbb{1}_{\{u_{k-1} < t \le u_k\}}\, dW_1(s)\, dW_2(t) \\
&\quad + \sum_{j=1}^{n} \int_0^1 \int_0^t M\left(u_{j-1}, u_{j-1}\right) \mathbb{1}_{\{u_{j-1} < s \le u_j\}} \mathbb{1}_{\{u_{j-1} < t \le u_j\}}\, dW_1(s)\, dW_2(t),
\end{aligned} \tag{71}$$

where the equality holds because the term with indices satisfying $j > k$ is 0. The first term on the right-hand side of (71) is

$$\begin{aligned}
&\sum_{j<k} \int_0^1 M\left(u_{j-1}, u_{k-1}\right)\left(W_1\left(u_j\right) - W_1\left(u_{j-1}\right)\right) \mathbb{1}_{\{u_{k-1} < t \le u_k\}}\, dW_2(t) \\
&= \sum_{j<k} M\left(u_{j-1}, u_{k-1}\right)\left(W_1\left(u_j\right) - W_1\left(u_{j-1}\right)\right)\left(W_2\left(u_k\right) - W_2\left(u_{k-1}\right)\right).
\end{aligned} \tag{72}$$

The second term on the right-hand side of (71) is

$$\sum_{j=1}^{n}\int_0^1 M\left(u_{j-1},u_{j-1}\right)\left(W_1(t)-W_1\left(u_{j-1}\right)\right)\mathbb{1}_{\{u_{j-1}<t\le u_j\}}\,dW_2(t)$$
$$=\sum_{j=1}^{n} M\left(u_{j-1},u_{j-1}\right)\left(\int_{u_{j-1}}^{u_j} W_1(t)\,dW_2(t)-W_1\left(u_{j-1}\right)\left(W_2\left(u_j\right)-W_2\left(u_{j-1}\right)\right)\right). \quad (73)$$

Combining (71), (72) and (73) yields

$$\int_0^1\int_0^t M_n(s,t)\,dW_1(s)\,dW_2(t)$$
$$=\sum_{j<k} M\left(u_{j-1},u_{k-1}\right)\left(W_1\left(u_j\right)-W_1\left(u_{j-1}\right)\right)\left(W_2\left(u_k\right)-W_2\left(u_{k-1}\right)\right)$$
$$+\sum_{j=1}^{n} M\left(u_{j-1},u_{j-1}\right)\left(\int_{u_{j-1}}^{u_j} W_1(t)\,dW_2(t)-W_1\left(u_{j-1}\right)\left(W_2\left(u_j\right)-W_2\left(u_{j-1}\right)\right)\right). \quad (74)$$

Similarly,

$$\int_0^1\int_0^s M_n(s,t)\,dW_2(t)\,dW_1(s)$$
$$=\sum_{j<k} M\left(u_{j-1},u_{k-1}\right)\left(W_2\left(u_j\right)-W_2\left(u_{j-1}\right)\right)\left(W_1\left(u_k\right)-W_1\left(u_{k-1}\right)\right)$$
$$+\sum_{j=1}^{n} M\left(u_{j-1},u_{j-1}\right)\left(\int_{u_{j-1}}^{u_j} W_2(t)\,dW_1(t)-W_2\left(u_{j-1}\right)\left(W_1\left(u_j\right)-W_1\left(u_{j-1}\right)\right)\right). \quad (75)$$

Combining (74) and (75) and rearranging terms gives

$$\begin{aligned}A_n=&\sum_{j,k=1}^{n} M\left(u_{j-1},u_{k-1}\right)\left(W_2\left(u_j\right)-W_2\left(u_{j-1}\right)\right)\left(W_1\left(u_k\right)-W_1\left(u_{k-1}\right)\right)\\&+\sum_{j=1}^{n} M\left(u_{j-1},u_{j-1}\right)\left[\int_{u_{j-1}}^{u_j} W_1(t)\,dW_2(t)+\int_{u_{j-1}}^{u_j} W_2(t)\,dW_1(t)\right.\\&\left.-\left(W_1\left(u_j\right)W_2\left(u_j\right)-W_1\left(u_{j-1}\right)W_2\left(u_{j-1}\right)\right)\right].\end{aligned} \quad (76)$$

Applying Itô's lemma to $W_1(t)W_2(t)$ yields

$$W_1\left(u_j\right)W_2\left(u_j\right)-W_1\left(u_{j-1}\right)W_2\left(u_{j-1}\right)=\int_{u_{j-1}}^{u_j} W_2(t)\,dW_1(t)+\int_{u_{j-1}}^{u_j} W_1(t)\,dW_2(t).$$

Together with (76), (68) follows. Since (69) and (70) are symmetric, we need only prove (69), and then (70) will follow similarly. By a similar argument to that of the derivation of (76), B_n equals

$$\sum_{j,k=1}^{n} M\left(u_{j-1},u_{k-1}\right)\left(W_1\left(u_j\right)-W_1\left(u_{j-1}\right)\right)\left(W_1\left(u_k\right)-W_1\left(u_{k-1}\right)\right)$$
$$+\sum_{j=1}^{n} M\left(u_{j-1},u_{j-1}\right)\left[2\int_{u_{j-1}}^{u_j} W_1(t)\,dW_1(t)-\left(W_1^2\left(u_j\right)-W_1^2\left(u_{j-1}\right)\right)+\frac{1}{n}\right].$$

Note that by applying Itô's lemma to $W_1^2(t)$

$$W_1^2\left(u_j\right) - W_1^2\left(u_{j-1}\right) = 2\int_{u_{j-1}}^{u_j} W_1(t)\,dW_1(t) + \frac{1}{n},$$

from which (69) immediately follows.

References

[1] E.S. Andersen, On sums of symmetrically dependent random variables, Scand. Actuar. J. 1953 (sup1) (1953) 123–138.
[2] E.S. Andersen, On the fluctuations of sums of random variables, Math. Scand. 1 (2) (1954) 263–285.
[3] E.S. Andersen, On the fluctuations of sums of random variables II, Math. Scand. 2 (2) (1955) 195–223.
[4] K.L. Chung, R.J. Williams, Introduction to Stochastic Integration, Springer, 1990.
[5] S. Douissi, K. Es-Sebaiy, F.G. Viens, Asymptotics of Yule's "nonsense correlation" for Ornstein-Uhlenbeck paths: A Wiener chaos approach, Electron. J. Stat. 16 (1) (2022) 3176–3211.
[6] P. Erdős, M. Kac, On certain limit theorems of the theory of probability, Bull. Amer. Math. Soc. 52 (4) (1946) 292–302.
[7] P.A. Ernst, L.C.G. Rogers, Q. Zhou, Yule's "nonsense correlation" solved: Part II, 2022, arXiv preprint arXiv:1909.02546.
[8] P.A. Ernst, L.A. Shepp, A.J. Wyner, Yule's "nonsense correlation" solved!, Ann. Statist. 45 (4) (2017) 1789–1809.
[9] C.S. Kubrusly, Spectral Theory of Bounded Linear Operators, Springer, 2020.
[10] H. Kuo, Introduction to Stochastic Integration, in: Universitext, Springer, ISBN: 9780387310572, 2006.
[11] J.R. Magnus, The exact moments of a ratio of quadratic forms in normal variables, Ann. D'Econ. Stat. 4 (1986) 95–109.
[12] M.B. Marcus, J. Rosen, Markov Processes, Gaussian Processes, and Local Times, Cambridge University Press, 2006.
[13] I. Nourdin, G. Peccati, Normal Approximations with Malliavin Calculus: From Stein's Method to Universality, Cambridge University Press, 2012.
[14] I. Nourdin, G. Peccati, The optimal fourth moment theorem, Proc. Amer. Math. Soc. 143 (7) (2015) 3123–3133.
[15] D. Nualart, E. Nualart, Introduction to Malliavin Calculus, Cambridge University Press, 2018.
[16] G. Orosi, A simple derivation of Faulhaber's formula, Appl. Math. E-Notes 18 (2018) 124–126.
[17] P.C.B. Phillips, Understanding spurious regressions in econometrics, J. Econometrics 33 (3) (1986) 311–340.
[18] P.C.B. Phillips, New tools for understanding spurious regressions, Econometrica 66 (6) (1998) 1299–1325.
[19] V. Pipiras, M.S. Taqqu, Long-Range Dependence and Self-Similarity, Cambridge University Press, 2017.
[20] S.S. Vallender, Calculation of the Wasserstein distance between probability distributions on the line, Theory Probab. Appl. 18 (4) (1974) 784–786.
[21] G.U. Yule, Why do we sometimes get nonsense-correlations between Time-Series?–a study in sampling and the nature of time-series, J. R. Stat. Soc. 89 (1) (1926) 1–63.