Yule's "nonsense correlation": moments and density

PHILIP A. ERNST[1,a], L.C.G. ROGERS[2,b] and QUAN ZHOU[3,c]

[1]*Department of Mathematics, Imperial College London, London, UK.* [a]*p.ernst@imperial.ac.uk*
[2]*Department of Mathematics, University of Cambridge, Cambridge, UK.* [b]*chris@statslab.cam.ac.uk*
[3]*Department of Statistics, Texas A&M University, College Station, TX, USA.* [c]*quan@stat.tamu.edu*

In 1926, G. Udny Yule ([29]) considered the following problem: given a sequence of pairs of random variables $\{X_k, Y_k\}$ ($k = 1, 2, \ldots, n$), and letting $X_i = S_i$ and $Y_i = S'_i$ where S_i and S'_i are the partial sums of two independent random walks, what is the distribution of the empirical correlation coefficient

$$\rho_n = \frac{\sum_{i=1}^n S_i S'_i - \frac{1}{n}(\sum_{i=1}^n S_i)(\sum_{i=1}^n S'_i)}{\sqrt{\sum_{i=1}^n S_i^2 - \frac{1}{n}(\sum_{i=1}^n S_i)^2}\sqrt{\sum_{i=1}^n (S'_i)^2 - \frac{1}{n}(\sum_{i=1}^n S'_i)^2}}?$$

Yule empirically observed the distribution of this statistic to be heavily dispersed and frequently large in absolute value, leading him to call it "nonsense correlation." This unexpected finding led to his formulation of two concrete questions, each of which would remain open for more than ninety years: (i) Find (analytically) the variance of ρ_n as $n \to \infty$ and (ii): Find (analytically) the higher order moments and the density of ρ_n as $n \to \infty$. In 2017, Ernst, Shepp and Wyner [12] considered the empirical correlation coefficient

$$\rho := \frac{\int_0^1 W_1(t)W_2(t)dt - \int_0^1 W_1(t)dt \int_0^1 W_2(t)dt}{\sqrt{\int_0^1 W_1^2(t)dt - \left(\int_0^1 W_1(t)dt\right)^2}\sqrt{\int_0^1 W_2^2(t)dt - \left(\int_0^1 W_2(t)dt\right)^2}}$$

of two *independent* Wiener processes W_1, W_2, the limit to which ρ_n converges weakly, as was first shown by the author of [19]. Using tools from integral equation theory, Ernst, Shepp and Wyner [12] closed question (i) by explicitly calculating the second moment of ρ to be .240522. This paper adopts a completely different approach to the same question, rooted in an earlier literature on the laws of quadratic functionals of Gaussian diffusions (in particular, [7,22]). This allows us to develop an Itô-formula approach from which we calculate expressions for the Laplace transform of ρ, leading to expressions for the moments which we evaluate up to order 16, thereby closing question (ii). This leads, for the first time, to an approximation to the density of Yule's nonsense correlation. The broad applicability of this approach is demonstrated by answering the corresponding questions when the pair of independent Brownian motions is replaced by a pair of correlated Brownian motions, or by two independent Ornstein-Uhlenbeck processes, or by two independent Brownian bridges. We conclude by extending the definition of ρ to the time interval $[0, T]$ for any $T > 0$ and prove a Central Limit Theorem for the case of two independent Ornstein-Uhlenbeck processes.

Keywords: Nonsense correlation; Ornstein-Uhlenbeck processes; volatile correlation; Wiener processes

1. Introduction

Given a sequence of pairs of random variables $\{X_k, Y_k\}$ ($k = 1, 2, \ldots, n$), how can we measure the strength of the dependence of X and Y? The classical Pearson correlation coefficient addresses this question in the case when the sequence $(X_k, Y_k)_{k=1}^n$ is an IID sequence, but when blindly calculated for sequences which are not IID, as arise (for example) when the sequence $(X_k, Y_k)_{k=1}^n$ is a random walk

or other time series, the resulting statistic is meaningless. Letting $X_i = S_i$ and $Y_i = S_i'$, where S_i and S_i' are the partial sums of two independent random walks, Yule [29] considered the distribution of the empirical correlation coefficient

$$\rho_n = \frac{\sum_{i=1}^n S_i S_i' - \frac{1}{n}(\sum_{i=1}^n S_i)(\sum_{i=1}^n S_i')}{\sqrt{\sum_{i=1}^n S_i^2 - \frac{1}{n}(\sum_{i=1}^n S_i)^2}\sqrt{\sum_{i=1}^n (S_i')^2 - \frac{1}{n}(\sum_{i=1}^n S_i')^2}},$$

which he found to be both heavily dispersed and frequently large in absolute value. This led Yule to call this distribution "nonsense correlation" and to formulate two concrete questions about its distribution, each of which would remain open for more than ninety years:

1. Find (analytically) the variance of ρ_n as $n \to \infty$;
2. Find (analytically) the higher order moments and the density of ρ_n as $n \to \infty$.

Ernst, Shepp and Wyner [12] refer to this "nonsense correlation" as "volatile" correlation since its distribution is both heavily dispersed and is frequently large in absolute value.

Despite the prominence of Yule's 1926 paper, his findings would remain "isolated" from the literature until 1986 (see [1]), when the authors of [15] and [19] confirmed many of the empirical claims of "spurious regression" made by the authors of [14]. In particular, [19] 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 ([20]), the same author provided an explanation of such spurious regressions in terms of orthonormal representations of the Karhunen Loève type.

In 2017, Ernst, Shepp and Wyner [12] considered the empirical correlation coefficient

$$\rho := \frac{\int_0^1 W_1(t)W_2(t)dt - \int_0^1 W_1(t)dt \int_0^1 W_2(t)dt}{\sqrt{\int_0^1 W_1^2(t)dt - \left(\int_0^1 W_1(t)dt\right)^2}\sqrt{\int_0^1 W_2^2(t)dt - \left(\int_0^1 W_2(t)dt\right)^2}}$$

of two *independent* Wiener processes W_1, W_2, the limit to which ρ_n converges weakly, as originally shown in [19]. Using tools from integral equation theory, Ernst, Shepp and Wyner [12] closed Question 1 by explicitly calculating the second moment of ρ to be .240522. The present paper adopts a completely different approach to the question, rooted in an earlier literature on the laws of quadratic functionals of Gaussian diffusions (in particular, [7,22]). This allows us to develop an Itô-formula approach from which we calculate expressions for the Laplace transform of ρ, leading to expressions for the moments which we evaluate up to order 16, thereby closing Question 2. This leads, for the first time, to an approximation to the density of Yule's nonsense correlation. The broad applicability of this approach is demonstrated by answering the corresponding questions when the pair of independent Brownian motions is replaced by a pair of correlated Brownian motions, or by two independent Ornstein-Uhlenbeck (OU) processes, or by two independent Brownian bridges. We conclude by extending the definition of ρ to the time interval $[0, T]$ for any $T > 0$ and prove a Central Limit Theorem for the case of two independent Ornstein-Uhlenbeck processes. Indeed, this closes all previously open problems raised in Section 3.3 of [12].

We proceed with some necessary notation. Let $(X_t)_{0 \le t \le T}$ be some process with values in $\mathbb{R}^d$, defined over a fixed time interval $[0, T]$. Define the random variables

$$\bar{X} := T^{-1} \int_0^T X_s \, ds, \qquad Y := \int_0^T (X_s - \bar{X})(X_s - \bar{X})^T \, ds, \qquad (1)$$

with values in $\mathbb{R}^d$ and $\mathbb{M}^d$ respectively, where $\mathbb{M}^d$ is the space of $d \times d$ real matrices. Let Y_{ij} be the (i, j)-th entry of the matrix Y. In the case $d = 2$, when X is a two-dimensional Wiener process, Yule's nonsense correlation can be expressed as

$$\rho := \frac{Y_{12}}{\sqrt{Y_{11}}\sqrt{Y_{22}}}. \tag{2}$$

Scaling properties of Brownian motion show that the law of ρ does not depend on the choice of $T > 0$. Ernst, Shepp and Wyner [12] explicitly calculated the second moment of Yule's nonsense correlation. However, their methodology did not successfully extend to explicitly calculating higher order moments of ρ.

The key vision of [12] was to calculate Laplace transform of the trivariate object formed of the three quadratic/bilinear forms of W, and to express that transform using Fredholm integral equations. In the present work, we rely instead on the characterization of the moment generating function of the random vector (Y_{11}, Y_{12}, Y_{22}). This approach inherits from an older and well-developed literature, on the laws of quadratic functionals of Brownian motion. There is a fine survey ([8]) which presents the state of the subject as it was in 1997. A range of techniques is available to characterize the laws of quadratic functionals of Brownian motion, including:

1. eigenfunction expansions — see, for example, [4,5,12,13,17,18];
2. identifying the covariance of the Gaussian process as the Green function of a symmetrizable Markov process — see, for example, [5,10];
3. stochastic Fubini relations — see, for example, [7,8];
4. Itô's formula — see [22].

The first of these techniques is historically the first; using it to deliver a simple closed-form solution depends on spotting a simpler form for an infinite expansion. The second works well if we can see a Markov process whose Green function is the covariance of the Gaussian process of interest. The third again requires an insight to transform the problem of interest into a simpler equivalent. The fourth, much less often exploited, deals conclusively with settings where the Gaussian process arises as the solution of a linear stochastic differential equation (SDE); this is the approach we use in the present paper.

Section 5 studies the asymptotic properties of ρ as $T \to \infty$. The tools used in this section are more straightforward, and the results of this section do not rely on the new techniques introduced in Section 2 and Section 3. For this discussion, we will write $X(T), Y(T)$ in place of X, Y defined at (1) and $\rho(T)$ in place of ρ defined at (2) to emphasize their dependence on the time horizon T. In the case of Wiener processes, by the property of self-similarity, it is straightforward to show that $\rho(1)$ and $\rho(T)$ have the same distribution. But for Gaussian processes which are not self-similar, $\rho(T)$ will depend on the value of T. Section 5 investigates this statistic's asymptotic behavior as $T \to \infty$. The key result is given by Theorem 5.2, which proves that, in the case of two independent Ornstein-Uhlenbeck processes, $\sqrt{T}\rho(T)$ converges in distribution as $T \to \infty$ to a zero-mean Gaussian.

To summarize then, the main contributions of this paper are, in order of importance:

1. We explicitly calculate all moments of Yule's nonsense correlation (up to order 16). This is only possible due to Proposition 1, in which the moments of ρ are given within explicit expressions involving partial derivatives of a trivariate Laplace transform. With these moments in hand, we provide the first density approximation to Yule's nonsense correlation. *This closes the final long-standing open question on the distribution of Yule's nonsense correlation from Yule's 1926 paper ([29]).*

2. We develop the most general version Theorem 2.1 of the 'mechanical' Itô-formula approach of [22], and show how it provides straightforward proofs of all the results of this paper.

3. We find the asymptotic behavior of $\rho(T)$ for the case of independent OU processes. We shall see that the variance of $\sqrt{T}\rho(T)$ shrinks to zero as the mean reversion parameter tends towards infinity. This shows that the empirical correlation coefficient may be viewed as "sensible" for testing independence of pairs of OU processes.

2. Quadratic functionals of Gaussian diffusions

We shall use the notation $\mathbb{S}_+^d$ for the space of *strictly* positive-definite symmetric $d \times d$ matrices, with the canonical ordering $A \geq B$ meaning that $A - B$ is non-negative definite. The notation $a \cdot b$ for $a, b \in \mathbb{R}^d$ denotes the scalar product of the two vectors. An "overdot" denotes the derivative with respect to t. In this manuscript, we shall use X_t and $X(t)$ interchangeably. The main result is the following.

Theorem 2.1. *Suppose that $\sigma : [0,T] \mapsto \mathbb{M}^d$ is a bounded measurable function, and that X solves*

$$dX_t = \sigma_t dW_t, \tag{3}$$

where W is d-dimensional Brownian motion. We write $\Sigma_t = \sigma_t \sigma_t^\top$. Suppose also that $Q : [0,T] \to \mathbb{S}_+^d$ and $z : [0,T] \to \mathbb{R}^d$ are bounded measurable functions such that Q^{-1} is also bounded. Define

$$\ell(t) := \frac{1}{2} X(t) \cdot Q(t) X(t) + z(t) \cdot X(t), \tag{4}$$

$$F(t,x) := E\left[\exp\left\{-\int_t^T \ell(s)ds - \ell(T)\right\} \,\middle|\, X(t) = x\right]. \tag{5}$$

Then $F(t,x)$ is given explicitly as

$$F(t,x) = \exp\left(-\frac{1}{2} x \cdot V(t)x - b(t) \cdot x - \gamma(t)\right), \tag{6}$$

where V, b, γ are obtained as the unique solutions to the system of ordinary differential equations (ODEs)

$$\dot{V} = V\Sigma V - Q, \tag{7}$$

$$\dot{b} = V\Sigma b - z, \tag{8}$$

$$2\dot{\gamma} = b^\top \Sigma b - \text{tr}\,(V\Sigma), \tag{9}$$

subject to the boundary conditions $V(T) = Q(T)$, $b(T) = z(T)$, $\gamma(T) = 0$.

Proof. (i) Notice that ℓ is bounded below by $-\frac{1}{2} z \cdot Q^{-1} z$, which by hypothesis is bounded below by some constant. Therefore F defined by (5) is bounded.

(ii) The ODE (7) has a unique solution up to possible explosion, as the coefficients are locally Lipschitz. We claim that this solution remains positive-definite for $t \leq T$. Since $Q(T) \in \mathbb{S}_+^d$, it has to be that there exists some $\varepsilon > 0$ such that $V(t) \in \mathbb{S}_+^d$ for all $t \in [T - \varepsilon, T]$. If V does not remain positive definite, then there exists some non-zero $w \in \mathbb{R}^d$ and a greatest $t^* \leq T - \varepsilon < T$ such that $w \cdot V(t^*)w \leq 0$. But we

see from (7) that $w \cdot \dot{V}(t^*)w \leq -w \cdot Q(t^*)w < 0$, contradicting the definition of t^*. Hence V remains positive-definite all the way back to possible explosion. However, we have that

$$V(t) = Q(T) + \int_t^T \{ Q(s) - V(s)\Sigma(s)V(s) \} \, ds \leq Q(T) + \int_t^T Q(s) \, ds.$$

So by hypothesis V is bounded above and no explosion happens. Since V is continuous on $[0,T]$ and positive-definite everywhere, it follows that V is *uniformly* positive-definite on $[0,T]$, that is, V^{-1} remains bounded.

It now follows easily that b and γ defined by (8) and (9) are unique, continuous and bounded.

(iii) Now define the process

$$Z_t = \tfrac{1}{2}X_t \cdot V_t X_t + b_t \cdot X_t + \gamma_t, \tag{10}$$

and develop

$$dZ_t = (V_t X_t + b_t, \sigma_t \, dW_t \, dt) + \tfrac{1}{2}\mathrm{tr}\,(V_t \Sigma_t)dt + \left\{ \tfrac{1}{2}X_t \cdot \dot{V}_t X_t + \dot{b}_t X_t + \dot{\gamma}_t \right\}dt,$$

$$d\langle Z \rangle_t = (V_t X_t + b_t) \cdot \Sigma_t (V_t X_t + b_t)dt.$$

Now consider the process

$$M(t) = \exp\left(-\tfrac{1}{2} \int_0^t \ell(s) \, ds - Z(t) \right). \tag{11}$$

Notice that M is bounded, because ℓ is bounded below, and so is Z since we have proved that V^{-1}, b and γ are all bounded on $[0,T]$. Developing M using Itô's formula, with the symbol $\doteq$ denoting that the two sides of the equation differ by a local martingale and omitting explicit appearance of the time parameter, we obtain

$$\frac{dM_t}{M_t} = -dZ_t + \tfrac{1}{2}d\langle Z \rangle_t - \tfrac{1}{2}X_t \cdot Q_t X_t dt - z_t \cdot X_t dt$$

$$\doteq \Big\{ -\tfrac{1}{2}\mathrm{tr}\,(V_t \Sigma_t) - \tfrac{1}{2}X_t \cdot \dot{V}_t X_t - \dot{b}_t X_t - \dot{\gamma}_t +$$

$$+ \tfrac{1}{2}(V_t X_t + b_t) \cdot \Sigma_t (V_t X_t + b_t) - \tfrac{1}{2}X_t \cdot Q_t X_t - z_t \cdot X_t \Big\} \, dt$$

$$= 0$$

because of (7), (8) and (9). Thus M is a local martingale, which is also bounded on $[0,T]$ because of the boundedness assumptions. Thus M is a bounded martingale, and the result follows. □

Remark 1. *Theorem 2.1 is in part inspired by Rogers and Shi ([22, Lemma 1]), which exhibits (in the simplest possible setting) the general principle that expectations of exponentials of quadratic functionals of Gaussians will be exponential-quadratic. The statement of the lemma is recalled next.*

Lemma 2.2. *[22, Lemma 1] Let X be a Gaussian random vector in $\mathbb{R}^d$ with mean 0 and covariance V, and let S be a non-negative definite symmetric matrix. Then, for any $a \in \mathbb{R}^d$,*

$$E\left[\exp\left\{-\frac{1}{2}(X+a)^T S(X+a)\right\}\right] = det(I+SV)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}a^T(I+SV)^{-1}Sa.\right\}$$

Theorem 2.1 extends easily to the situation where X is the solution of a linear SDE.

Theorem 2.3. *Suppose that σ, $B : [0,T] \mapsto \mathbb{M}^d$ and $\delta : [0,T] \mapsto \mathbb{R}^d$ are bounded measurable functions, and that X solves*

$$dX_t = \sigma_t\, dW_t + (B_t X_t + \delta_t)\, dt. \tag{12}$$

Suppose that $Q : [0,T] \to \mathbb{S}_+^d$ and $z : [0,T] \to \mathbb{R}^d$ are bounded measurable functions such that Q^{-1} is also bounded, and suppose that ℓ and F are defined as before at (4), (5). Then $F(t,x)$ is given explicitly as

$$F(t,x) = \exp\left(-\frac{1}{2}x \cdot V(t)x - b(t)\cdot x - \gamma(t)\right), \tag{13}$$

where V, b, γ are obtained as the unique solutions to the system of ODEs,

$$\dot{V} = V\Sigma V - (VB + B^\top V) - Q, \tag{14}$$

$$\dot{b} = (V\Sigma - B^\top)b - V\delta - z, \tag{15}$$

$$2\dot{\gamma} = b^\top\Sigma b - \text{tr}(V\Sigma) - \delta^\top b, \tag{16}$$

subject to the boundary conditions $V(T) = Q(T)$, $b(T) = z(T)$, $\gamma(T) = 0$.

Proof. The coefficients of the SDE (12) are globally Lipschitz, so it is a standard result (see, for example, [23] Theorem V.11.2) that the SDE has a unique strong solution. If we now set

$$\tilde{X}_t = A_t X_t + c_t, \tag{17}$$

where A and c solve

$$\dot{A}_t + A_t B_t = 0, \qquad A(0) = I, \tag{18}$$

$$\dot{c}_t + A_t \delta_t = 0, \qquad c(0) = 0, \tag{19}$$

then a few simple calculations show that

$$d\tilde{X}_t = A_t \sigma_t\, dW_t,$$

and Theorem 2.1 applies. The equations (14), (15) and (16) are easily checked to be the analogs of (7), (8) and (9) respectively.

$\square$

Remark 2. *We will want to apply Theorem 2.1 to situations where $Q(T) = 0$. This is a simple limiting case of the problem where we take $Q(T) = \varepsilon I$ and let $\varepsilon \downarrow 0$. In a little more detail, we let V^ε, b^ε, γ^ε denote the solution to (14)-(16) with boundary condition $Q(T) = \varepsilon I$, and we write*

$$q_t^\varepsilon : x \mapsto \tfrac{1}{2}x \cdot V^\varepsilon(t)x + b^\varepsilon(t)\cdot x + \gamma^\varepsilon(t), \tag{20}$$

for the quadratic form $-\log F(t,x)$. Evidently $q_t^\varepsilon(x)$ is decreasing in ε for each x and t, and from this it follows easily that limits of $V^\varepsilon(t), b^\varepsilon(t), \gamma^\varepsilon(t)$ exist for each t and determine F for the limit case when $Q(T) = 0$.

Remark 3. *Theorem 2.3 is a special case of the Feynman-Kac formula; the fact that the process M defined in (11) is a martingale is equivalent to the Feynman-Kac formula, and is valid for any additive functional ℓ of the diffusion X. However, without the special linear form of the SDE for X and the quadratic form of the additive functional ℓ it is rare that any explicit solution can be found for F.*

Remark 4. *If σ is constant, we may assume that $\sigma = I$, the identity matrix. To see this, let $\hat{X}_t = \Sigma_t^{-1/2} X_t$, and note that the diffusion process $\hat{X}$ solves the linear SDE*

$$d\hat{X}_t = (\Sigma_t^{-1/2} B_t \Sigma_t^{1/2} \hat{X}_t + \Sigma_t^{-1/2} \delta_t) dt + dW_t.$$

Letting $\hat{Q} = \Sigma^{1/2} Q \Sigma^{1/2}$ and $\hat{z} = \Sigma^{1/2} z$ we obtain

$$\ell(t) = \frac{1}{2} X^\top(t) Q(t) X(t) + z^\top(t) X(t) = \frac{1}{2} \hat{X}(t)^\top \hat{Q}(t) \hat{X}(t) + \hat{z}^\top(t) \hat{X}(t),$$

and thus we can work with the process $\hat{X}$ instead of X. However, it seems simpler to provide the full form of the solution for the SDE (12) rather than a reduced form which then requires a translation back to the original problem.

Remark 5. *Although Theorem 2.3 deals with the general case where Q, z are measurable functions, in the remainder of this paper we only need invoke Theorem 2.3 for the special case where Q and z are constants. For this reason, we will sometimes use the alternative expanded notation*

$$F(t,x) := F(t,x;Q,z) \tag{21}$$

when we want to make explicit the dependence of F on the coefficients Q and z appearing in ℓ.

3. Computing the moments of ρ

Henceforth, we deal exclusively with cases where

$$d = 2.$$

Recall the definition (1) of the 2×2 random matrix Y. Let ϕ be the moment generating function of the joint distribution of (Y_{11}, Y_{12}, Y_{22}), which can be expressed using quadratic functionals of X as

$$\phi(S) := E\left[\exp\left\{-\frac{1}{2}(s_{11} Y_{11} + 2 s_{12} Y_{12} + s_{22} Y_{22})\right\}\right]$$
$$= E\left[\exp\left\{-\frac{1}{2} \int_0^T (X(u) - \bar{X}) \cdot S(X(u) - \bar{X}) du\right\}\right]. \tag{22}$$

Here, S is a 2×2 positive-definite symmetric matrix with entries denoted by s_{ij} ($i, j = 1, 2$). As we shall show in the following proposition, the function ϕ is all we shall need to evaluate the moments of ρ.

Proposition 1. *Let ρ be as given in (2) and $\phi(s_{11}, s_{12}, s_{22}) = \phi(S)$ be as given in (22). For $k = 0, 1, 2, \ldots$, we have*

$$E\rho^k = \frac{(-1)^k}{2^k \Gamma(k/2)^2} \int_0^\infty \int_0^\infty s_{11}^{k/2-1} s_{22}^{k/2-1} \frac{\partial^k \phi}{\partial s_{12}^k} (s_{11}, 0, s_{22}) \, ds_{11} \, ds_{22}. \tag{23}$$

Proof. It is well known that the moments of a random variable can be obtained by differentiating the moment generating function, given it exists ([2]). Now note that for any fixed nonnegative s_{11}, s_{22}, there exists $\epsilon > 0$ such that $S = [s_{ij}]$ is positive semi-definite for any $s_{12} \in [-\epsilon, \epsilon]$ and thus $\phi(s_{11}, s_{12}, s_{22}) \leq 1$. Hence, the partial derivative with respect to s_{12} exists at $s_{12} = 0$. Applying Fubini's Theorem we obtain

$$(-1)^k \frac{\partial^k \phi}{\partial s_{12}^k} (s_{11}, 0, s_{22}) = E\left[Y_{12}^k \exp\left\{ -\frac{1}{2}(s_{11} Y_{11} + s_{22} Y_{22}) \right\} \right].$$

Next, recall that by the definition of Gamma function, for any $\alpha > 0$,

$$y^{-\alpha} = \frac{1}{\Gamma(\alpha)} \int_0^\infty t^{\alpha-1} e^{-ty} dt = \frac{1}{2^\alpha \Gamma(\alpha)} \int_0^\infty s^{\alpha-1} e^{-sy/2} ds.$$

Since $\rho^k = Y_{12}^k Y_{11}^{-k/2} Y_{22}^{-k/2}$, we can apply the above formula to obtain (23) (by Tonelli's Theorem, the order of integration can always be exchanged). □

Remark 6. *The idea of using the moment generating function to compute negative moments or moments of the ratio of two random variables has been widely used in the literature. See [6,16,24].*

So we see that the distribution of ρ is determined by (22), from which moments can in principle be derived using Proposition 1; but we need to get hold of the expression in (22). This is where Theorem 2.3 comes in. If X is a solution of a linear SDE (12), starting at $X_0 = 0$ to fix the discussion, and we set

$$Q(t) = S, \quad z(t) = a \in \mathbb{R}^2 \; \forall \; 0 \leq t < T, \qquad Q(T) = 0, \quad z(T) = 0,$$

then Theorem 2.3 tells us how to compute

$$F(0, 0; a) = E\left[\exp\left\{ -\int_0^T \{\tfrac{1}{2} X(u) \cdot SX(u) + a \cdot X(u)\} \, du \right\} \right] \tag{24}$$

$$= \exp(-\gamma(0; a)), \tag{25}$$

where we have written $F(t, x; a)$ and $\gamma(0; a)$ to emphasize dependence on a. If we now integrate over a with a $N(0, T^{-1}S)$ distribution the right-hand side of (24) becomes

$$E\left[\exp\left\{ -\int_0^T \tfrac{1}{2} X(u) \cdot SX(u) \, du + \tfrac{1}{2} T\bar{X} \cdot S\bar{X} \right\} \right] = \phi(S). \tag{26}$$

The strategy now should be clear. In any particular application, we use Theorem 2.3 to obtain $\gamma(t; a)$ as explicitly as possible, and then we integrate (25) over a to find $\phi(S)$.

4. Examples

In this section we will carry out the program just outlined in four examples, and obtain remarkably explicit expressions for everything we need.

In the first three examples, the two-dimensional diffusion process X has two special properties:

(i) The law of $(RX_t)_{0 \le t \le 1}$ is the same as the law of $(X_t)_{0 \le t \le 1}$ for any fixed rotation matrix R;
(ii) The two components of X are independent.

Consequently, if we abbreviate $X^1(t) = x_t$, $\bar{x} = \int_0^1 x_s \, ds$, and define

$$\psi(v) = E \left[\exp\left\{ -\tfrac{1}{2} \int_0^1 v(x_u - \bar{x})^2 du \right\} \right], \tag{27}$$

it follows that the function $\phi(S)$ defined at (22) simplifies to the product

$$\phi(S) = \psi(\theta_1^2) \, \psi(\theta_2^2), \tag{28}$$

where θ_1^2, θ_2^2 are the eigenvalues of S. This observation simplifies the solution of the differential equations (14)-(16) considerably, reducing everything to a one-dimensional problem.

The final example, that of correlated Brownian motion, reduces to the Brownian example by a linear transformation.

4.1. Brownian motion

For a standard one-dimensional Brownian motion $x(t)$, consider the function $F(t, x; \theta^2, z)$ where $\theta \ge 0$ and $z \in \mathbb{R}$. By Theorem 2.3, the solution has the following form (the subscript "Bm" is Brownian motion)

$$F_{\text{Bm}}(t, x; \theta^2, z) = \exp\left\{ -\tfrac{1}{2} V(t) x^2 - b(t) x - \gamma(t) \right\},$$

which leads to the following system of ordinary differential equations

$$\dot{V} - V^2 + \theta^2 = 0,$$

$$\dot{b} - Vb + z = 0,$$

$$2\dot{\gamma} - b^2 + V = 0.$$

Using the boundary condition $V(T) = 0$, we obtain

$$V(t) = \theta \tanh \theta\tau,$$

where $\tau = T - t$. Using the condition $b(T) = 0$, one can show that the solution for b is

$$b(t) = \frac{z}{\theta^2} V(t) = \frac{z}{\theta} \tanh \theta\tau.$$

Solving the third ODE, we obtain

$$2\gamma(t) = \log \cosh \theta\tau + \frac{z^2}{\theta^3} \left(-\theta\tau + \tanh \theta\tau \right),$$

Table 1. Numerical values of the moments of Yule's nonsense correlation for two independent Wiener processes ($T = 1$).

k	2	4	6	8
$E\rho^k$	0.240522	0.109177	0.060862	0.037788

k	10	12	14	16
$E\rho^k$	0.025114	0.017504	0.012641	0.009385

and thus

$$F(0,0;\theta^2, z) = \exp\left\{-\frac{z^2}{2\theta^3}\left(-\theta T + \tanh\theta T\right) - \frac{1}{2}\log\cosh\theta T\right\}.$$

As at (26), we now mix this expression over $z \sim N(0, \theta^2/T)$ to discover that in this example the function ψ (defined at (27)) takes the simple explicit form

$$\psi_{\mathrm{Bm}}(\theta^2) = \left(\frac{\theta T}{\sinh\theta T}\right)^{1/2}, \tag{29}$$

which (after appropriate scaling), is identical to the relation in the third display on p.577 of [7], which is in fact a special case of the result (3.10) on p. 251 of [5].

From (28) therefore, the moment generating function $\phi(S)$ is given by

$$\phi_{\mathrm{Bm}}(S) = \left(\frac{\theta_1\theta_2 T^2}{\sinh\theta_1 T \sinh\theta_2 T}\right)^{1/2}, \tag{30}$$

where θ_1^2, θ_2^2 are the eigenvalues of S. These eigenvalues are given in terms of the entries of S as

$$\theta_i^2 = \frac{1}{2}\left(s_{11} + s_{22} \pm \sqrt{(s_{11} - s_{22})^2 + 4s_{12}^2}\right), \tag{31}$$

where s_{ij} is the (i, j)-th entry of S. We note that the formula in (30) is given in more generality in formula (3.b) on page 578 of [7].

Consider $E(\rho^k)$ for $k = 0, 1, 2, \ldots$. Note that for any k, the expectation always exists since $\rho \in [-1, 1]$. Further, all the odd moments, i.e. $E(\rho^{2k+1})$, are zero by symmetry. To compute an even moment of ρ, we apply formula (23). For example, consider the second moment. Straightforward but tedious calculations yield

$$E\rho^2 = \int_0^\infty \int_0^v \frac{uv\sqrt{uv}}{(v^2 - u^2)\sqrt{\sinh u \sinh v}}\left(\frac{1}{u\tanh u} - \frac{1}{v\tanh v} - \frac{1}{u^2} + \frac{1}{v^2}\right) du\, dv, \tag{32}$$

where we have applied a change of variables, $u = \sqrt{s_{11}}, v = \sqrt{s_{22}}$. Note that this is exactly the same as the formula provided in Ernst, Shepp and Wyner [12, Proposition 3.4].

For higher-order moments, the calculation of $\partial^k\phi/\partial s_{12}^k$ is extremely laborious. We use `Mathematica` to perform symbolic high-order differentiation and then the two-dimensional numerical integration. The numerical results are summarized in Table 1. The choice of T is irrelevant since the distribution of $\rho(T)$ does not depend on T.

We proceed to use the numerical values of $E(\rho^k)$ to approximate the probability density function of ρ, which we denote by f. Consider a polynomial approximation

$$\hat{f}_k(\rho) = a_{k,0} + a_{k,1}\rho + a_{k,2}\rho^2 + \cdots + a_{k,k}\rho^k.$$

The coefficients $(a_{k,0}, \ldots, a_{k,k})$ can be computed by matching the first $k + 1$ moments of ρ (including the zero moment which is always equal to 1). This is also known as the Legendre series expansion of f, which minimizes the integrated squared error among all polynomial approximants with degree k ([3,21]). The rate of convergence depends on the modulus of continuity of f (see, for example, Saxena [25],Suetin [26],Wang [27],Wang and Xiang [28]); the theoretical properties of the latter are difficult to investigate via the moment generating function ϕ. Below we perform some numerical experiments to show that this polynomial approximation strategy indeed provides an efficient solution to calculating the distribution of ρ.

Recalling that $E(\rho^k) = 0$ for odd k, we have, for for $k = 0, 2, 4, \ldots$, that $\hat{f}_k = \hat{f}_{k+1}$. It thus suffices to consider $\hat{f}_k$, for $k = 0, 2, 4, \ldots$. To determine whether the Legendre series expansion has "converged," a commonly used diagnostic is the quantity $\hat{\varepsilon}_k = \sup_{\rho \in (-1,1)} |\hat{f}_k(\rho) - \hat{f}_{k-1}(\rho)|$, which needs to be sufficiently small for the algorithm to stop. In the left panel of Figure 1, we show how $\hat{\varepsilon}_k$ changes with k for two independent Wiener processes, from which we see that $\hat{\varepsilon}_k$ quickly tends to zero as k increases. The plot also suggests that $\hat{f}_4$ may give a reasonably good approximation to f since $\hat{\varepsilon}_6 = |\hat{f}_6 - \hat{f}_4|_\infty$ is just about 0.01. Indeed, we observe that the overall shape of $\hat{f}_4$ is very similar to that of $\hat{f}_{12}$, but $\hat{f}_4$ is more rough. Below we give the expressions for $\hat{f}_4, \hat{f}_6$ and $\hat{f}_8$, which constitute *the first density approximations* to Yule's nonsense correlation, and thereby *solves the second (and final) of the two longstanding open questions raised by Yule's 1926 paper ([29])!*

$$\hat{f}_4(\rho) = 0.59081 + 0.31001\rho^2 - 0.97075\rho^4,$$

$$\hat{f}_6(\rho) = 0.60057 + 0.10518\rho^2 - 0.35627\rho^4 - 0.45062\rho^6,$$

$$\hat{f}_8(\rho) = 0.61200 - 0.30638\rho^2 + 1.9073\rho^4 - 4.3742\rho^6 + 2.1019\rho^8.$$

Finally, we compare our moment-based polynomial approximation with Monte Carlo estimates. We fix $T = 1$ and discretize time using step size 10^{-4}. The gray bars in the right panel of Figure 1 give the histogram of ρ from 10^7 replicates. The red curve is the 12th-order approximation $\hat{f}_{12}$, which agrees very well with the empirical distribution. Note that the moment-based polynomial approximation is much more efficient and accurate than the empirical density function obtained from sampling; the latter has two sources of errors, one from Monte Carlo sampling and the other from time discretization. From the plot, we see that the distribution of ρ is heavily dispersed and frequently large in absolute value, and that the density remains approximately constant for $\rho \in (-0.5, 0.5)$.

We have only provided the numerical values of $E(\rho^k)$ up to $k = 16$. This has been done for two reasons. Firstly, for practical purposes such density approximation, moments of even higher orders are of much less interest. Secondly, the calculations of the derivative $\partial^k \phi / \partial s_{12}^k$ and the double integral in (23) become extremely slow and require massive memory for $k \geq 16$.

4.2. Ornstein-Uhlenbeck process

Consider a one-dimensional Ornstein-Uhlenbeck (OU) process which starts from $X(0) = 0$ and evolves according to the following stochastic differential equation

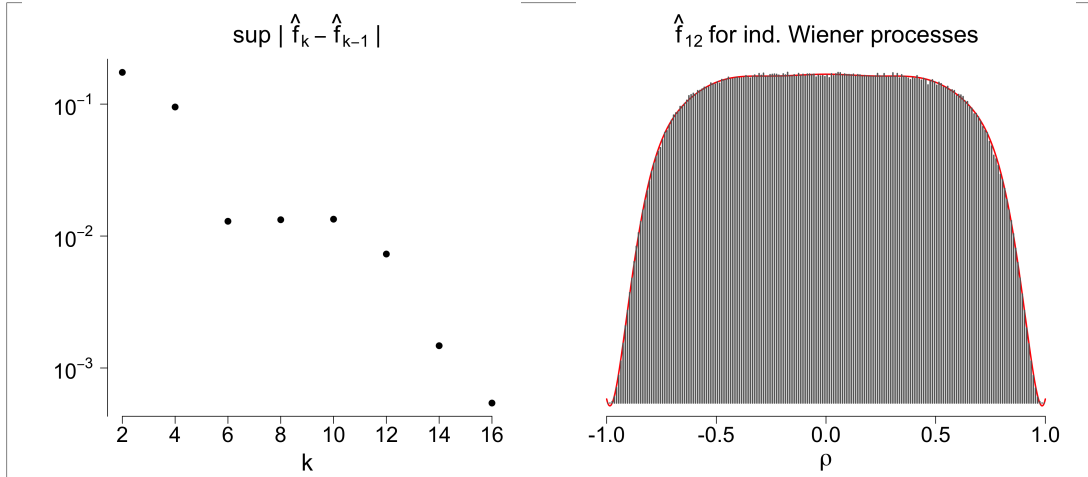$$dX_t = -rX_t dt + dW_t, \qquad r \in (0, \infty). \tag{33}$$

Figure 1: Moment-based polynomial approximations for the density function of ρ for two independent Wiener processes. $\hat{f}_k$ denotes the approximant with degree k. In the right-panel, the red curve corresponds to $\hat{f}_{12}$. The gray bars represent the empirical frequencies from 10^7 Monte Carlo simulations.

By Theorem 2.3, the solution has the form

$$F_{\mathrm{OU}}(t, x; \theta^2, z) = \exp\left\{-\frac{1}{2}V(t)x^2 - b(t)x - \gamma(t)\right\},$$

which can be obtained by solving the following system of ODEs

$$\dot{V} - 2rV - V^2 + \theta^2 = 0,$$

$$\dot{b} - (V + r)b + z = 0,$$

$$2\dot{\gamma} - b^2 + V = 0.$$

Using $V(T) = 0$, we solve the first equation to obtain

$$V(t) = \frac{\theta^2}{r + \eta \coth \eta\tau},$$

where $\eta = \sqrt{r^2 + \theta^2}$ and $\tau = T - t$. The second differential equation is first-order linear, so it can be solved explicitly; after some straightforward calculations we obtain

$$b(t) = \frac{z}{r + \eta \coth \eta\tau}\left(1 + \frac{r}{\eta}\tanh\frac{\eta\tau}{2}\right).$$

Finally, solving the last differential equation yields

$$2\gamma(t) = \frac{z^2}{\theta^2}\left\{\frac{\left(1 + \frac{r}{\eta}\tanh\frac{\eta\tau}{2}\right)^2}{r + \eta \coth \eta\tau} - \frac{r^2}{\eta^3}\tanh\frac{\eta\tau}{2} - \frac{\theta^2\tau}{\eta^2}\right\} - r\tau + \log\left(\cosh\eta\tau + \frac{r}{\eta}\sinh\eta\tau\right).$$

Table 2. Numerical values of the second moment of the empirical correlation coefficient for two independent Ornstein-Uhlenbeck processes with mean reversion parameter r ($T = 1$).

r	0.1	0.2	0.3	0.4	0.5	1
$E\rho^2$	0.23209	0.22438	0.21734	0.21091	0.20504	0.18231

r	2	5	10	20	50	100
$E\rho^2$	0.15583	0.11454	0.07627	0.04404	0.01907	0.00971
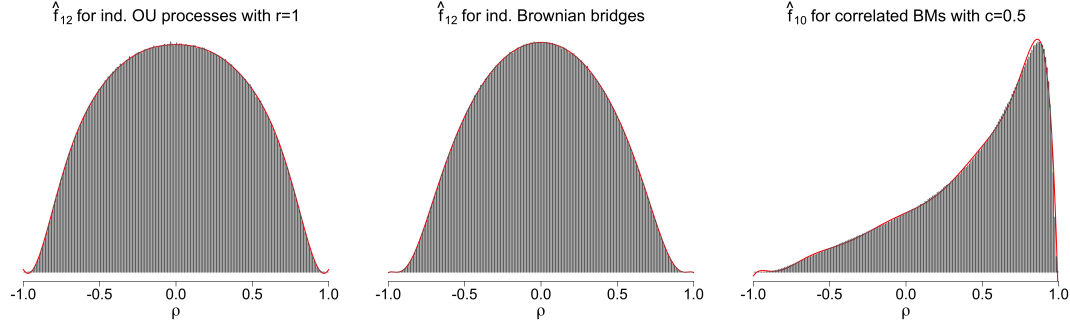
Figure 2: Moment-based polynomial approximations for the density function of ρ. The gray bars represent the empirical frequencies from 10^7 Monte Carlo simulations.

Mixing over z with a Gaussian law as before, and using $\tanh(x/2) = \coth x - \operatorname{csch} x$, we obtain

$$\psi_{\text{OU}}(\theta^2; r) = \sqrt{T} e^{rT/2} \left\{ \frac{\theta^2}{\eta^4} [2r(\cosh \eta T - 1) + \eta \sinh \eta T] + \frac{r^2 T}{\eta^3} [\eta \cosh \eta T + r \sinh \eta T] \right\}^{-1/2}.$$

If we have two independent Ornstein-Uhlenbeck processes $X_1(t), X_2(t)$ which both start at zero and have common mean reversion parameter r, one can check that an orthogonal transformation of $X = (X_1, X_2)$ leaves the joint distribution invariant. Indeed, the new two-dimensional process follows exactly the same SDE. Hence, the moment generating function in this case can be computed by

$$\phi_{\text{OU}}(S; r) = \psi_{\text{OU}}(\theta_1^2; r)\psi_{\text{OU}}(\theta_2^2; r),$$

where θ_1^2, θ_2^2 are the eigenvalues of S.

In Table 2 above we give the numerical values of $E\rho^2$ for independent Ornstein-Uhlenbeck processes with mean reversion parameter r ($T = 1$). Note that as $r \to \infty$, the processes converge to constant zero and thus $E\rho^2$ (the variance of ρ) goes to zero. Our numerical results show that $E\rho^2$ decreases slowly.

As in the case of independent Wiener processes, we can use a Legendre series expansion to approximate the density function of ρ. For $r = 1$, the result is shown in the first panel of Figure 2, from which we see that the 12-th order approximation is very accurate.

4.3. Brownian bridge

For a more complicated example, consider a standard Brownian bridge (denoted by "$\mathbb{B}$") which satisfies $X(0) = X(1) = 1$. In this case, we must fix $T = 1$ and let $\tau = 1 - t$. The dynamics of $X(t)$ can be descr

ibed by (see for example [23] Theorem IV.40.3)

$$dX_t = -\frac{X_t}{1-t}dt + dW_t.$$

Though this SDE has the linear form, the drift coefficient $-(1-t)^{-1}$ explodes at $t = 1$. Hence, it does not satisfy the conditions required in Theorem 2.3. However, the singularity can easily be isolated, by freezing everything at $t = 1 - \varepsilon$ and applying Theorem 2.3 to that; we can then let $\varepsilon \downarrow 0$ and we find the instances of the ODEs (14)-(16) to be

$$\dot{V} - 2V/(1-t) - V^2 + \theta^2 = 0,$$

$$\dot{b} - [V + (1-t)^{-1}]b + z = 0,$$

$$2\dot{\gamma} - b^2 + V = 0.$$

Solving the first differential equation with $\lim_{t\to 1} V(t) = 0$ yields

$$V(t) = \frac{\theta\tau \cosh \theta\tau - \sinh \theta\tau}{\tau \sinh \theta\tau}.$$

One can check that $\lim_{t\to 1} \dot{V}(t) = -\theta^2/3$. Similarly, the solution to the second ODE is given by

$$b(t) = \frac{z(\cosh \theta\tau - 1)}{\theta \sinh \theta\tau}.$$

Though at first sight this might appear to have a singularity at $\tau = 0$ it is in fact analytic. The solution to the third differential equation is given by

$$2\gamma(t) = \frac{z^2}{\theta^2}\left(\frac{2(\cosh \theta\tau - 1)}{\theta \sinh \theta\tau} - \tau\right) + \log \frac{\sinh \theta\tau}{\theta\tau}.$$

One can also check that $\lim_{t\to 1} \gamma(t) = \lim_{t\to 1} \dot{\gamma}(t) = 0$. Using this, we have

$$F_{\mathbb{B}}(t, x; \theta^2, z) = \exp\left\{-\frac{1}{2}V(t)x^2 - b(t)x - \gamma(t)\right\}.$$

Hence

$$F_{\mathbb{B}}(0, 0; \theta^2, z) = \exp\left\{-\gamma(0)\right\}$$

$$= \sqrt{\frac{\theta}{\sinh \theta}} \exp\left\{-\frac{z^2}{2\theta^2}\left(\frac{2(\cosh \theta - 1)}{\theta \sinh \theta} - 1\right)\right\}.$$

Mixing over $z \sim N(0, \theta^2)$ gives the one-dimensional generating function

$$\psi_{\mathbb{B}}(\theta^2) = \frac{\theta}{2 \sinh(\theta/2)},$$

which matches the formula in the second display on p.577 of [7]. As in the case case of Ornstein-Uhlenbeck processes, the moment generating function is $\phi_{\mathbb{B}}(S) = \psi_{\mathbb{B}}(\theta_1^2)\psi_{\mathbb{B}}(\theta_2^2)$.

In Table 3 we provide the moments of ρ for independent Brownian bridges. Comparing with Table 1, we can see that ρ has smaller variance for two Brownian bridges. Intuitively, this is because Brownian

Table 3. Numerical values of the moments of the empirical correlation coefficient for two independent Brownian bridges.

k	2	4	6	8	10
$E\rho^k$	0.149001	0.047864	0.0201829	0.009876	0.005321

bridges are forced to fluctuate around zero more frequently than Brownian motions: a Brownian bridge has to return to zero at $t = 1$ but a Brownian motion is likely to make long excursions away from zero. The second panel of Figure 2 shows the 12-th order polynomial approximation of the density function of ρ.

4.4. Correlated Brownian motion

Let $X_1(t)$, $X_2(t)$ be two Brownian motions with constant correlation c, represented by the following SDE

$$dX_1(t) = dW_1(t), \qquad dX_2(t) = c\, dW_1(t) + \sqrt{1 - c^2}\, dW_2(t).$$

To compute the moment generating function $\phi(S)$, we take the approach outlined in Remark 4. Define a matrix M as

$$M = M(c) = \begin{bmatrix} 1 & 0 \\ -c(1 - c^2)^{-1/2} & (1 - c^2)^{-1/2} \end{bmatrix}.$$

Then the process $MX(t)$ is a two-dimensional Brownian motion with independent coordinates. The inverse of M is

$$M^{-1} = M^{-1}(c) = \begin{bmatrix} 1 & 0 \\ c & \sqrt{1 - c^2} \end{bmatrix}.$$

We now transform the problem to the uncorrelated case by

$$\phi_{\text{cBm}}(S) = \phi_{\text{Bm}}((M^{-1})^\top S M^{-1}),$$

where we use "cBm" to indicate that X is a correlated two-dimensional Brownian motion. The solution may be expressed as

$$\phi_{\text{cBm}}(S; c) = \left(\frac{\lambda_1 \lambda_2}{\sinh \lambda_1 \sinh \lambda_2} \right)^{1/2}, \tag{34}$$

where λ_1^2, λ_2^2 are the eigenvalues of the matrix $(M^{-1})^\top S M^{-1}$. Straightforward calculations yield

$$\lambda_i^2 = \frac{1}{2} \left\{ s_{11} + s_{22} + 2c s_{12} \pm \sqrt{(s_{11} - s_{22})^2 + 4(c s_{11} + s_{12})(c s_{22} + s_{12})} \right\}.$$

In Table 4 we give the first and second moments of ρ for two-dimensional correlated Brownian motion with correlation coefficient c. Observe that $E(\rho)$ is always slightly smaller than c if $c \in (0, 1)$. The variance of ρ, computed as $\text{Var}(\rho) = E\rho^2 - (E\rho)^2$, is decreasing (as c increases) but very slowly. Indeed, the standard deviation of ρ is 0.49 for $c = 0$, 0.41 for $c = 0.5$ and 0.25 for $c = 0.8$. In Table 5 we give high-order moments of ρ for $c = 0.5$. It is somewhat surprising that $E(\rho^k)$ remains close to 0.1 even when $k = 10$.

Table 4. Numerical values of the moments of the empirical correlation coefficient for two correlated Brownian motions with correlation coefficient c ($T = 1$).

c	0	0.1	0.2	0.3	0.4
$E\rho$	0	0.08873	0.17792	0.26804	0.35963
$E\rho^2$	0.24052	0.24550	0.26061	0.28636	0.32368
$\text{Var}(\rho)$	0.2405	0.2376	0.2290	0.2145	0.1943

c	0.5	0.6	0.7	0.8	0.9
$E\rho$	0.45338	0.55004	0.65071	0.75698	0.87151
$E\rho^2$	0.37407	0.43986	0.52477	0.63509	0.78298
$\text{Var}(\rho)$	0.1685	0.1373	0.1013	0.0621	0.0235

Table 5. Numerical values of higher moments of the empirical correlation coefficient for two correlated Brownian motions with correlation coefficient $c = 0.5$ and $T = 1$.

k	1	2	3	4	5
$E\rho^k$	0.4534	0.3741	0.2603	0.2221	0.17137

k	6	7	8	9	10
$E\rho^k$	0.1515	0.1253	0.1109	0.0948	0.0848

Approximating the density function of ρ is more challenging than in the previous three examples due to the slow decay of high-order moments of ρ and the asymmetry of the density function of ρ. In the last panel of Figure 2, we plot the 10-th order approximation for $c = 0.5$, which agrees with the empirical Monte Carlo estimates well, although there appears to be some slight difference near the mode; the difference may be caused by the slow convergence of $\hat{f}_k$ and/or the time discretization scheme used in simulation.

5. Asymptotics of $\rho(T)$ as $T \to \infty$

We now extend the definition of the empirical correlation coefficient to the time interval $[0, T]$ for any $T > 0$. In this context, the empirical correlation coefficient may be written as

$$\rho(T) := \frac{Y_{12}(T)}{\sqrt{Y_{11}(T)Y_{22}(T)}}, \tag{35}$$

where the random variables $Y_{ij}(T)$ ($i, j = 1, 2$) are defined as

$$Y_{ij}(T) := \int_0^T X_i(u)X_j(u)\,du - T\bar{X}_i\bar{X}_j, \quad \bar{X}_i := \frac{1}{T}\int_0^T X_i(u)\,du. \tag{36}$$

The random variable $\bar{X}_i$ is the time average of the process X_i.

The fundamental reason that the statistic $\rho(T)$ has been called "nonsense correlation" is because, in the case of two independent Wiener processes, its asymptotic distribution is heavily dispersed and frequently large in absolute value, leading to high variance (.240522). Further, its asymptotic distribution is very different than that of the nominal t-distribution. This begs the following question: might $\rho(T)$ be useful for testing the independence of some *other pair* of Gaussian processes? In fact, the answer is *yes*; $\rho(T)$ may be used to test independence of two *Ornstein-Uhlenbeck processes*. We shall prove this claim by first showing a Strong Law result (Theorem 5.1), that for two independent Ornstein-Uhlenbeck

processes, $\rho(T)$ converges almost surely to 0 as $T \to \infty$. We next prove a Central Limit result (Theorem 5.2), that $\sqrt{T}\rho(T)$ converges in distribution as $T \to \infty$ to a zero-mean Gaussian with variance that shrinks to zero as the mean reversion parameter tends towards infinity. Of course the Strong Law result Theorem 5.1 is not needed to prove the Central Limit result Theorem 5.2, but as the proof is simple we record it.

Theorem 5.1. *For two independent Ornstein-Uhlenbeck processes, $X_1(t)$ and $X_2(t)$, which both follow the SDE* (33) *with $r > 0$, $\rho(T)$ converges almost surely to zero as $T \to \infty$.*

Proof. If $X_1(0)$ and $X_2(0)$ are both distributed according to the invariant $N(0, 1/2r)$ distribution of the OU process (33), then the bivariate process (X_1, X_2) is ergodic, so, by Birkhoff's Ergodic Theorem, time-averages converge almost surely to expectations. Thus (recalling (35) and (36)) we have

$$\bar{X}_i \to E[X_i(0)] \qquad = 0 \quad \text{a.s. as } T \to \infty, i = 1, 2;$$

$$T^{-1}Y_{12}(T) \to E[X_1(0)X_2(0)] = 0 \quad \text{a.s. as } T \to \infty;$$

$$T^{-1}Y_{ii}(T) \to E[X_i(0)^2] \qquad = (2r)^{-1} \quad \text{a.s. as } T \to \infty, i = 1, 2.$$

Dividing the numerator and denominator of $\rho(T)$ defined at (35) by T, it is immediate that $\rho(T)$ converges almost surely to 0 if the initial distribution is the invariant distribution.

If the initial distribution is something else, then we still have these results by coupling with an independent stationary copy of the OU process - see [23] Theorem V.54.5, which proves that the two diffusions couple in finite time almost surely, so that the long-time averages have the same limits. $\square$

We now prove a central limit theorem for $\rho(T)$ as $T \to \infty$. Two years after the first version of the present paper appeared on arXiv ([11]), a paper by Douissi et al. ([9]) would study in detail the asymptotics of Yule's nonsense correlation for two independent Ornstein-Uhlenbeck processes.

Theorem 5.2. *For two independent Ornstein-Uhlenbeck processes, $X_1(t)$ and $X_2(t)$, which both follow the SDE* (33) *with $r > 0$, we have that*

$$\sqrt{T}\rho(T) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{2r}\right).$$

Proof. Firstly, as we proved in the previous result, we have

$$\frac{Y_{11}(T)}{T} = \frac{1}{T}\int_0^T X_1(s)^2 ds - \bar{X}_1(T)^2$$

$$\xrightarrow{a.s.} E[X_1(0)^2] = \frac{1}{2r}.$$

We now need to obtain weak convergence of

$$\frac{Y_{12}(T)}{\sqrt{T}} = \int_0^T X_1(s)X_2(s)\frac{ds}{\sqrt{T}} - \frac{T\bar{X}_1(T)\bar{X}_2(T)}{\sqrt{T}}. \tag{37}$$

Let us first consider the second term of the right-hand side of the above equation. For simplicity, assume $X_1(0) = 0$ and then

$$X_1(t) = e^{-rt} \int_0^t e^{rs} dW_s,$$

so that

$$T\bar{X}_1(T) = \int_0^T X_1(t)dt = \int_0^T e^{-rt} \int_0^t e^{rs} dW_s \, dt$$

$$= r^{-1} \int_0^T e^{rs} \left(-e^{-rT} + e^{-rs} \right) dW_s$$

$$= r^{-1} \int_0^T \left(1 - e^{-r(T-s)} \right) dW_s.$$

Hence

$$E[\bar{X}_1(T)^2] = \frac{1}{r^2 T^2} \int_0^T \left(1 - e^{-r(T-s)} \right)^2 ds \le \frac{1}{r^2 T},$$

and so

$$E\left[\left(\sqrt{T} \bar{X}_1(T) \bar{X}_2(T) \right)^2 \right] \le \frac{1}{r^4 T} \to 0.$$

Thus $\sqrt{T}\bar{X}_1(T)\bar{X}_2(T)$ converges in L^2 to 0, and so we need now only consider the first term of the right-hand side of equation (37). For $\theta \in \mathbb{R}$, let us evaluate the characteristic function by firstly conditioning on X_2:

$$E \exp\left\{ \frac{i\theta}{\sqrt{T}} \int_0^T X_1(s) X_2(s) \, ds \right\}$$

$$= E \exp\left\{ \frac{i\theta}{\sqrt{T}} \int_0^T e^{-rs} X_2(s) \int_0^s e^{ru} dW_1(u) \, ds \right\}$$

$$= E \exp\left\{ \frac{i\theta}{\sqrt{T}} \int_0^T e^{ru} \int_u^T e^{-rs} X_2(s) \, ds \, dW_1(u) \right\}$$

$$= E \exp\left\{ -\frac{\theta^2}{2T} \int_0^T \left(\int_u^T e^{-r(s-u)} X_2(s) ds \right)^2 du \right\}.$$

Again by the ergodic theorem, we have

$$\frac{1}{T} \int_0^T \left(\int_u^T e^{-r(s-u)} X_2(s) ds \right)^2 du \xrightarrow{a.s.} E\left[\left(\int_0^\infty e^{-rs} X_2(s) ds \right)^2 \right] = \frac{1}{8r^3}.$$

We thus obtain

$$\frac{1}{\sqrt{T}} \int_0^T X_1(s) X_2(s) ds \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{8r^3} \right),$$

from which the stated result follows. $\qquad \square$

Acknowledgments

Funding

References

[1] ALDRICH, J. (1995). Correlations genuine and spurious in Pearson and Yule. *Statist. Sci.* **10** 364-376.

[2] BILLINGSLEY, P. (2008). *Probability and Measure*. John Wiley & Sons.

[3] BURDEN, R. L. and FAIRES, J. D. (1997). Numerical Analysis. *Cole, Belmont*.

[4] CHAN, T. (1991). Indefinite quadratic functionals of Gaussian processes and least-action paths. *Ann. Inst. Henri Poincaré Probab. Stat.* **27** 239–271.

[5] CHAN, T., DEAN, D. S., JANSONS, K. M. and ROGERS, L. C. G. (1994). On polymer conformations in elongational flows. *Comm. Math. Phys.* **160** 239–257.

[6] CRESSIE, N., DAVIS, A. S. and FOLKS, J. L. (1981). The moment-generating function and negative integer moments. *Amer. Statist.* **35** 148–150.

[7] DONATI-MARTIN, C. and YOR, M. (1993). On some examples of quadratic functionals of Brownian motion. *Adv. in Appl. Probab.* **25** 570–584.

[8] DONATI-MARTIN, C. and YOR, M. (1997). Some Brownian functionals and their laws. *Ann. Probab.* **25** 1011–1058.

[9] DOUISSI, S., ES-SEBAIY, K. and VIENS, F. G. (2022). Asymptotics of Yule's nonsense correlation for Ornstein-Uhlenbeck paths: A Wiener chaos approach. *Electron. J. Stat.* **16** 3176–3211.

[10] DYNKIN, E. B. (1980). Markov processes and random fields. *Bull. Amer. Math. Soc.* **3** 975–999.

[11] ERNST, P. A., ROGERS, L. C. G. and ZHOU, Q. (2019). The distribution of Yule's "nonsense correlation". *arXiv preprint arXiv:1909.02546*.

[12] ERNST, P. A., SHEPP, L. A. and WYNER, A. J. (2017). Yule's "nonsense correlation" solved! *Ann. Statist.* **45** 1789–1809.

[13] FIXMAN, M. (1962). Radius of gyration of polymer chains. *J. Chem. Phys.* **36** 306–310.

[14] GRANGER, C. and NEWBOLD, D. (1974). Spurious regression in econometrics. *J. Econometrics* **2** 111-120.

[15] HENDRY, D. F. (1986). Economic modelling with cointegrated variables: an overview. *Oxford Bull. Econom. Statist.* **48** 201-212.

[16] JONES, M. C. (1987). Inverse factorial moments. *Statist. Probab. Lett.* **6** 37–42.

[17] LÉVY, P. (1951). Wiener's random function, and other Laplacian random functions. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. The Regents of the University of California.

[18] MAC AONGHUSA, P. and PULE, J. V. (1989). An extension of Lévy's stochastic area formula. *Stochastics* **26** 247–255.

[19] PHILLIPS, P. C. B. (1986). Understanding spurious regressions in econometrics. *J. Econometrics* **33** 311-340.

[20] PHILLIPS, P. C. B. (1998). New tools for understanding spurious regressions. *Econometrica* **66** 1299–1325.

[21] PROVOST, S. B. (2005). Moment-based density approximants. *Math. J.* **9** 727–756.

[22] ROGERS, L. C. G. and SHI, Z. (1992). Quadratic functionals of Brownian motion, optimal control, and the "Colditz" example. *Stochastics* **41** 201–218.

[23] ROGERS, L. C. G. and WILLIAMS, D. (2000). *Diffusions, Markov Processes and Martingales: Volume 2, Itô Calculus*. Cambridge University Press.

[24] SAWA, T. (1972). Finite-sample properties of the k-class estimators. *Econometrica* **40** 653.

[25] SAXENA, R. (1967). Expansion of continuous differentiable functions in Fourier Legendre series. *Canad. J. Math.* **19** 823–827.

[26] SUETIN, P. K. (1964). On the representation of continuous and differentiable functions by Fourier series in Legendre polynomials. In *Dokl. Akad. Nauk* **158** 1275–1277. Russian Academy of Sciences.

[27] WANG, H. (2021). How much faster does the best polynomial approximation converge than Legendre projection? *Numer. Math.* **147** 481–503.

[28] WANG, H. and XIANG, S. (2012). On the convergence rates of Legendre approximation. *Math. Comp.* **81** 861–877.

[29] YULE, G. U. (1926). 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** 1–63.