The distribution of Yule's "nonsense correlation"

Philip A. Ernst[*], L.C.G. Rogers[†], and Quan Zhou[‡]

September 29, 2019

Abstract

In 2017, the authors of Ernst et al. [2017] explicitly computed the second moment of Yule's "nonsense correlation," offering the first mathematical explanation of Yule's 1926 empirical finding of nonsense correlation. [Yule, 1926]. The present work closes the final longstanding open question on the distribution of Yule's nonsense correlation

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

by explicitly calculating all moments of ρ (up to order 16) for two *independent* Wiener processes, W_1, W_2. These lead to an approximation to the density of Yule's nonsense correlation, apparently for the first time. We proceed to explicitly compute higher moments of Yule's nonsense correlation when the two independent Wiener processes are replaced by two *correlated* Wiener processes, two independent Ornstein-Uhlenbeck processes, and 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; Wiener processes; Volatile correlation

MSC 2010 Codes: Primary 60J65, 60G15; Secondary 60G05

[*]Department of Statistics, Rice University, Houston, Texas 77005, U.S.A.

[†]Statistical Laboratory, Center for Mathematical Sciences, Cambridge CB3 0WB, U.K.

[‡]Department of Statistics, Texas A&M University, College Station, Texas 77843, U.S.A.

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 offers a solution that is standard yet often problematic, particularly when it is calculated between two time series. Indeed, one may observe "volatile" correlation in independent time series, where the correlation is volatile in the sense that its distribution is heavily dispersed and is frequently large in absolute value. Yule observed this empirically in his 1926 seminal paper [Yule, 1926], calling it "nonsense" correlation, writing that "we sometimes obtain between quantities varying with time (time-variables) quite high correlations to which we cannot attach any physical significance whatever, although under the ordinary test the correlation would be held to be certainly significant."

Yule's finding would remain "isolated" from the literature until 1986 [Aldrich, 1995], when the authors of Hendry [1986] and Phillips [1986] confirmed many of the empirical claims of "spurious regression" made by the authors of Granger and Newbold [1974]. In particular, Phillips [1986] 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 Phillips [1998], the same author provided an explanation of such spurious regressions in terms of orthonormal representations of the Karhunen Loève type.

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 random variables

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

with values in $\mathbb{R}^d$ and $\mathbb{M}^d$ respectively, where $\mathbb{M}^d$ is the space of $d \times d$ real matrices. Y_{ij} is the (i, j)-th entry of the matrix Y. In 2017, Ernst et al. [2017] explicitly calculated the second moment of Yule's nonsense correlation, defined as

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

in the case $d = 2$, when X is a two-dimensional Wiener process[1]. This calculation provided the first mathematical explanation of Yule's 1926 empirical finding. The present work closes the final long-standing open question on the distribution of Yule's nonsense correlation by determining the law of ρ and explicitly calculating all moments (up to order 16). With these moments in hand, we provide the first density approximation to Yule's nonsense correlation. We then proceed to explicitly compute higher moments of Yule's nonsense correlation when the two independent Wiener processes are replaced by two *correlated* Wiener processes, two independent Ornstein-Uhlenbeck processes, and two independent Brownian bridges. This closes all previously open problems raised in Section 3.3 of Ernst et al. [2017].

Our method for explicitly calculating all moments (up to order 16) of Yule's nonsense correlation (Section 3), relies 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 [Donati-Martin and Yor, 1997] which presents the state of the subject as it was in 1997, and as it has substantially remained since then. A range of techniques is available to characterize the laws of quadratic functionals of Brownian motion, including:

1. eigenfunction expansions — see, for example, Lévy [1951], Fixman [1962], Mac aonghusa and Pule [1989], Chan [1991], Chan et al. [1994], Ernst et al. [2017];

2. identifying the covariance of the Gaussian process as the Green function of a symmetrizable Markov process — see, for example, Chan et al. [1994], Dynkin [1980];

3. stochastic Fubini relations — see, for example, Donati-Martin and Yor [1997];

4. Itô's formula — see Rogers and Shi [1992].

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

[1]Scaling properties of Brownian motion show that the law of ρ does not depend on the choice of $T > 0$.

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, not so 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 this paper. It has the advantage that no clever insight is required — it is a mechanical calculation, as we shall see when we extend the Brownian motion result to correlated Brownian motions, Ornstein-Uhlenbeck processes and the Brownian bridge; we make the obvious changes to the ODEs to be solved and that is all there is.

The final part of the paper studies the asymptotic properties of ρ as $T \to \infty$. 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 4, 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.

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 main result is the following.

Theorem 1. *Suppose that $\sigma : [0, T] \mapsto \mathbb{M}^d$ is a bounded measurable function, and that X solves*[2]

$$dX = \sigma dW, \tag{3}$$

where W is d-dimensional Brownian motion. We write $\Sigma = \sigma\sigma^\top$.

Suppose that $Q : [0, T] \to \mathbb{S}^d_+$ and $z : [0, T] \to \mathbb{R}^d$ are bounded measurable

[2] For notational simplicity, we will often omit the independent variable t.

functions such that Q^{-1} is also bounded. Define

$$\ell := \frac{1}{2} X \cdot QX + z \cdot X, \tag{4}$$

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

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),[3]

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

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

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

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.

[3]We use an "overdot" to denote the derivative with respect to t.

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

$$\begin{aligned} dZ_t &= (V_t X_t + b_t, \sigma_t \, dW_t \, dt) + \tfrac{1}{2} \mathrm{tr}\,(V_t \Sigma_t) dt + \left\{ \tfrac{1}{2} X_t \cdot \dot{V}_t X_t + \dot{b}_t X_t + \dot{\gamma}_t \right\} dt, \\ d\langle Z \rangle_t &= (V_t X_t + b_t) \cdot \Sigma_t (V_t X_t + b_t) dt. \end{aligned}$$

Now consider the process

$$M_t = \exp\biggl(-\tfrac{1}{2} \int_0^t \ell(s) \; ds - Z_t \biggr). \tag{11}$$

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

$$\begin{aligned} \frac{dM_t}{M_t} &= -dZ + \tfrac{1}{2} d\langle Z \rangle - \tfrac{1}{2} X \cdot QX dt - z \cdot X dt \\ &\doteq \Big\{ -\tfrac{1}{2} \mathrm{tr}\,(V\Sigma) - \tfrac{1}{2} X \cdot \dot{V} X - \dot{b} X - \dot{\gamma} + \\ &\qquad + \tfrac{1}{2} (VX + b) \cdot \Sigma (VX + b) - \tfrac{1}{2} X \cdot QX - z \cdot X \Big\} \, dt \\ &= 0 \end{aligned}$$

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

$\square$

Theorem 1 extends easily to the situation where X is the solution of a linear SDE.

Theorem 2. *Suppose that σ, $B : [0, T] \mapsto \mathbb{M}^d$ and $\delta :\mapsto \mathbb{R}^d$ are bounded measurable functions, and that X solves*

$$dX = \sigma \, dW + (BX + \delta) \, dt, \tag{12}$$

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 ordinary differential equations (ODEs),[4]

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

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

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

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, Rogers and Williams [2000] 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

$$d\tilde{X} = A\sigma\, dW$$

and Theorem 1 applies. The equations (14), (15) and (16) are easily checked to be the analogues of (7), (8) and (9) respectively.

□

Remark 1. We will want to apply Theorem 1 to situations where $Q(T) = 0$. This is a simple limiting case of the problem where we take $Q(T) = \varepsilon I$ and

[4] We use an "overdot" to denote the derivative with respect to t.

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 2. Theorem 2 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 3. If σ is constant, we may assume that $\sigma = I$, the identity matrix. To see this, let $\hat{X} = \Sigma^{-1/2} X$, and note that the diffusion process $\hat{X}$ solves the linear SDE,

$$d\hat{X} = (\Sigma^{-1/2} B \Sigma^{1/2} \hat{X} + \Sigma^{-1/2} \delta) dt + dW.$$

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

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

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 4. Although Theorem 2 deals with the general case where Q, z are measurable functions, in the remainder of this paper we only need invoke Theorem 2 for the special case in which Q and z are constants. For this reason, we will sometimes use the alternative expanded notation

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

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 (10) 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

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

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 $\rho = \rho(T)$ 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 [Billingsley, 2008]. 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). $\square$

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 (22). This is where Theorem 2 comes in. If X is a solution of a linear SDE (12), starting at $X_0 = 0$ to fix the discussion, and we set

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

then Theorem 2 tells us how to compute

$$\begin{aligned} F(0,0;a) &= E\Big[\exp\Big\{ -\int_0^T \{\tfrac{1}{2}X(u)\cdot SX(u) + a\cdot X(u)\}\, du \Big\}\Big] & (24)\\ &= \exp(-\gamma(0;a)), & (25) \end{aligned}$$

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

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

The strategy now should be clear. In any particular application, we use Theorem 2 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

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

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 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 $\theta \geq 0$ and $z \in \mathbb{R}$. By Theorem 2, the solution has the following form (the subscript "Bm" is Brownian motion)

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

which leads to the following system of ordinary differential equations

$$\begin{aligned} \dot{V} - V^2 + \theta^2 &= 0, \\ \dot{b} - Vb + z &= 0, \\ 2\dot{\gamma} - b^2 + V &= 0. \end{aligned}$$

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

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

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

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

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

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

where s_{ij} is the (i,j)-th entry of S.

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

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

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 et al. [2017, 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. For example, we may consider a polynomial approximation

$$\hat{f}(\rho) = a_0 + a_1 \rho + a_2 \rho^2 + \cdots + a_k \rho^k.$$

k	2	4	6	8
$E\rho^k$	0.240523	0.109177	0.060862	0.037788
k	10	12	14	16
$E\rho^k$	0.025114	0.017504	0.012641	0.009385

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

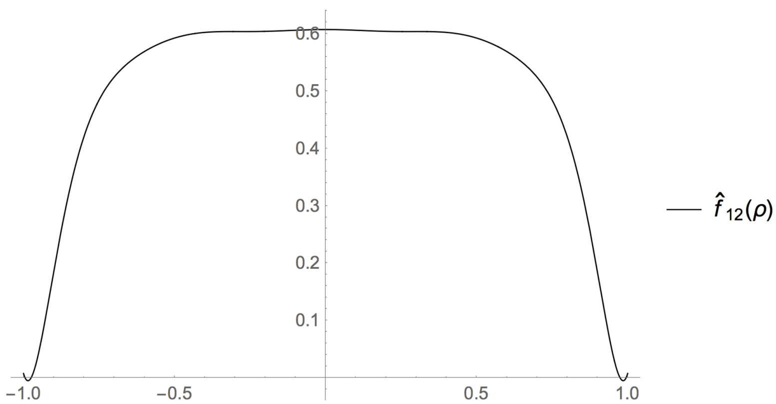

Figure 1: The 12th-order polynomial approximation to the probability density function of ρ for two independent Wiener processes.

The coefficients $(a_0, a_1, \dots, a_k)$ 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 polynomial approximation to the density function [Provost, 2005]. The 4th-order, 6th-order and 8th-order polynomial approximations are provided below.

$$\hat{f}_4(\rho) = 0.59081 + 0.31001\rho^2 - 0.97075\rho^4,$$
$$\hat{f}_6(\rho) = 0.60057 + 0.10518\rho^2 - 0.35627\rho^4 - 0.45062\rho^6,$$
$$\hat{f}_8(\rho) = 0.61200 - 0.30638\rho^2 + 1.9073\rho^4 - 4.3742\rho^6 + 2.1019\rho^8.$$

The 12th-order approximation is drawn in Figure 1 above, which looks almost the same as the 8th-order one (not shown here). The above expressions constitute *the first density approximations* to Yule's nonsense correlation. It can be seen that the distribution of ρ is dispersed: 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 order 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 become extremely slow and require massive memory.

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:

$$dX(t) = -rX(t)dt + dW(t), \qquad r \in (0, \infty). \tag{32}$$

By Theorem 2, the solution has the form

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

which can be obtained by solving the following system of ODEs

$$\begin{aligned} \dot{V} - 2rV - V^2 + \theta^2 &= 0, \\ \dot{b} - (V + r)b + z &= 0, \\ 2\dot{\gamma} - b^2 + V &= 0. \end{aligned}$$

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

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

where $\eta = \sqrt{r^2 + \theta^2}$. The second differential equation is first-order linear, so can be solved explicitly; after some routine 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

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

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

Table 2: Numerical values of the second moment of Yule's nonsense correlation for two independent Ornstein-Uhlenbeck processes with mean reversion parameter r ($T = 1$).

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

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

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

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

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.

4.3 Brownian bridge.

For a more complicated example, consider a standard Brownian bridge (denoted by "Bb") 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 described by (see for example Rogers and Williams [2000] 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. However, the singularity can easily be isolated, by freezing everything at $t = 1-\varepsilon$ and applying Theorem 2 to that; we can then let $\varepsilon \downarrow 0$ and we find the instances of the ODEs (14)-(16) to be

$$\begin{aligned}
\dot{V} - 2V/(1-t) - V^2 + \theta^2 = 0, \\
\dot{b} - [V + (1-t)^{-1}]b + z = 0, \\
2\dot{\gamma} - b^2 + V = 0.
\end{aligned}$$

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

$$F_{\mathrm{Bb}}(t, x;\, \theta^2, z) = \exp\left\{-\frac{1}{2}Vx^2 - bx - \gamma\right\}.$$

Hence

$$\begin{aligned}
F_{\mathrm{Bb}}(0, 0;\, \theta^2, z) &= \exp\{-\gamma(0)\} \\
&= \sqrt{\frac{\theta}{\sinh\theta}}\exp\left\{-\frac{z^2}{2\theta^2}\left(\frac{2(\cosh\theta - 1)}{\theta\sinh\theta} - 1\right)\right\}
\end{aligned}$$

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

$$\psi_{\mathrm{Bb}}(\theta^2) = \frac{\theta}{2\sinh(\theta/2)}.$$

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

Table 3: Numerical values of the moments of Yule's nonsense correlation for two independent Brownian bridges.

As in the case case of Ornstein-Uhlenbeck processes, the moment generating function is $\phi_{\mathrm{Bb}}(S) = \psi_{\mathrm{Bb}}(\theta_1^2)\psi_{\mathrm{Bb}}(\theta_2^2)$.

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 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.

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 3. 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

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

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
$\mathrm{Var}(\rho)$	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
$\mathrm{Var}(\rho)$	0.1685	0.1373	0.1013	0.0621	0.0235

Table 4: Numerical values of the moments of Yule's nonsense correlation for two correlated Brownian motions with correlation coefficient c ($T = 1$).

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

$$\phi_{\mathrm{cBm}}(S; c) = \left(\frac{\lambda_1 \lambda_2}{\sinh \lambda_1 \sinh \lambda_2} \right)^{1/2}, \tag{33}$$

where λ_1^2, λ_2^2 are the eigenvalues of the matrix $(M^{-1})^\top S M^{-1}$. Routine calculation yields

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

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 $\mathrm{Var}(\rho) = E\rho^2 - (E\rho)^2$, 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 $c = 0.8$.

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

The fundamental reason that the statistic $\rho(T)$ has been called "nonsense correlation" is that, in the case of two independent Wiener processes, its asymptotic variance is quite large, rendering it useless for statistical inference. This being said, might $\rho(T)$ be *useful* for testing the independence of some other pair of stochastic processes? In fact, the answer is *yes*; $\rho(T)$ may be used to test independence of two *Ornstein-Uhlenbeck processes.* We prove this claim by first showing a Strong Law result, 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,that $\sqrt{T}\rho(T)$ converges in distribution as $T \to \infty$ to a zero-mean Gaussian[5].

Theorem 3. *For two independent Ornstein-Uhlenbeck processes, $X_1(t)$ and $X_2(t)$, which both follow the SDE* (32) *with $r > 0$, the "nonsense" correlation statistic $\rho(T)$ 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 (32), then the bivariate process (X_1, X_2) is ergodic, so, by Birkhoff's Ergodic Theorem, time-averages converge almost surely to expectations. Thus (recall the notations (**??**) and (**??**)) we have

$$\begin{aligned}
\bar{X}_i &\to E[X_i(0)] &&= 0 \quad \text{a.s. as } T \to \infty,\ i = 1,2 \\
T^{-1}Y_{12}(T) &\to E[X_1(0)X_2(0)] &&= 0 \quad \text{a.s. as } T \to \infty \\
T^{-1}Y_{ii}(T) &\to E[X_i(0)^2] &&= (2r)^{-1} \quad \text{a.s. as } T \to \infty,\ i = 1,2
\end{aligned}$$

Dividing the numerator and denominator of $\rho(T)$ defined at (**??**) 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 Rogers and Williams [2000] 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.

□

We now prove a central limit theorem for $\rho(T)$ as $T \to \infty$.

Theorem 4. *For two independent Ornstein-Uhlenbeck processes, $X_1(t)$ and $X_2(t)$, which both follow the SDE* (32) *with $r > 0$, we have that*

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

[5]Of course, the Strong Law result Theorem 3 is not needed to prove the Central Limit result Theorem 4, but as the proof is very simple we record it.

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

$$\begin{aligned}\frac{Y_{11}(T)}{T} &= \frac{1}{T}\int_0^T X_1(s)^2 ds - \bar{X}_1(T)^2 \\ &\xrightarrow{a.s.} E[X_1(0)^2] = \frac{1}{2r}.\end{aligned}$$

We now need to obtain weak convergence of

$$\frac{Y_{12}(T)}{\sqrt{T}} = \int_0^T X_1(s)X_2(s)\frac{ds}{\sqrt{T}} - \frac{T\bar{X}_1(T)\bar{X}_2(T)}{\sqrt{T}}. \tag{34}$$

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

$$\begin{aligned}T\bar{X}_1(T) = \int_0^T X_1(t)dt &= \int_0^T e^{-rt}\int_0^t e^{rs} dW_s\,dt \\ &= r^{-1}\int_0^T e^{rs}\left(-e^{-rT} + e^{-rs}\right) dW_s \\ &= r^{-1}\int_0^T \left(1 - e^{-r(T-s)}\right) dW_s.\end{aligned}$$

Hence

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

and so

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

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 (34). For $\theta \in \mathbb{R}$, let us evaluate the characteristic function by firstly conditioning on X_2:

$$
\begin{aligned}
&E\exp\left\{ \frac{i\theta}{\sqrt{T}} \int_0^T X_1(s)X_2(s)\,ds \right\} \\
&= E\exp\left\{ \frac{i\theta}{\sqrt{T}} \int_0^T e^{-rs} X_2(s) \int_0^s e^{ru} dW_1(u)\,ds \right\} \\
&= E\exp\left\{ \frac{i\theta}{\sqrt{T}} \int_0^T e^{ru} \int_u^T e^{-rs} X_2(s)\,ds\,dW_1(u) \right\} \\
&= E\exp\left\{ -\frac{\theta^2}{2T} \int_0^T \left(\int_u^T e^{-r(s-u)} X_2(s) ds \right)^2 du \right\}.
\end{aligned}
$$

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. $\square$

Acknowledgments Philip Ernst is very grateful to Professor Frederi Viens for many illuminating conversations about this work. Philip Ernst acknowledges support from Army Research Office Young Investigator Program grant ARO-YIP-71636-MA, National Science Foundation (NSF) grant DMS-1811936, and Office of Naval Research (ONR) grant N00014-18-1-2192.

References

J. Aldrich. Correlations genuine and spurious in Pearson and Yule. *Statistical Science*, 10(4): 364–376, 1995.

P. Billingsley. *Probability and Measure*. John Wiley & Sons, 2008.

T. Chan. Indefinite quadratic functionals of gaussian processes and least-action paths. *Annales de l'IHP Probabilités et Statistiques*, 27(2):239–271, 1991.

T. Chan, D. S. Dean, K. M. Jansons, and L. C. G. Rogers. On polymer conformations in elongational flows. *Communications in Mathematical Physics*, 160(2):239–257, 1994.

C. Donati-Martin and M. Yor. Some Brownian functionals and their laws. *The Annals of Probability*, 25(3):1011–1058, 1997.

E. B. Dynkin. Markov processes and random fields. *Bulletin of the American Mathematical Society*, 3(3):975–999, 1980.

P. A. Ernst, L. A. Shepp, and A. J. Wyner. Yules "nonsense correlation" solved! *The Annals of Statistics*, 45(4):1789–1809, 2017.

M. Fixman. Radius of gyration of polymer chains. *The Journal of Chemical Physics*, 36(2):306–310, 1962.

C. Granger and D. Newbold. Spurious regression in econometrics. *Journal of Econometrics*, 2: 111–120, 1974.

D.F. Hendry. Economic modelling with cointegrated variables: an overview. *Oxford Bulletin of Economics and Statistics*, 48(3): 201–212, 1986.

P. Lévy. 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, 1951.

P. Mac aonghusa and J. V. Pule. An extension of Lévy's stochastic area formula. *Stochastics: An International Journal of Probability and Stochastic Processes*, 26(4):247–255, 1989.

P. C. B. Phillips. Understanding spurious regressions in econometrics. *Journal of Econometrics*, 33(3): 311–340, 1986.

P. C. B. Phillips. New tools for understanding spurious regressions. *Econometrica*, 66(6):1299–1325, 1998.

S. B. Provost. Moment-based density approximants. *Mathematica Journal*, 9(4):727–756, 2005.

L. C. G. Rogers and Z. Shi. Quadratic functionals of Brownian motion, optimal control, and the Colditz example. *Stochastics: An International Journal of Probability and Stochastic Processes*, 41(4):201–218, 1992.

L. C. G. Rogers and David Williams. *Diffusions, Markov Processes and Martingales: Volume 2, Itô Calculus*. Cambridge University Press, 2000.

G. U. Yule. Why do we sometimes get nonsense-correlations between time-series?–a study in sampling and the nature of time-series. *Journal of the Royal Statistical Society*, 89(1):1–63, 1926.