



Brownian Forgery of Statistical Dependences

Vincent Wens^{1,2*}

¹ Laboratoire de Cartographie fonctionnelle du Cerveau, ULB Neurosciences Institute, Université libre de Bruxelles, Brussels, Belgium, ² Magnetoencephalography Unit, Department of Functional Neuroimaging, Service of Nuclear Medicine, CUB – Hôpital Erasme, Brussels, Belgium

The balance held by Brownian motion between temporal regularity and randomness is embodied in a remarkable way by Levy's forgery of continuous functions. Here we describe how this property can be extended to forge arbitrary dependences between two statistical systems, and then establish a new Brownian independence test based on fluctuating random paths. We also argue that this result allows revisiting the theory of Brownian covariance from a physical perspective and opens the possibility of engineering nonlinear correlation measures from more general functional integrals.

Keywords: Brownian distance covariance, Brownian motion, nonlinear correlation, Levy's forgery theorem, statistical independence

OPEN ACCESS

Edited by:

Lixin Shen,
Syracuse University, United States

Reviewed by:

Gabor J. Szekely,
National Science Foundation (NSF),
United States
Xin Guo,
Hong Kong Polytechnic University,
Hong Kong

*Correspondence:

Vincent Wens
vwens@ulb.ac.be

Specialty section:

This article was submitted to
Mathematics of Computation and
Data Science,
a section of the journal
Frontiers in Applied Mathematics and
Statistics

Received: 27 March 2018

Accepted: 17 May 2018

Published: 05 June 2018

Citation:

Wens V (2018) Brownian Forgery of
Statistical Dependences.
Front. Appl. Math. Stat. 4:19.
doi: 10.3389/fams.2018.00019

1. INTRODUCTION

The modern theory of Brownian motion provides an exceptionally successful example of how physical models can have far-reaching consequences beyond their initial field of development. Since its introduction as a model of particle diffusion, Brownian motion has indeed enabled the description of a variety of phenomena in cell biology, neuroscience, engineering, and finance [1]. Its mathematical formulation, based on the Wiener measure, also represents a fundamental prototype of continuous-time stochastic process and serves as powerful tool in probability and statistics [2, 3]. Following a similar vein, we develop in this note a new way of applying Brownian motion to the characterization of statistical independence.

Our connection between Brownian motion and independence is motivated by recent developments in statistics, more specifically the unexpected coincidence of two different-looking dependence measures: distance covariance, which characterizes independence fully thanks to its built-in sensitivity to all possible relationships between two random variables [4], and Brownian covariance, a version of covariance that involves nonlinearities randomly generated by a Brownian process [5]. Their equivalence provides a realization of the aforementioned connection, albeit in a somewhat indirect way that conceals its naturalness. Our goal is to explicit how Brownian motion can reveal statistical independence by relying directly on the geometry of its sample paths.

The *brute force* method to establish the dependence or independence of two real-valued random variables X and Y consists in examining all potential relations between them. More formally, it is sufficient to measure the covariances $\text{cov}[f(X), g(Y)]$ associated with transformations f, g that are bounded and continuous (see, e.g., Theorem 10.1 Jacod and Protter [6]). The question pursued here is whether using sample paths of Brownian motion in place of bounded continuous functions also allows to characterize independence, and we shall demonstrate that the answer is yes. In a nutshell, the statistical fluctuations of Brownian paths B, B' enable the *stochastic* covariance index $\text{cov}[B(X), B'(Y)]$ to probe arbitrary dependences between the random variables X and Y .

Our strategy to realize this idea consists in establishing that, given any pair f, g of bounded continuous functions and any level of accuracy, the covariance $\text{cov}[f(X), g(Y)]$ can be approximated *generically* by $\text{cov}[\mathbf{B}(X), \mathbf{B}'(Y)]$. Crucially, the notion of genericity used here refers to the fact that the probability of picking paths \mathbf{B}, \mathbf{B}' fulfilling this approximation is nonzero, which ensures that an appropriate selection of stochastic covariance can be achieved by finite sampling of Brownian motion. This core result of the paper will be referred to as the *forgery of statistical dependences*, in analogy with Levy's classical forgery theorem [2].

Actually Levy's remarkable theorem, which states that any continuous function can be approximated on a finite interval by generic Brownian paths, provides an obvious starting point of our analysis. Indeed, it stands to reason that if the paths \mathbf{B} and \mathbf{B}' approach the functions f and g , respectively, then $\text{cov}[\mathbf{B}(X), \mathbf{B}'(Y)]$ should approach $\text{cov}[f(X), g(Y)]$ as well. A technical difficulty, however, lies with the restriction to finite intervals since the random variables X and Y may be unbounded. Although it turns out that intervals can not be prolonged as such without ruining genericity, we shall describe first a suitable extension of Levy's forgery that holds on infinite domains. Our forgery of statistical dependences will then follow.

From a practical standpoint, using Brownian motion to establish independence turns out to be advantageous. Indeed, exploring all bounded continuous transformations exhaustively is realistically impossible. (This practical difficulty motivates the use of reproducing kernel Hilbert spaces, see e.g., Gretton and Györfi [7] for a review). Generating all possible realizations of Brownian motion obviously poses the same problem, but this unwieldy task can be bypassed by averaging directly over sample paths. In this way, and quite amazingly, the measurement of an uncountable infinity of covariance indices can be replaced by a *single* functional integral. We shall discuss how this idea leads back to the concept of Brownian covariance and how the forgery of statistical dependences clarifies the way it does characterize independence, without reference to the equivalence with distance covariance. Brownian covariance represents a very promising tool for modern data analysis [8, 9] but appears to be still scarcely used in applications (with seminal exceptions for nonlinear time series [10] or brain connectomics [11]). Our approach based on random paths is both physically grounded and mathematically rigorous, so we believe that it may help further disseminate this method and establish it as a standard tool of statistics.

2. MAIN RESULTS

Here we motivate and describe our main results, with sufficient precision to provide a self-contained presentation of the ideas introduced above while avoiding technical details, which are then developed in the dedicated section 3. We also use here assumptions that are slightly stronger than is necessary, and some generalizations are relegated to the Supplementary Material S1.

2.1. Extension of Levy's Forgery

Imagine recording the movement of a free Brownian particle in a very large number of trials. In essence, Levy's forgery ensures that

one of these traces will follow closely a predefined test trajectory, at least for some time. To formulate this more precisely, let us focus for definiteness on *standard* Brownian motion \mathbf{B} , whose initial value is set to $\mathbf{B}(0) = 0$ and variance at time t is normalized to $\langle \mathbf{B}(t)^2 \rangle = |t|$. We fix a real-valued continuous function f with $f(0) = 0$ (the test trajectory) and consider the *uniform* approximation event that a Brownian path \mathbf{B} fits f tightly up to a constant distance $\delta > 0$ on the time interval $[-T, T]$,

$$\mathcal{U}_{f,\delta,T} = \{|\mathbf{B}(t) - f(t)| < \delta, \forall |t| \leq T\}. \quad (1)$$

Levy's forgery theorem states that this event is generic, i.e., it occurs with probability $\mathbf{P}(\mathcal{U}_{f,\delta,T}) > 0$ (see Chapter 1, Theorem 38 in [2]). This result requires both the randomness and the continuity of Brownian motion. Neither deterministic processes nor white noises satisfy this property.

In all trials though, the particle will eventually drift away to infinity and thus deviate from any bounded test trajectory. Indeed, let us further assume that the function f is bounded and examine what happens when $T \rightarrow \infty$. If the limit event $\mathcal{U}_{f,\delta,\infty} = \bigcap_{T>0} \mathcal{U}_{f,\delta,T}$ occurs, the path \mathbf{B} must be bounded too since the particle is forever trapped in a finite-size neighborhood of the test trajectory. However Brownian motion is almost surely (a.s.) unbounded at long times [2], so that $\mathbf{P}(\mathcal{U}_{f,\delta,\infty}) = 0$. Hence Levy's forgery theorem does not work on infinite time domains.

To accommodate this asymptotic behavior, we should thus allow the particle to diverge from the test trajectory, at least in a controlled way. Let us recall that the escape to infinity is a.s. slower for Brownian motion than for any movement at constant velocity (which is one way to state the law of large numbers [2]). This suggests adjoining to event (1) the loose approximation event

$$\mathcal{E}_{f,v,T} = \{|\mathbf{B}(t) - f(t)| < v|t|, \forall |t| \geq T\} \quad (2)$$

whereby the particle is confined to a neighborhood of the test trajectory that *expands* at finite speed $v > 0$.

Asymptotic forgery theorem. *Let f be bounded and continuous, and $v, T > 0$. Then $\mathbf{P}(\mathcal{E}_{f,v,T}) > 0$.*

An elegant, albeit slightly abstract, proof rests on a short/long time duality between the classes of events (1) and (2), which maps Levy's forgery and this asymptotic version onto each other (see section 3.1). For a more concrete approach, let us focus on the large T limit that will be used to study statistical dependences. Since the path $\mathbf{B}(t)$ and the neighborhood size $v|t|$ both diverge, the bounded term $f(t)$ can be neglected in Equation (2) and the event $\mathcal{E}_{f,v,T}$ thus merely requires not to outrun deterministic particles moving at speed v . The asymptotic forgery thus reduces to the law of large numbers, which ensures that $\mathbf{P}(\mathcal{E}_{f,v,T})$ is close to one for all $T \gg 1$. This probability decreases continuously as T is lowered [since the defining condition in Equation (2) becomes stricter] but does not drop to zero until $T = 0$ is reached. This line of reasoning can be completed and also generalized to allow slower expansions (see Supplementary Material S1).

We now combine Levy's forgery and the asymptotic version to obtain an extension valid at all timescales. Specifically, let us

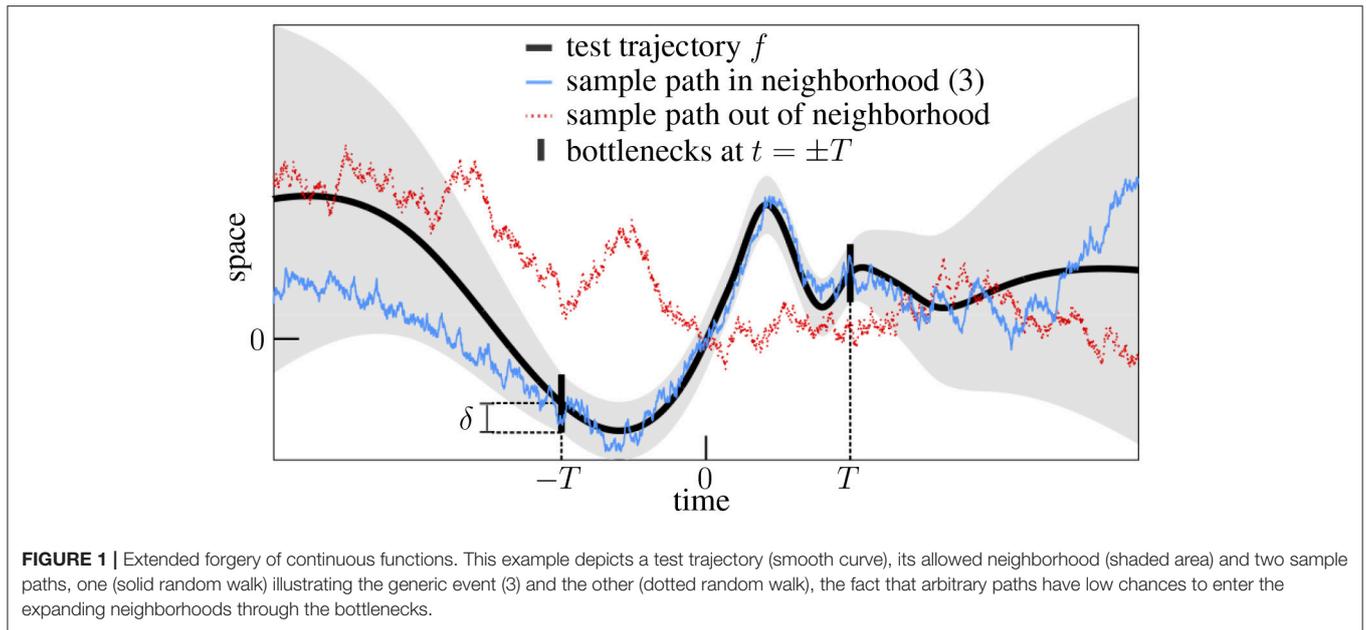


FIGURE 1 | Extended forgery of continuous functions. This example depicts a test trajectory (smooth curve), its allowed neighborhood (shaded area) and two sample paths, one (solid random walk) illustrating the generic event (3) and the other (dotted random walk), the fact that arbitrary paths have low chances to enter the expanding neighborhoods through the bottlenecks.

examine the *joint* approximation event

$$\mathcal{I}_{f,\delta,T} = \mathcal{U}_{f,\delta,T} \cap \mathcal{E}_{f,v,T} \quad \text{with } v = \delta/T. \quad (3)$$

In words, the particle is constrained to follow closely the test trajectory for some time but is allowed afterwards to deviate slowly from it (Figure 1).

Extended forgery theorem. *Let f be bounded and continuous with $f(0) = 0$, and $\delta, T > 0$. Then $\mathbf{P}(\mathcal{I}_{f,\delta,T}) > 0$.*

This result relies on the suitable integration of a “local” version of the theorem (see section 3.2), but it can also be understood rather intuitively as follows. Imagine for a moment that the events (1) and (2) were independent. Their joint probability would merely be equal to the product of their marginal probabilities, which are positive by Levy’s forgery and the asymptotic forgery, and genericity would then follow. Actually they do interact because the associated neighborhoods are connected through the narrow bottlenecks at $t = \pm T$ (Figure 1) but this should only increase their joint probability, i.e.,

$$\mathbf{P}(\mathcal{I}_{f,\delta,T}) > \mathbf{P}(\mathcal{U}_{f,\delta,T}) \mathbf{P}(\mathcal{E}_{f,v,T}). \quad (4)$$

The reason lies in the temporal continuity of Brownian motion. A particle staying in the uniform neighborhood while $|t| \leq T$ necessarily passes through the bottlenecks, and is thus more likely to remain within the expanding neighborhood than arbitrary particles, which have low chances to even meet the bottlenecks (Figure 1). In other words, the proportion $\mathbf{P}(\mathcal{I}_{f,\delta,T})/\mathbf{P}(\mathcal{U}_{f,\delta,T})$ of sample paths $\mathbf{B} \in \mathcal{E}_{f,v,T}$ among all those sample paths $\mathbf{B} \in \mathcal{U}_{f,\delta,T}$ should be larger than the unconstrained probability $\mathbf{P}(\mathcal{E}_{f,v,T})$, hence the bound (4).

2.2. Forgery of Statistical Dependences

We now turn to the analysis of statistical relations using Brownian motion. Let us fix two random variables X, Y and a

pair of bounded test trajectories f, g . Consider the *covariance* approximation event

$$\mathcal{C}_{X,Y,f,g,\varepsilon} = \{ |\text{cov}[\mathbf{B}(X), \mathbf{B}'(Y)] - \text{cov}[f(X), g(Y)]| < \varepsilon \} \quad (5)$$

whereby the stochastic covariance $\text{cov}[\mathbf{B}(X), \mathbf{B}'(Y)]$, built by picking independently two sample paths \mathbf{B}, \mathbf{B}' , coincides with the test covariance $\text{cov}[f(X), g(Y)]$ up to a small error $\varepsilon > 0$ (Figure 2). We argue that this event is generic too.

The first step is to ensure that the set (5) is measurable so that its probability is meaningful. Physically, this technical issue is rooted once again in the escape of Brownian particles to infinity. The stochastic covariance can be expressed as a difference of two averages $\langle \mathbf{B}(X) \mathbf{B}'(Y) \rangle$ and $\langle \mathbf{B}(X) \rangle \langle \mathbf{B}'(Y) \rangle$ (computed at *fixed* sample paths) involving the coordinates $\mathbf{B}(t), \mathbf{B}'(t')$ at random moments $t = X, t' = Y$. If long times and thus large coordinates are sampled too often, the two terms may diverge and lead to an ill-defined covariance, i.e., $\infty - \infty$. To avoid this situation, we should therefore assume that asymptotic values of X and Y are unlikely enough. Actually we shall adopt hereafter the sufficient condition that X, Y are L_2 , i.e., they have finite mean and variance. (See Supplementary Material S2 for a derivation of measurability).

Forgery theorem of statistical dependences. *Let X, Y be L_2 random variables, f, g be bounded and continuous, and $\varepsilon > 0$. Then $\mathbf{P}(\mathcal{C}_{X,Y,f,g,\varepsilon}) > 0$.*

The idea is that one way of realizing the event (5) is to pick sample paths $\mathbf{B} \in \mathcal{I}_{f,\delta,T}, \mathbf{B}' \in \mathcal{I}_{g,\delta,T}$ that fit the test trajectories f, g tightly ($\delta \ll 1$) over a long time period ($T \gg 1$), see Figure 2b. Indeed, as shown below, we have then

$$\text{cov}[\mathbf{B}(X), \mathbf{B}'(Y)] - \text{cov}[f(X), g(Y)] = \mathcal{O}(\delta) + \mathcal{O}(v) \quad (6)$$

with $v = \delta/T$. This rough estimate explains why the event (5) must occur whenever δ and v are small enough [see section 3.3,

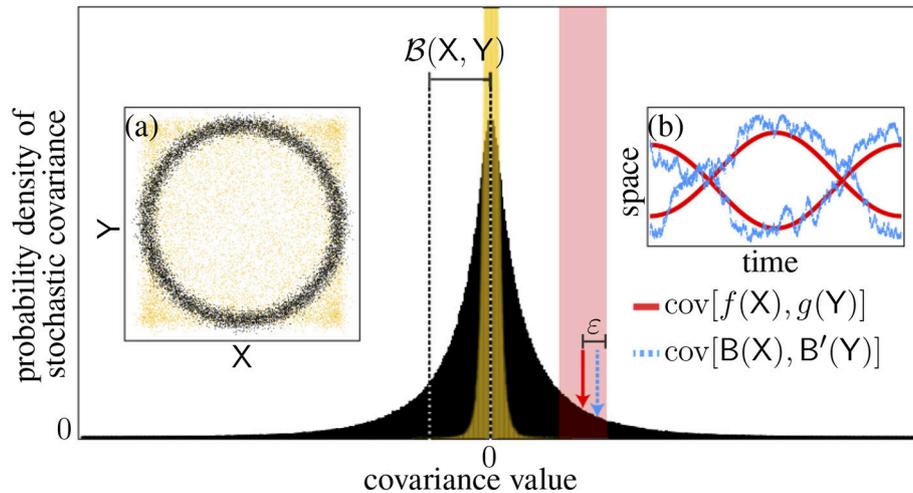


FIGURE 2 | Forgery of statistical dependences. The distribution of the stochastic covariance (black histogram) and its standard deviation $[B(X, Y)]$ are shown for simulated dependent random variables [$n = 10^4$ black dots, Insert (a)] as well as a test covariance (plain arrow) and a sample (dotted arrow) falling within the allowed error (shaded area). Insert (b) shows the associated functions. The case of independent variables is superimposed (yellow) for comparative purposes.

in particular Equations (27) and (28) for a more precise error bound and the nested forgery lemma (29) for a full derivation]. In turn, the extended forgery ensures the genericity of this selection of sample paths B, B' and thus of the event $\mathcal{C}_{X,Y,f,g,\varepsilon}$ as well [the necessary condition $f(0) = g(0) = 0$ can indeed be assumed without loss of generality, see Equation (31) in section 3.3].

To understand Equation (6), imagine first that the random times were bounded with $|X|, |Y| \leq T$. Then $B(X), B'(Y)$ differ from $f(X), g(Y)$ by less than δ for all times X, Y (Equation 1) so the covariance error must be $\mathcal{O}(\delta)$ at most. Now for unbounded random times the distance between sample paths and test trajectories may exceed δ and must actually diverge at long times, which could have led to an infinitely large covariance error if not for the fact that the occurrence of $|X| \geq T$ or $|Y| \geq T$ is very unlikely. So the fit divergence $\nu|t|$ (see Equation 2) is counterbalanced within averages by the fast decay of long times probability [e.g., $\mathbf{P}(|X| \geq |t|) \leq \langle X^2 \rangle / t^2$]. The contribution of $|X| \geq T$ or $|Y| \geq T$ to the covariance error is thus finite and scales as $\mathcal{O}(\nu)$, which leads to Equation (6).

This forgery theorem allows us to probe enough possible relationships to establish statistical dependence or independence, as we explain now. Consider the two probability densities of stochastic covariance shown in **Figure 2**, which were generated using simulations differing only by the presence or absence of coupling between X and Y . The distribution appears significantly wider for the dependent variables, so this suggests that width is the key indicator of a relation. Actually, for the independent variables the narrow peak observed reflects an underlying Dirac delta function (its nonzero width in **Figure 2** is due to finite sampling errors in the covariance estimates). Indeed the vanishing of all stochastic covariances is a necessary condition of independence. The impossibility of sampling nonzero values also turns out to be sufficient.

Brownian independence test. *Two L_2 random variables X, Y are independent iff $\text{cov}[B(X), B'(Y)] = 0$ a.s.*

To prove sufficiency, we show that the hypothesis $\text{cov}[B(X), B'(Y)] = 0$ a.s. implies that all test covariances vanish, which is equivalent to the independence of X and Y (Theorem 10.1 Jacod and Protter [6]). This can be understood concretely using the following thought experiment. Imagine that $\text{cov}[f(X), g(Y)] \neq 0$ for some pair of test trajectories f, g and let us fix, say, $\varepsilon = |\text{cov}[f(X), g(Y)]|/4$ (as in **Figure 2**). We then generate sequentially samples of stochastic covariance until the approximation event $\mathcal{C}_{X,Y,f,g,\varepsilon}$ (Equation 5, shaded area in **Figure 2**) occurs. The forgery of statistical dependences ensures that this sequence stops eventually and our choice of ε , that the last covariance sample is nonzero. However this contradicts our hypothesis, which imposes that all trials result in vanishing covariance (see also section 3.4 for a set-theoretic argument).

Figure 2 suggests a straightforward manner to implement the Brownian independence test in practice. The sample distribution of the stochastic covariance $\text{cov}[B(X), B'(Y)]$ can be generated by drawing a large number of sample paths B, B' , which can be approximated numerically via random walks (black histogram in **Figure 2**). Likewise, a null distribution can be built under the hypothesis of independence, e.g., by accompanying each walk simulation with a random permutation of the sample orderings within X and Y so as to break any dependency, hence leading to a finite-sample version of the Dirac delta (yellow histogram in **Figure 2**). These two distributions can then be compared statistically using, e.g., a Kolmogorov-Smirnov test. Actually, it is sufficient to focus on a comparison of their variances since width is the key parameter here. As it turns out, this provides a more efficient implementation because it is possible to integrate out B, B' analytically in the variance statistic (i.e., all sample paths are probed exhaustively without the need of random walk

simulations). This idea leads back to the notion of Brownian covariance.

2.3. Brownian Covariance Revisited

The forgery of statistical dependences provides an alternative approach to the theory of Brownian covariance, hereafter denoted $\mathcal{B}(X, Y)$. This dependence index emerges naturally in our context as the root mean square (r.m.s.) of the stochastic covariance (or equivalently its standard deviation since the mean $\langle \text{cov}[\mathbf{B}(X), \mathbf{B}'(Y)] \rangle$ vanishes identically by symmetry $\mathbf{B} \longleftrightarrow -\mathbf{B}$, see also **Figure 2**). Thus

$$\mathcal{B}(X, Y)^2 = \langle \text{cov}[\mathbf{B}(X), \mathbf{B}'(Y)]^2 \rangle, \quad (7)$$

which is only a slight reformulation of the definition in Székely GJ and Rizzo [5]. For L_2 random variables X, Y , the quadratic Gaussian functional integrals over the sample paths \mathbf{B}, \mathbf{B}' can be computed analytically and the result reduces to distance covariance, so $\mathcal{B}(X, Y)$ inherits all its properties [5]. Alternatively, we argue here that the central results of the theory follow in a natural manner from Equation (7).

The first key property is that X, Y are independent iff $\mathcal{B}(X, Y) = 0$. This mirrors directly the Brownian independence test since the r.m.s. (7) measures precisely the deviations of stochastic covariance from zero. This argument thus replaces the formal manipulations on the regularized singular integrals that underlie the theory of distance covariance [4]. Furthermore the forgery of statistical dependences clarifies how this works physically: Brownian motion fluctuates enough to make the functional integral (7) probe all the possible test covariances between X and Y .

The second key aspect is the straightforward sample estimation of $\mathcal{B}(X, Y)$ using a parameter-free, algebraic formula. This is an important practical advantage over other dependency measures such as, e.g., mutual information [12]. Instead of relying on the sample formula of distance covariance [4, 5], Equation (7) prompts us to estimate the stochastic covariance (or rather, its square) before averaging over the Brownian paths. So, given n joint samples X_i, Y_i ($i = 1, \dots, n$) of X, Y and an expression for their sample covariance $\widehat{\text{cov}}_n$, the functional integral

$$\widehat{\mathcal{B}}_n(X, Y)^2 = \langle \widehat{\text{cov}}_n[\mathbf{B}(X), \mathbf{B}'(Y)]^2 \rangle_{\text{samples } X_1, Y_1, \dots, X_n, Y_n \text{ fixed}} \quad (8)$$

determines an estimator $\widehat{\mathcal{B}}_n(X, Y)$ of Brownian covariance. If X, Y are L_2 random variables, this procedure allows to build the estimation theory of Brownian (and thus, distance) covariance from that of the elementary covariance. For instance, the rather intricate algebraic formula for the unbiased sample distance covariance [13, 14] is recovered by using, quite naturally from Equations (7) and (8), an unbiased estimator $\widehat{\text{cov}}_n[\cdot, \cdot]^2$ of the covariance squared (see section 3.5 for explicit expressions of these estimators and a derivation of this statement). The unbiasedness property $\langle \widehat{\text{cov}}_n[\cdot, \cdot]^2 \rangle = \text{cov}[\cdot, \cdot]^2$ is then automatically transferred to the corresponding estimator (8), i.e.,

$$\langle \widehat{\mathcal{B}}_n(X, Y)^2 \rangle = \mathcal{B}(X, Y)^2, \quad (9)$$

because the L_2 convergence hypothesis is sufficient to ensure that averaging over the samples X_i, Y_i commutes with the functional integration over the Brownian paths \mathbf{B}, \mathbf{B}' .

Coming back to the implementation of the Brownian independence test, once an estimator $\widehat{\mathcal{B}}_n(X, Y)$ is found, its null distribution under the hypothesis of independence should be derived for formal statistical assessment. This may be done using large n approximations to obtain parametrically the asymptotic distribution or nonparametric methods also valid at small n (such as sample ordering permutations, as suggested above). Asymptotic tests are derived explicitly in Székely et al. [4], Székely and Rizzo [5, 13], where both parametric and nonparametric approaches are also illustrated on examples motivating the usefulness of this nonlinear correlation index in data analysis (including comparisons to linear correlation tests and assessments of statistical power).

It is noteworthy that our construction of Brownian covariance and its estimator can be generalized by formally replacing the Brownian paths \mathbf{B}, \mathbf{B}' with other stochastic processes or fields (in which case we may consider multivariate variables X, Y). This determines a simple rule to engineer a wide array of dependence measures via functional integrals, and opens the question of what processes allow to characterize independence. Our approach relied on the ability to probe generically all possible test covariances but, critically, the class of processes satisfying a forgery of statistical dependences might be relatively restricted. On the other hand, the original theory of Brownian covariance does extend to multidimensional Brownian fields or fractional Brownian motion [5], which are not continuous or Markovian (two properties central for the forgery theorems). So the forgery of statistical dependences provides a new and elegant tool to establish independence, but it may only represent a particular case of a more general theory of functional integral-based correlation measures.

3. MATHEMATICAL ANALYSIS

We now proceed with a more detailed examination of our results. The most technical parts of the proofs are relegated to the Supplementary Material S3.

3.1. Asymptotic Forgery and Duality

We sketched in section 2.1 a derivation of the asymptotic forgery theorem using the law of large numbers. A generalization can actually be developed fully (see Supplementary Material S1). However, this particular case enjoys a concise proof based on a symmetry argument.

The short/long time duality. We start by establishing the duality relation

$$\mathbf{P}(\mathcal{E}_{f,v,T}) = \mathbf{P}(\mathcal{U}_{f_D,v,1/T}), \quad (10)$$

where the dual f_D of the function f is given by

$$f_D(t) = \begin{cases} |t|f(1/t) & \text{for } t \neq 0, \\ 0 & \text{for } t = 0. \end{cases} \quad (11)$$

Proof. By the time-inversion symmetry of Brownian motion [2], replacing $\mathbf{B}(t)$ by $|t|\mathbf{B}(1/t)$ in the right-hand side of Equation (2) determines a new event with the same probability as $\mathcal{E}_{f,v,T}$. Explicitly, this event is

$$\begin{aligned} & \{ |t|\mathbf{B}(1/t) - f(t) < v|t|, \forall |t| \geq T \} \\ & = \{ |\mathbf{B}(t') - f_D(t')| < v, \forall 0 < |t'| \leq 1/T \}, \end{aligned} \quad (12)$$

where $t' = 1/t$ and Equation (11) were used. The condition at $t' = 0$ holds identically since $\mathbf{B}(0) = f_D(0) = 0$, so (12) coincides with $\mathcal{U}_{f_D,v,1/T}$. This yields Equation (10).

Proof of the asymptotic forgery theorem. The theorem naturally follows from this duality and Levy’s forgery. The explicit formula (11) establishes that f_D is continuous [for $t \neq 0$ this corresponds to the continuity of f , and for $t = 0$ to its boundedness $\sup_{t \in \mathbb{R}} |f(t)| \leq M$ since then $|tf(1/t)| \leq M|t| \rightarrow 0$ as $t \rightarrow 0$]. Levy’s forgery theorem thus applies and shows that the right-hand side of Equation (10) is nonzero.

3.2. Local Extended Forgery

We now describe an analytical derivation of the extended forgery theorem that formalizes the intuitive argument given in section 2.1. The ensuing bound for the probability of the joint event (3) does not quite reach that in Equation (4) but is sufficient to ensure genericity.

Restriction to positive-time events. As a preliminary, it will be useful to consider the positive- and negative-time events $\mathcal{U}_{f,\delta,T}^\pm$ and $\mathcal{E}_{f,v,T}^\pm$, which correspond to (1) and (2) except that their defining conditions are enforced only for $0 \leq \pm t \leq T$ and $\pm t \geq T$, respectively. These events are generic too because they are less constrained (e.g., $\mathcal{U}_{f,\delta,T}^\pm$ contains $\mathcal{U}_{f,\delta,T} = \mathcal{U}_{f,\delta,T}^+ \cap \mathcal{U}_{f,\delta,T}^-$) so that

$$\mathbf{P}(\mathcal{U}_{f,\delta,T}^\pm) \geq \mathbf{P}(\mathcal{U}_{f,\delta,T}) > 0, \quad (13)$$

$$\mathbf{P}(\mathcal{E}_{f,v,T}^\pm) \geq \mathbf{P}(\mathcal{E}_{f,v,T}) > 0, \quad (14)$$

by monotonicity of the probability measure $\mathbf{P}(\cdot)$, Levy’s forgery, and the asymptotic forgery.

We are going to focus below on the derivation of

$$\mathbf{P}(\mathcal{U}_{f,\delta,T}^+ \cap \mathcal{E}_{f,v,T}^+) > 0. \quad (15)$$

This is sufficient to prove the extended forgery because the backward-time ($t \leq 0$) part of Brownian motion is merely an independent copy of its forward-time ($t \geq 0$) part. Hence a similar result necessarily holds for the negative-time events and, in turn,

$$\mathbf{P}(\mathcal{J}_{f,\delta,T}) = \prod_{\sigma=\pm} \mathbf{P}(\mathcal{U}_{f,\delta,T}^\sigma \cap \mathcal{E}_{f,v,T}^\sigma) > 0. \quad (16)$$

The integral formula. The first step in the demonstration of Equation (15) relies on the following explicit expression for the joint probability as a functional integral. Let us introduce the family of auxiliary events

$$\mathcal{A}_{f,\delta,T}^x = \{ |\mathbf{B}(t) + x - f(T+t)| < \delta + vt, \forall t \geq 0 \} \quad (17)$$

and denote their probability by

$$p_{f,\delta,T}(x) = \mathbf{P}(\mathcal{A}_{f,\delta,T}^x) \quad (18)$$

for all $x \in \mathbb{R}$. Then

$$\mathbf{P}(\mathcal{U}_{f,\delta,T}^+ \cap \mathcal{E}_{f,v,T}^+) = \langle \mathbf{1}_{\mathcal{U}_{f,\delta,T}^+} p_{f,\delta,T}(\mathbf{B}(T)) \rangle. \quad (19)$$

The first factor in this expectation value denotes the indicator function of the event $\mathcal{U}_{f,\delta,T}^+$ and enforces the constraint that all considered sample paths \mathbf{B} must lie within the uniform neighborhood (1) for $0 \leq t \leq T$. In the second factor the function (18) is evaluated at the position of the random path $\mathbf{B}(t)$ at $t = T$. As we explain below this function is Borel measurable, so $p_{f,\delta,T}(\mathbf{B}(T))$ represents a proper random variable and the expectation value is well defined.

Proof. The formula (19) is a direct consequence of the two following statements, each involving the conditional probability $\mathbf{P}[\mathcal{E}_{f,v,T}^+ | \mathbf{B}(T)]$ that the event $\mathcal{E}_{f,v,T}^+$ occurs under the constraint that the Brownian motion passes at time $t = T$ through a given location x randomly distributed as $\mathbf{B}(T)$:

$$\mathbf{P}(\mathcal{U}_{f,\delta,T}^+ \cap \mathcal{E}_{f,v,T}^+) = \langle \mathbf{1}_{\mathcal{U}_{f,\delta,T}^+} \mathbf{P}[\mathcal{E}_{f,v,T}^+ | \mathbf{B}(T)] \rangle, \quad (20)$$

and

$$\mathbf{P}[\mathcal{E}_{f,v,T}^+ | \mathbf{B}(T)] = p_{f,\delta,T}(\mathbf{B}(T)) \quad \text{a.s.} \quad (21)$$

They follow from fairly standard arguments about Brownian motion [2, 3] that we detail in the Supplementary Materials S3.1 and S3.2. The first implements the Markov property that $\mathcal{E}_{f,v,T}^+$ depends on its past only via $\mathbf{B}(t)$ at the boundary time $t = T$. The second provides an explicit representation of the random variable $\mathbf{P}[\mathcal{E}_{f,v,T}^+ | \mathbf{B}(T)]$ (see also Chapter 1, Theorem 12 in Kraskov et al. [2]).

The next step is to show that the integrand in the right-hand side of Equation (19) cannot vanish identically, which provides a “local” version of the extended forgery. The full theorem will follow by integration.

Extended forgery theorem (local version). *There exists a subinterval J of the bottleneck interval $f(T) - \delta < x < f(T) + \delta$ at $t = T$ such that*

$$\mathbf{P}(\mathcal{U}_{f,\delta,T}^+ \cap \{ \mathbf{B}(T) \in J \}) \geq \mathbf{P}(\mathcal{W}_{g,\ell,T}^+) \quad (22)$$

for some continuous function g satisfying $g(0) = 0$ and some distance parameter $\ell > 0$, and

$$p_{f,\delta,T}(x) > \mathbf{P}(\mathcal{E}_{f,v,T}^+) \quad (23)$$

for all $x \in J$.

By Equations (13) and (14), both lower bounds are positive. The two parts of the theorem are closely related to Levy’s forgery and the asymptotic forgery, and we treat them separately.

Proof of (22). This property actually holds for arbitrary subintervals J , which we shall choose open and parameterized as

$x_0 - \ell < x < x_0 + \ell$ with $|x_0 - f(T)| \leq \delta - \ell$ to ensure inclusion in the bottleneck. We also define the continuous function g by $g(t) = f(t) + t|x_0 - f(T)|/T$. This setup ensures that $\mathcal{W}_{g,\ell,T}^+ \subset \mathcal{W}_{f,\delta,T}^+$ and $\mathcal{W}_{g,\ell,T}^+ \subset \{\mathbf{B}(T) \in J\}$, as the condition $|\mathbf{B}(t) - g(t)| < \ell$ for all $0 \leq t \leq T$ implies

$$|\mathbf{B}(t) - f(t)| \leq |\mathbf{B}(t) - g(t)| + |g(t) - f(t)| < \ell + |x_0 - f(T)| \leq \delta$$

and $|\mathbf{B}(T) - x_0| = |\mathbf{B}(T) - g(T)| < \ell$ as $g(T) = x_0$. The bound (22) follows from these two inclusions by the monotonicity of $\mathbf{P}(\cdot)$.

Two properties of the function $p_{f,\delta,T}$. For the second part it is useful to interject here the following simple statements about the function (18):

- (i) $p_{f,\delta,T}$ vanishes identically outside the bottleneck,
- (ii) $p_{f,\delta,T}$ is continuous within the bottleneck (and therefore, it is Borel measurable as well).

The first claim rests on the observation that the set (17) is empty whenever $|x - f(T)| \geq \delta$ (since its defining condition at $t = 0$ cannot be satisfied, as $\mathbf{B}(0) = 0$), so we find $p_{f,\delta,T}(x) = \mathbf{P}(\emptyset) = 0$.

The second claim may appear quite clear as well, since the set (17) should vary continuously with x , but this intuition is not quite right. For a more accurate statement, let us fix a point x in the bottleneck and consider an arbitrary sequence x_n converging to it. Then the limit event of $\mathcal{A}_{f,\delta,T}^{x_n}$ (assuming it exists) coincides with $\mathcal{A}_{f,\delta,T}^x$ modulo a zero-probability set, so by continuity of $\mathbf{P}(\cdot)$

$$\lim_{n \rightarrow \infty} p_{f,\delta,T}(x_n) = p_{f,\delta,T}(x). \tag{24}$$

The full proof appears rather technical, so we only sketch the key ideas here and relegate the details to the Supplementary Material S3.3. The limit event imposes that sample paths lie within the expanding neighborhood associated with (17) but are allowed to reach its boundary [essentially because the large n limit of the strict inequalities (17) at $x = x_n$ yields a nonstrict inequality]. However the latter hitting event a.s. never happens because the typical roughness of Brownian motion forbids meeting a boundary curve without crossing it and thus, leaving the neighborhood (see Supplementary Material S3.4 for this “boundary-crossing law”, which generalizes Lemma 1 on page 283 of [3] to time-dependent levels.).

Proof of (23). The key observation here is that

$$\mathbf{P}(\mathcal{E}_{f,v,T}^+) = \int_{-\infty}^{\infty} dx \frac{\exp(-x^2/2T)}{\sqrt{2\pi T}} p_{f,\delta,T}(x), \tag{25}$$

which is merely the integral formula (19) with the constraint $\mathbf{B} \in \mathcal{W}_{f,\delta,T}^+$ removed (equivalently, this is Supplemental Equation S.18 with $S = \mathbb{R}$) and the expectation over $\mathbf{B}(T)$ made explicit. If $p_{f,\delta,T}(x) \leq \mathbf{P}(\mathcal{E}_{f,v,T}^+)$ everywhere, then for almost all x this inequality must saturate to an equality by Equation (25) and thus $p_{f,\delta,T}(x) > 0$ by Equation (14). This contradicts property (i) of $p_{f,\delta,T}$, so that Equation (23) must hold for at least one point x

in the bottleneck, and thus also on some subinterval J by the continuity property (ii).

Proof of the extended forgery theorem. Finally we combine the local version of the theorem with the integral formula (19) to obtain

$$\mathbf{P}(\mathcal{W}_{f,\delta,T}^+ \cap \mathcal{E}_{f,v,T}^+) > \mathbf{P}(\mathcal{W}_{g,\ell,T}^+) \mathbf{P}(\mathcal{E}_{f,v,T}^+). \tag{26}$$

Indeed, further constraining the paths to $\mathbf{B}(T) \in J$ allows to bound the right-hand side of (19) from below by

$$\begin{aligned} \langle \mathbf{1}_{\mathcal{W}_{f,\delta,T}^+ \cap \{\mathbf{B}(T) \in J\}} p_{f,\delta,T}(\mathbf{B}(T)) \rangle &> \mathbf{P}(\mathcal{W}_{f,\delta,T}^+ \cap \{\mathbf{B}(T) \in J\}) \mathbf{P}(\mathcal{E}_{f,v,T}^+) \\ &\text{by (23),} \\ &\geq \mathbf{P}(\mathcal{W}_{g,\ell,T}^+) \mathbf{P}(\mathcal{E}_{f,v,T}^+) \\ &\text{by (22).} \end{aligned}$$

In passing through the first inequality we also used the identity $\langle \mathbf{1}_{\mathcal{M}} \rangle = \mathbf{P}(\mathcal{M})$ valid for any event \mathcal{M} . The theorem (15) follows from the inequality (26) together with Equations (13) and (14).

It is noteworthy that we stated and proved the extended forgery under the assumption that $v = \delta/T$ for convenience, but in the above arguments this restriction was actually artificial so the theorem holds for arbitrary parameters $\delta, v, T > 0$.

3.3. Nested Forgery of Statistical Dependences

We showed in section 2.2 that the forgery of statistical dependences is induced by the extended forgery using the somewhat rough estimate (6) of the covariance error. We provide now an exact upper bound and a complete proof of the theorem.

Error bound. For arbitrary sample paths $\mathbf{B} \in \mathcal{I}_{f,\delta,T}$ and $\mathbf{B}' \in \mathcal{I}_{g,\delta,T}$, we have

$$|\text{cov}[\mathbf{B}(X), \mathbf{B}'(Y)] - \text{cov}[f(X), g(Y)]| < \varepsilon_B(\delta, v) \tag{27}$$

where $v = \delta/T$ and the error bound $\varepsilon_B(\delta, v)$ is given by

$$\begin{aligned} \varepsilon_B(\delta, v) &= 2(M_f + M_g)\delta + 2\delta^2 + 2(M_g \langle |X| \rangle + M_f \langle |Y| \rangle)v \\ &\quad + 2(\langle |X| \rangle + \langle |Y| \rangle)\delta v + (\langle |XY| \rangle + \langle |X| \rangle \langle |Y| \rangle)v^2, \end{aligned} \tag{28}$$

with $M_f = \sup_{t \in \mathbb{R}} |f(t)|$ and $M_g = \sup_{t \in \mathbb{R}} |g(t)|$.

The derivation of this estimate relies on a relatively straightforward application of a series of inequalities and is relegated to the Supplementary Material S3.5. Its significance rests on the fact that, when the polynomial coefficients in (28) are finite, the covariance error can be made arbitrarily small by taking $\delta, v \rightarrow 0$. Hence we obtain the following key intermediate result.

The nested forgery lemma. Assume that $M_f, M_g, \langle |X| \rangle, \langle |Y| \rangle$, and $\langle |XY| \rangle$ are finite, and let $\varepsilon > 0$. Then there exists $\delta, T > 0$ such that

$$\mathcal{I}_{f,\delta,T} \times \mathcal{I}_{g,\delta,T} \subset \mathcal{C}_{X,Y,f,g,\varepsilon}, \tag{29}$$

where the prime on event $\mathcal{I}_{g,\delta,T}$ indicates that it applies to the process B' .

Proof of the forgery theorem of statistical dependences. This is a direct corollary. The L_2 hypothesis and the boundedness assumption on f, g ensure that the lemma (29) applies. Using the monotonicity of $\mathbf{P}(\cdot)$ and the independence of B, B' then yields

$$\mathbf{P}(\mathcal{C}_{X,Y,f,g,\varepsilon}) \geq \mathbf{P}(\mathcal{I}_{f,\delta,T} \times \mathcal{I}'_{g,\delta,T}) = \mathbf{P}(\mathcal{I}_{f,\delta,T}) \mathbf{P}(\mathcal{I}'_{g,\delta,T}). \quad (30)$$

It is now tempting to invoke the extended forgery theorem, but here the conditions $f(0) = g(0) = 0$ were not imposed. This is however not an issue thanks to the translation invariance of covariance, i.e., $\text{cov}[f(X), g(Y)] = \text{cov}[f(X) - f(0), g(Y) - g(0)]$. Thus

$$\begin{aligned} \mathbf{P}(\mathcal{C}_{X,Y,f,g,\varepsilon}) &= \mathbf{P}(\mathcal{C}_{X,Y,f-f(0),g-g(0),\varepsilon}) \\ &\geq \mathbf{P}(\mathcal{I}_{f-f(0),\delta,T}) \mathbf{P}(\mathcal{I}'_{g-g(0),\delta,T}) > 0, \end{aligned} \quad (31)$$

where we used Equation (30) and the extended forgery theorem in the second line.

Note that the L_2 assumption used in section 2.2 was slightly stronger than needed, as made clear by the nested forgery lemma.

3.4. Brownian Independence Test and Dichotomy

To prove that the assertion $\text{cov}[B(X), B'(Y)] = 0$ a.s. implies $\text{cov}[f(X), g(Y)] = 0$, we used in section 2.2 a discrete sampling method for which it is straightforward that the covariance approximation event (5) must occur simultaneously with the zero covariance event

$$\mathcal{Z}_{X,Y} = \{\text{cov}[B(X), B'(Y)] = 0\}. \quad (32)$$

Their compatibility, which was indeed central in our derivation, is actually a general property of probability theory and we use it here to provide an alternative, set-theoretic argument.

The dichotomy lemma. For arbitrary functions f, g and parameter ε , we have

$$\mathcal{C}_{X,Y,f,g,\varepsilon} \cap \mathcal{Z}_{X,Y} = \begin{cases} \mathcal{Z}_{X,Y} & \text{if } |\text{cov}[f(X), g(Y)]| < \varepsilon \\ \emptyset & \text{otherwise.} \end{cases} \quad (33)$$

Proof. It is a direct consequence of Equations (5) and (32) that the intersection is characterized equivalently by the two conditions $\text{cov}[B(X), B'(Y)] = 0$ and $|\text{cov}[f(X), g(Y)]| < \varepsilon$.

Second proof of sufficiency for the Brownian independence test. By the forgery of statistical dependences the event $\mathcal{C}_{X,Y,f,g,\varepsilon}$ is generic for f, g bounded and continuous and $\varepsilon > 0$, and by hypothesis the event $\mathcal{Z}_{X,Y}$ occurs a.s. Since the intersection of a generic event and an almost sure event is never empty [if it were empty, the generic event would be a subset of a zero probability event (i.e., the complement of the almost sure event), which is forbidden by monotonicity of $\mathbf{P}(\cdot)$], the second possibility in Equation (33) is ruled out so $|\text{cov}[f(X), g(Y)]| < \varepsilon$ must hold true. Since $\varepsilon > 0$ was arbitrary, we conclude that $\text{cov}[f(X), g(Y)] = 0$.

3.5. Unbiased Estimation of Brownian Covariance

We now exemplify how an explicit estimator of Brownian covariance can be derived from the functional integral (8). Our construction enforces unbiasedness at finite sampling and allows to recover the unbiased sample formula of distance covariance [13, 14] that we review first.

The unbiased estimator. Let us introduce the distance $a_{ij} = |X_i - X_j|$ between the samples X_i of X as well as its ‘‘U-centered’’ version

$$A_{ij} = a_{ij} - \frac{\sum_k (a_{ik} + a_{kj})}{n-2} + \frac{\sum_{k,l} a_{kl}}{(n-1)(n-2)}, \quad (34)$$

where $1 \leq i, j, k, l \leq n$. The analogous matrices for the corresponding samples of Y are denoted b_{ij} and B_{ij} , respectively. With these notations and assuming $n \geq 4$,

$$\widehat{B}_n(X, Y)^2 = \frac{1}{4n(n-3)} \sum_{\substack{\text{distinct} \\ ij}} A_{ij} B_{ij}. \quad (35)$$

This expression differs from the formula given in Székely and Rizzo [13] and Rizzo and Székely [14] by a trivial factor 1/4 due to our use of the standard normalization for Brownian motion. The asymptotic distribution of this estimator under the null hypothesis of independence was worked out in Székely and Rizzo [13]. The unbiasedness property (9) of Equation (35) was also proven in Székely and Rizzo [13] and Rizzo and Székely [14], but here it will follow directly from our derivation based on the functional integral (8), which starts naturally from an unbiased estimation of the covariance squared.

Estimation of covariance squared. It is convenient to introduce the variables $x = B(X)$, $y = B'(Y)$ and their samples $x_i = B(X_i)$, $y_i = B'(Y_i)$, all being defined at fixed sample paths B, B' . An unbiased estimator for $\text{cov}(x, y)$ itself is well-known from elementary statistics, but it can be checked by developing its square that the estimation of $\text{cov}(x, y)^2$ is then hampered by systematic errors of order $1/n$. Here, given $n \geq 4$, we shall rather define

$$\widehat{\text{cov}}_n(x, y)^2 = \frac{(n-4)!}{n!} \sum' x_i x_j (y_i y_j - y_i y_k - y_j y_k + y_k y_l), \quad (36)$$

where the primed sum is taken over all distinct indices $1 \leq i, j, k, l \leq n$. This expression differs from the aforementioned development by $\mathcal{O}(1/n)$ and is indeed free from finite sampling biases.

This can be proven by averaging Equation (36) over the n joint samples x_i, y_i . Indeed, using the identities $\langle x_i x_j y_i y_j \rangle = \langle xy \rangle^2$, $\langle x_i x_j y_i y_k \rangle = \langle xy \rangle \langle x \rangle \langle y \rangle$, and $\langle x_i x_j y_k y_l \rangle = \langle x \rangle^2 \langle y \rangle^2$ and the fact that the primed sum contains $n!/(n-4)!$ terms, we find

$$\langle \widehat{\text{cov}}_n(x, y)^2 \rangle = \langle xy \rangle^2 - 2 \langle xy \rangle \langle x \rangle \langle y \rangle + \langle x \rangle^2 \langle y \rangle^2 = \text{cov}(x, y)^2. \quad (37)$$

Derivation of Equation (35) from Equation (8). We now average Equation (36) with $x_i = B(X_i)$ and $y_i = B'(Y_i)$ over

the independent random paths \mathbf{B}, \mathbf{B}' , while the samples X_i, Y_i are being kept constant. The computation factors into the functional integration of $x_i x_j = \mathbf{B}(X_i)\mathbf{B}(X_j)$ and $y_i y_j = \mathbf{B}'(Y_i)\mathbf{B}'(Y_j)$, so $\widehat{\mathcal{B}}_n(\mathbf{X}, \mathbf{Y})^2$ is obtained from the right-hand side of Equation (36) by replacing $x_i x_j$ and $y_i y_j$ with the autocorrelation functions [2]

$$\begin{aligned} \langle \mathbf{B}(X_i)\mathbf{B}(X_j) \rangle &= \frac{1}{2}(|X_i| + |X_j| - a_{ij}), \\ \langle \mathbf{B}'(Y_i)\mathbf{B}'(Y_j) \rangle &= \frac{1}{2}(|Y_i| + |Y_j| - b_{ij}). \end{aligned} \quad (38)$$

This substitution rule can actually be simplified further to $x_i x_j \rightarrow -a_{ij}/2$ because the terms involving $|X_i|$ cancel out thanks to the algebraic identity

$$\sum_{\substack{\text{all distinct} \\ j, k, l \neq i}} (y_i y_j - y_i y_k - y_j y_k + y_k y_l) = 0. \quad (39)$$

Similar cancellations also allow to use $y_i y_j \rightarrow -b_{ij}/2$. We thus obtain the unbiased estimator

$$\widehat{\mathcal{B}}_n(\mathbf{X}, \mathbf{Y})^2 = \frac{(n-4)!}{4n!} \sum' a_{ij} (b_{ij} - b_{ik} - b_{kj} + b_{kl}). \quad (40)$$

The equivalence with Equation (35) is not obvious at first sight. To make contact with it, let us fix i, j and consider the sum of the second factor over k, l , all indices being distinct. This contribution is the sum of the following four terms:

$$\begin{aligned} \sum'_{k,l} b_{ij} &= (n-2)(n-3)b_{ij}, \\ -\sum'_{k,l} b_{ik} &= (n-3)(b_{ij} - \sum_k b_{ik}), \\ -\sum'_{k,l} b_{kj} &= (n-3)(b_{ij} - \sum_k b_{kj}), \\ \sum'_{k,l} b_{kl} &= \sum_{k,l} b_{kl} - 2\sum_k (b_{ik} + b_{kj}) + 2b_{ij}, \end{aligned}$$

REFERENCES

- Sethna JP. *Statistical Mechanics: Entropy, Order Parameters and Complexity*. New York, NY : Oxford University Press (2006).
- Freedman D. *Brownian Motion and Diffusion*. New York, NY: Springer-Verlag (1983).
- Gikhman II, Skorokhod AV. *Introduction to the Theory of Random Processes*. Mineola, NY: Dover (1969).
- Székely GJ, Rizzo ML, Bakirov NK. Measuring and testing dependence by correlation of distances. *Ann Stat.* (2007) **35**:2769–94. doi: 10.1214/009053607000000505
- Székely GJ, Rizzo ML. Brownian distance covariance. *Ann Appl Stat.* (2009) **3**:1236–65. doi: 10.1214/09-AOAS312
- Jacod J, Protter PE. *Probability Essentials*. 2nd ed. Berlin: Universitext. Springer-Verlag (2013).
- Gretton A, Györfi L. Consistent nonparametric tests of independence. *J Mach Learn Res.* (2010) **11**:1391–423.
- Newton MA. Introducing the discussion paper by Székely and Rizzo. *Ann Appl Stat.* (2009) **3**:1233–5. doi: 10.1214/09-AOAS34INTRO
- Székely GJ, Rizzo ML. Rejoinder: brownian distance covariance. *Ann Appl Stat.* (2010) **3**:1303–8. doi: 10.1214/09-AOAS312REJ
- Zhou Z. Measuring nonlinear dependence in time series, a distance correlation approach. *J Time Ser Anal.* (2012) **33**:438–57. doi: 10.1111/j.1467-9892.2011.00780.x

where the sums in the right-hand sides now run over unconstrained indices $1 \leq k, l \leq n$. The total sum thus reduces to $(n-1)(n-2)B_{ij}$, so Equation (40) can be rewritten as

$$\widehat{\mathcal{B}}_n(\mathbf{X}, \mathbf{Y})^2 = \frac{1}{4n(n-3)} \sum_{\substack{\text{distinct} \\ i,j}} a_{ij} B_{ij}. \quad (41)$$

Finally we recover Equation (35) because a_{ij} can be replaced by its U-centering A_{ij} . Indeed the extra terms cancel in the sum thanks to the centering property $\sum_{j \neq i} B_{ij} = 0$ (for all i fixed), which can be checked from the definition (34).

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

ACKNOWLEDGMENTS

This research was carried over in the context of methodological developments for the MEG project at the CUB – Hôpital Erasme (Université libre de Bruxelles, Brussels, Belgium), which is financially supported by the Fonds Erasme (convention Les Voies du Savoir, Fonds Erasme, Brussels, Belgium).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fams.2018.00019/full#supplementary-material>

- Geerligs L, Cam-Can, Henson RN. Functional connectivity and structural covariance between regions of interest can be measured more accurately using multivariate distance correlation. *NeuroImage* (2016) **135**:16–31. doi: 10.1016/j.neuroimage.2016.04.047
- Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. *Phys Rev E* (2004) **69**:066138. doi: 10.1103/PhysRevE.69.066138
- Székely GJ, Rizzo ML. The distance correlation t -test of independence in high dimension. *J Multivar Anal.* (2013) **117**:193–213. doi: 10.1016/j.jmva.2013.02.012
- Rizzo ML, Székely GJ. Partial distance correlation with methods for dissimilarities. *Ann Stat.* (2014) **42**:2382–412. doi: 10.1214/14-AOS125

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Wens. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.