A DIFFERENT CONSTRUCTION OF GAUSSIAN FIELDS FROM MARKOV CHAINS: DIRICHLET COVARIANCES

PERSI DIACONIS AND STEVEN N. EVANS

ABSTRACT. We study a class of Gaussian random fields with negative correlations. These fields are easy to simulate. They are defined in a natural way from a Markov chain that has the index space of the Gaussian field as its state space. In parallel with Dynkin's investigation of Gaussian fields having covariance given by the Green's function of a Markov process, we develop connections between the occupation times of the Markov chain and the prediction properties of the Gaussian field. Our interest in such fields was initiated by their appearance in random matrix theory.

1. Introduction

Let $\mathcal{X}$ be a finite set. Our goal is to define and study a rather general class of mean zero Gaussian fields $\mathbf{Z} = \{Z_x\}_{x \in \mathcal{X}}$ with *negative* correlations. These fields may be used for smoothing, interpolation and Bayesian prediction as in [Ylv87, Bes74, BG93, BGHM95, BK95, BH99], where there are extensive further references.

The definition begins with a reversible Markov chain $\mathbf{X}$ with state space $\mathcal{X}$. Our development in the body of the paper will be for continuous time chains, because the exposition is somewhat cleaner in continuous time, but we will first explain the construction in the discrete time setting. Let $P(x,y)$ be the transition matrix of a (conservative) discrete time Markov chain $\mathbf{X}$ with state space $\mathcal{X}$. Assume for simplicity that the chain $\mathbf{X}$ has no holding (that is, the one-step transitions of $\mathbf{X}$ are always to another state). Thus, $P(x,y) \geq 0$ for all $x, y \in \mathcal{X}$, $\sum_{y \in \mathcal{X}} P(x,y) = 1$ for all $x \in \mathcal{X}$, and $P(x,x) = 0$ for all $x \in \mathcal{X}$. Suppose further that the chain $\mathbf{X}$ is reversible with respect to some probability vector $(\pi(x))_{x,y \in \mathcal{X}}$; that is, $\pi(x)P(x,y) = \pi(y)P(y,x)$ for all $x, y \in \mathcal{X}$. The matrix Σ given by

$$\Sigma(x,y) := \begin{cases} \pi(x), & \text{if } x = y, \\ -\pi(x)P(x,y), & \text{if } x \neq y, \end{cases}$$

is positive semi–definite, and hence is the covariance matrix of a mean zero Gaussian field $\mathbf{Z} = \{Z_x\}_{x \in \mathcal{X}}$. Note that the dependence structure of the field $\mathbf{Z}$ accords with the local neighbourhood structure defined by the transition matrix P: if $P(x,y) = 0$ (that is, the chain $\mathbf{X}$ is unable to go from x to y in one step), then the Gaussian random variables Z_x and Z_y are independent.

Date: April 10, 2001.

Key words and phrases. random field, Dirichlet form, negative correlation, simulation, Markov process, potential theory.

Research supported in part by NSF grant DMS-9504379.

Research supported in part by NSF grants DMS-9703845 and DMS-0071468.

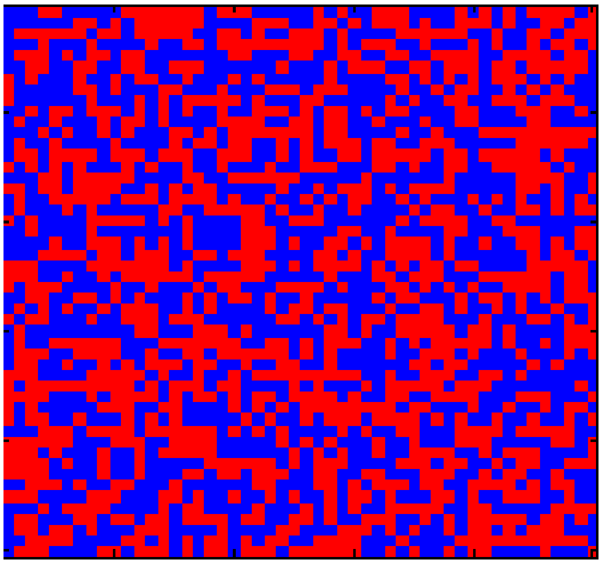

FIGURE 1. Signs of the Gaussian field arising from the simple random walk on the 50×50 discrete torus

Example 1.1. Let $\mathcal{X}$ be the points of the $n \times n$ discrete torus (that is, $\mathbb{Z}_n \times \mathbb{Z}_n$, where $\mathbb{Z}_n$ is the group of integers modulo n), and let P be the transition matrix of nearest neighbour random walk on $\mathcal{X}$. Thus, $P(x, y) = \frac{1}{4}$ if x and y are adjacent (that is, $x - y \in \{(\pm 1, 0), (0, \pm 1)\}$) and $P(x, y) = 0$ otherwise. This chain is reversible with respect to the uniform distribution $\pi(x) = \frac{1}{n^2}$. A realisation of the resulting field is shown in Figure 1 for the case $n = 50$. Sites at which the corresponding Gaussian variable is positive (resp. negative) are coloured black (resp. grey).

For the sake of comparison, the corresponding picture for a field of i.i.d. Gaussian random variables is shown in Figure 2. Note that the negative correlation is apparent to the eye as a more clustered pattern. This phenomenon is an example of the Julesz conjecture, which claims that the eye can only distinguish first and second order statistical features (densities and correlations). A review of the literature on this conjecture and its connections to de Finetti's theorem – an early joint interest of Bretagnolle and Dacunha-Castelle – is in [DF81].

1.1. **Continuous time and Dirichlet forms.** As we noted above, it will be more convenient to work with continuous time Markov chains. To this end, let $\mathbf{X}$ now be a continuous time Markov chain on the finite state space $\mathcal{X}$. Write Q for the associated infinitesimal generator and suppose that $\mathbf{X}$ is reversible with respect to the probability measure π (that is, $\pi(x)Q(x, y) = \pi(y)Q(y, x)$ for all $x, y \in \mathcal{X}$). We do not suppose that $\mathbf{X}$ is conservative. That is, we allow $\sum_y Q(x, y) < 0$, in which case the chain is killed at rate $-\sum_y Q(x, y)$ when it is in state x.

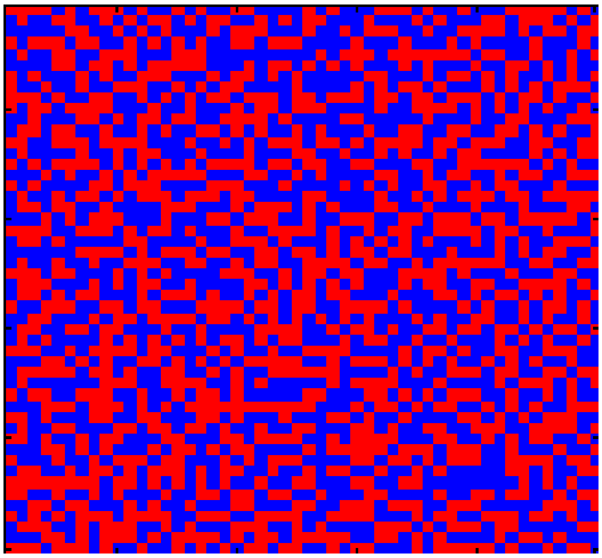

FIGURE 2. Signs of the i.i.d. Gaussian field on the 50×50 discrete torus

Set $L^2(\mathcal{X}, \pi) := \{f : \mathcal{X} \to R\}$ equipped with the inner product $\langle f \mid g \rangle := \sum_x f(x)g(x)\pi(x)$. The kernel Q operates on L^2 by

$$Qf(x) = \sum_y Q(x,y)f(y).$$

Reversibility of $\mathbf{X}$ is equivalent to requiring that the operator Q is self–adjoint on L^2. Of course, if P is the transition matrix of a reversible, conservative, discrete time chain with no holding as above, then $Q = P - I$ is the infinitesimal generator of a reversible, conservative, continuous time chain, namely the chain that exits state $x \in \mathcal{X}$ at rate 1 and jumps to state $y \neq x$ with probability $P(x,y)$ upon exiting. Consequently, the discrete time construction above can be subsumed under the more general construction we are now considering.

The usual quadratic form associated with Q is the Dirichlet form:

$$\begin{aligned}
\mathcal{E}(f,g) &:= -\langle Qf \mid g \rangle \\
&= \frac{1}{2} \sum_{x,y} (f(x) - f(y))(g(x) - g(y))\pi(x)Q(x,y) \\
&\quad + \sum_x f(x)g(x)\kappa(x),
\end{aligned} \tag{1.1}$$

where

$$\kappa(x) := -\sum_y \pi(x)Q(x,y) = -\sum_y \pi(y)Q(y,x) \geq 0.$$

It is clear from (1.1) that the Dirichlet form is a positive semi–definite, self–adjoint quadratic form. The two terms on the right–hand side of (1.1) are called, respectively, the jump part and the killing part of the form. If $\mathbf{X}$ is conservative (that is, no killing occurs), then $\sum_y Q(x,y) = 0$ for all $x \in \mathcal{X}$ and $\kappa = 0$. A standard reference for Dirichlet forms is [FOT94], but we find the original paper by Beurling and Deny [BD58] useful and readable.

It follows from (1.1) that

$$\Sigma(x, y) := -\pi(x)Q(x, y) \tag{1.2}$$

is a positive semi-definite self–adjoint matrix. Hence Σ is the covariance of a mean zero Gaussian field $\mathbf{Z} = \{Z_x\}_{x\in\mathcal{X}}$ indexed by $\mathcal{X}$.

Example 1.2. Set $\mathcal{X} = \mathbb{Z}_n$, the integers modulo n. Take $\mathbf{X}$ to be nearest neighbour random walk with unit jump rate, so

$$Q(x, y) = \begin{cases} -1, & \text{if } x = y, \\ \frac{1}{2}, & \text{if } x - y = \pm 1, \\ 0, & \text{otherwise.} \end{cases}$$

Then $\pi(x) = \frac{1}{n}$, $\Sigma(x, x) = \frac{1}{n}$, $\Sigma(x, x \pm 1) = -\frac{1}{2n}$, and $\Sigma(x, y) = 0$ otherwise. When $n = 5$ the matrix $\Sigma(x, y)$ appears as the circulant

$$\frac{1}{5}\begin{pmatrix} 1 & -\frac{1}{2} & 0 & 0 & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{2} & 0 & 0 & -\frac{1}{2} & 1 \end{pmatrix}.$$

1.2. **Outline of the rest of the paper.** In Section 2 we develop some properties of this construction. We give a simple procedure for simulating the field $\mathbf{Z}$ using independent Gaussian random variables associated with the "edges" (x, y) such that $Q(x, y) > 0$. Generating realisations of Gaussian fields on grids or graphs with general covariances can be a complex enterprise. A useful review of the literature is in [GS89].

In Section 3 we show how the problem of using the observations $\{Z_y\}_{y\in B}$ to predict Z_x for $x \notin B \subset \mathcal{X}$ is intimately related to the properties of the occupation times of the Markov chain $\mathbf{X}$. In Section 4 we indicate how certain questions that involve minimising the variance of a linear combination $\sum_x f(x)Z_x$ subject to constraints can be related to the potential theory of $\mathbf{X}$.

We conclude this Introduction with some comments on the background and context of this paper.

1.3. **A random matrix connection.** Our interest in this construction began with some results in random matrix theory. Let U_n be the unitary group of $n\times n$ matrices M with $MM^* = I$. Elements of U_n have eigenvalues on the unit circle $\mathbb{T}$ in the complex plane. The study of the distribution of these eigenvalues under Haar measure makes up a chapter of random matrix theory (see, for example, [Meh91]).

Let $H_2^{\frac{1}{2}}$ denote the space of functions $f \in L^2(\mathbb{T})$ such that

$$\|f\|_{\frac{1}{2}}^2 := \sum_{j\in\mathbb{Z}} |\hat{f}_j|^2|j| < \infty,$$

and define an inner product on $H_2^{\frac{1}{2}}$ by

$$\langle f, g\rangle_{\frac{1}{2}} := \sum_{j\in\mathbb{Z}} \hat{f}_j\overline{\hat{g}_j}|j|.$$

Alternatively, $H_2^{\frac{1}{2}}$ is the space of functions $f \in L^2(\mathbb{T})$ such that

$$\frac{1}{16\pi^2} \iint \frac{(f(\phi) - f(\theta))^2}{\sin^2\left(\frac{\phi-\theta}{2}\right)} \, d\theta \, d\phi < \infty, \tag{1.3}$$

and, moreover,

$$\langle f, g \rangle_{\frac{1}{2}} = \frac{1}{16\pi^2} \iint \frac{(f(\phi) - f(\theta))(g(\phi) - g(\theta))}{\sin^2\left(\frac{\phi-\theta}{2}\right)} \, d\theta \, d\phi$$

(see Equations (1.2.18) and (1.2.21) of [FOT94]).

In independent work, Johansson [Joh97] and Diaconis-Shahshahani [DS94] proved the following result which was extended in [DE01].

Theorem 1.3. *Choose $M \in U_n$ from Haar measure. For f in $H_2^{\frac{1}{2}}$ let $W_f(M) = \sum_{j=1}^n f(e^{i\theta_j})$, where $\theta_1, \ldots, \theta_n$ are the eigenvalues of M. Then, as n tends to infinity, for any finite collection $f_1, f_2, \ldots f_K$ in $H_2^{\frac{1}{2}}$*

$$\{W_{f_k}(M)\}_{k=1}^K \Rightarrow \{Z_{f_k}\}_{k=1}^K,$$

where $\{Z_f : f \in H_2^{\frac{1}{2}}\}$ is a mean zero Gaussian field with covariance $\mathbb{E}[Z_f Z_g] = \langle f, g \rangle_{\frac{1}{2}}$.

The space $H_2^{\frac{1}{2}}$ is an example of a Bessel–potential function space and it coincides with the Besov space $B_{2,2}^{\frac{1}{2}}$, the Sobolev–Lebesgue space $F_{2,2}^{\frac{1}{2}}$ and the Lipschitz space $\Lambda_{2,2}^{\frac{1}{2}}$ (see Equations (18) and (19) in §3.5.4 and Equation (13) in §3.5.1 of [ST87]). However, for our purposes the interesting observation is that the space $H_2^{\frac{1}{2}}$ equipped with the inner product $\langle \cdot, \cdot \rangle_{\frac{1}{2}}$ is nothing other than the Dirichlet space and Dirichlet form of the symmetric Cauchy process on the circle (see Example 1.4.2 of [FOT94]). (The symmetric Cauchy process on the circle is just the usual symmetric Cauchy process on the line wrapped around the circle.) It is possible to carry through much of what we do in the discrete state space setting of this paper to Dirichlet forms of Markov processes on general state spaces, but we do not pursue that direction here.

Note also that if we take the complex Poisson integral of $f \in L^2(\mathbb{T})$, namely

$$\begin{aligned} \mathcal{P}f(z) &:= \frac{1}{2\pi} \int \frac{e^{i\theta} + z}{e^{i\theta} - z} f(\theta) \, d\theta \\ &= \hat{f}_0 + 2 \sum_{j=1}^{\infty} \hat{f}_j z^j, \quad |z| < 1, \end{aligned}$$

then, letting m denote Lebesgue measure on the disk $\{z \in \mathbb{C} : |z| < 1\}$,

$$\begin{aligned} \int \left| \frac{d\mathcal{P}f(z)}{dz} \right|^2 m(dz) &= \int_0^1 2\pi \left[4 \sum_{j=1}^{\infty} |\hat{f}_j|^2 j^2 r^{2(j-1)} \right] r \, dr \\ &= 2\pi \sum_{j \in \mathbb{Z}} |\hat{f}_j|^2 |j| \end{aligned}$$

Thus, $f \in H_2^{\frac{1}{2}}$ if and only if

$$\int \left| \frac{d\mathcal{P}f(z)}{dz} \right|^2 m(dz) < \infty,$$

and

$$\langle f, g \rangle_{\frac{1}{2}} = \frac{1}{2\pi} \int \frac{d\mathcal{P}f(z)}{dz} \overline{\frac{d\mathcal{P}g(z)}{dz}} \, m(dz), \quad f, g \in H_2^{\frac{1}{2}}.$$

The form

$$\frac{1}{2\pi} \int \frac{dF(z)}{dz} \overline{\frac{dG(z)}{dz}} \, m(dz)$$

is (up to a constant multiple) nothing other than the Dirichlet form of Brownian motion on the unit disk. There has been much recent interest in studying the Dirichlet form of Brownian motion on such restricted domains (see, for example, [Gro99]).

The above connections suggest that we should be able to find a Brownian motion or Cauchy process as a limit of objects defined in terms of the eigenvalues of random unitary matrices. We have so far failed in this attempt.

1.4. **Dynkin's isomorphism.** We conclude with a brief review of Dynkin's isomorphism [Dyn80, Dyn83, Dyn84b, Dyn84a]. Assume that the continuous time Markov chain $\mathbf{X}$ considered above is *transient*. The Green's function $G(x, y) = -Q^{-1}(x, y)\pi(y)^{-1}$ is positive semi-definite and so can serve as a covariance of a mean zero Gaussian field $\mathbf{Y}$ indexed by $\mathcal{X}$. Note the parallel: roughly, our basic construction uses $-Q$ to construct a covariance while Dynkin used $-Q^{-1}$. Dynkin related properties of the Gaussian field to the underlying Markov chain. Among other things he showed that the best prediction of the field at a point a given its values at sites $B \subset \mathcal{X}$ is a linear combination of the observed values at B with weights the first hitting distribution of the chain started at a when it first hits B. We have a parallel version in Proposition 3.1.

Dynkin also proved the following distributional identity. Let

$$\ell_t^x = \pi(x)^{-1} \int_0^t \mathbf{1}\{X_s = x\} \, ds, \quad x \in \mathcal{X},$$

denote the "local time" process for the chain $\mathbf{X}$ with respect to the measure π. Suppose that on some probability space with expectation $\mathbb{P}$ we have a mean zero Gaussian field $\mathbf{Y} = \{Y_x\}_{x \in \mathcal{X}}$ with covariance G and an independent copy of the Markov chain $\mathbf{X}$. The chain $\mathbf{X}$ is started at $x \in \mathcal{X}$ and conditioned to die upon hitting $y \in \mathcal{X}$. Then, for any bounded Borel function $F : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}$ we have

$$\mathbb{E}\left[Y_x Y_y F\left(\frac{Y^2}{2} \right) \right] = \mathbb{E}\left[F\left(\frac{Y^2}{2} + \ell_\infty \right) \right] G(x, y)$$

Here, $\mathbf{Y}^2 = \{Y_x^2\}_{x \in \mathcal{X}}$ is the pointwise square of the Gaussian field $\mathbf{Y}$ and $\ell_\infty = \{\ell_\infty^x\}_{x \in \mathcal{X}}$.

In a sustained sequence of papers Marcus and Rosen [MR92d, MR92a, MR92c, MR92b, MR93] have studied symmetric Markov processes by using Dynkin's isomorphism. The isomorphism is tight enough so that refined knowledge of Gaussian fields (e.g. continuity of sample paths) can be carried over to develop fine properties of Markov processes (e.g., continuity of local time). Sheppard [She85] gave a proof of the Ray–Knight theorem on the Markovianity of local times of one-dimensional diffusions that used Dynkin's result and the obvious Markovianity of the associated Gaussian field. We do not see such depth for our construction, but find the parallels tantalising.

Dynkin's construction has been used in statistical applications by Ylvisaker [Ylv87]. He used the Gaussian fields as Bayesian priors for prediction and design problems. Dynkin's fields only have positive correlations while the fields we construct have negative correlations; using independent sums of both constructions may prove useful.

The relationship between a Markov chain and the Gaussian field with covariance given by the associated Green's function was discovered independently by several people. In physics, there is work of Symanzik [Sym69] followed by work of Brydges et al. [BFS82]. In statistics, Ylvisaker [Ylv87] gives references to Hammersley's [Ham67] work on *harnesses* as followed up by Williams [Wil73], Kingman [Kin85, Kin86] and Dozzi [Doz81, Doz89]. Variants of Dynkin's isomorphism have been established by Eisenbaum [Eis95], as well as by Marcus and Rosen in the papers cited above. Markov chain representations of fields other than Gaussian ones have also been studied: a recent paper with an extensive bibliography is [Bol01].

Here are two lesser known alternative appearances of this connection. Bhattacharya [Bha82] establishes general results that specialise in our finite setting to the following. Suppose that the chain $\mathbf{X}$ is ergodic. For f in the range of Q, $\frac{1}{\sqrt{T}}\int_0^T f(X_s)\,ds$ converges in distribution as $T \to \infty$ to a Gaussian field with covariance G

In a more applied context, various authors (see, for example, [BM74, Bor79, Bor01]) have considered optimal estimates of height in surveying problems. There are n points and estimates of height differences are available for some pairs. Forming an undirected graph with the pairs as edges (assumed connected), they find the best linear unbiased estimates of the true heights $\widehat{h}_x$. Assuming one true height, say at site z, is known, they show that

$$\mathrm{Cov}(\widehat{h}_x, \widehat{h}_y) = \frac{1}{q(y)} G_z(x, y)$$

with $G_z(x, y)$ the expected number of times y is hit starting at x by a discrete time reversible Markov chain constructed from edge weights $1/\sigma^2(x, y)$. Here $\sigma^2(x, y)$ is the variance of the (x, y)th height difference measurement and $q(y) = \sum_x (1/\sigma^2(y, x))$. The walk is killed when it hits z. If the measurement errors are assumed Gaussian, then $\widehat{h}_x$ is a Gaussian field with covariance given by the Green's function: Known asymptotics of $G_z(x, y)$ in planar grids can then be used to understand how the covariances fall off with separation.

2. Finite State Spaces

The following result gives a representation of the field $\mathbf{Z}$ in terms of independent Gaussian random variables and hence furnishes a simple way to simulate such a field.

Proposition 2.1. *Let $\mathbf{X}$ be a reversible Markov chain with finite state space $\mathcal{X}$. Form a graph with vertex set $\mathcal{X}$ by placing an undirected edge from x to y if $Q(x, y) > 0$. Choose an orientation for each edge e of the graph, that is, a function from the edge set to $\{\pm 1\}$. This orientation may be chosen arbitrarily but, of course, $\epsilon(x, y) = \epsilon(y, x)$. Associate a mean zero, variance $-\Sigma(x, y)$, Gaussian* *random variable $W(x,y)$ to each edge and a mean zero, variance $\kappa(x)$, Gaussian random variable $W(x)$ to each vertex, with all of these random variables being independent. Set*

$$Z_x = \sum_{y:Q(x,y)>0} \epsilon(x,y)W(x,y) + W(x),$$

then the Gaussian field $\{Z_x\}_{x\in\mathcal{X}}$ has covariance Σ.

Proof. The random variable Z_x has mean zero and variance

$$\begin{aligned}-\sum_{y:Q(x,y)>0} \Sigma(x,y) + \kappa(x) &= \sum_{y\neq x} \pi(x)Q(x,y) - \sum_{y} \pi(x)Q(x,y) \\ &= -\pi(x)Q(x,x) = \Sigma(x,x).\end{aligned}$$

Further, for $x \neq y$,

$$\mathbb{E}[Z_x Z_y] = \sum_{a,b:Q(x,a)Q(y,b)>0} \epsilon(x,a)\epsilon(y,b)\mathbb{E}[W(x,a)W(y,b)].$$

The sum is zero unless $Q(x,y) > 0$, and then it contains the single term

$$-\epsilon(x,y)\epsilon(y,x)\Sigma(x,y) = \Sigma(x,y),$$

as desired. □

Remark 2.2.
i) The construction is not limited to Gaussian variables. It gives a 2nd order field with the prescribed covariance for other uncorrelated choices of $W(x,y)$ and $W(x)$ with the variances set out in Proposition 2.1. Vertices with no edge between them are uncorrelated.
ii) If $\kappa = 0$ (that is, there is no killing), then $\sum_{x\in\mathcal{X}} Z_x = 0$.
iii) For simple random walk on $\mathbb{Z}_n$ (Example 1.2), choose a clockwise orientation and let $Z_j = W_j - W_{j-1}$ (indices $\mod n$) with W_j independent $N(0, \frac{1}{2n})$ variables.
iv) Conversely, if $\Sigma = (\Sigma(x,y))_{x,y\in\mathcal{X}}$ is a covariance matrix with positive diagonal entries, non-positive off-diagonal entries, and non-negative row sums, then Σ can be realized as a matrix arising from the Markov chain construction **in many different ways**. Just take π to be an arbitrary probability measure on the finite set $\mathcal{X}$ with $\pi(x) > 0$ for all x and put $Q(x,y) = -\pi(x)^{-1}\Sigma(x,y)$.
v) The representation of Proposition 2.1 can be thought of as a factorisation $\Sigma = AA'$, where A is a matrix that has a row for each element of $\mathcal{X}$ and a column corresponding to each of the random variables $W(x,y)$ and $W(x)$. This factorisation should be compared to the Karhunen–Loeve decomposition $\Sigma = (\Gamma D^{\frac{1}{2}})(\Gamma D^{\frac{1}{2}})' = \Gamma D \Gamma'$ where the columns of Γ are the normalised eigenvectors of Σ corresponding to non–zero eigenvalues and $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_k)$, $k = \operatorname{rank}\Sigma$, is the diagonal matrix that has these eigenvalues down the diagonal. Of course, the Karhunen–Loeve decomposition leads to another representation of the field $\mathbf{Z}$, namely

$$Z_x = \sum_{\lambda} \Gamma_{x\lambda} V_\lambda, \quad x \in \mathcal{X},$$

where $V_{\lambda_1}, \ldots, V_{\lambda_k}$ are independent mean zero Gaussian random variables, with V_λ having variance λ.

For the $n \times n$ discrete torus field in Example 1.1, $\operatorname{rank} \Sigma \sim n^2$ whereas the construction of Proposition 2.1 requires $\sim 4n^2$ independent Gaussian random variables. However, the computation of a particular Z_x requires only 4 of these variables, whereas the Karhunen–Loeve expansion requires the use of all $\operatorname{rank} \Sigma \sim n^2$ variables. Thus simulation the entire field $\mathbf{Z}$ requires $\sim 4n^2$ additions using our representation and the order of n^4 multiplications and additions for the Karhunen–Loeve representation. In this particular example, the fast Fourier transform can be used to cut the latter number of operations down to the order of $n^2 \log n$, but this improvement is not available for general chains that lack such group structure.

vi) Most constructions of Gaussian fields lead to positive correlations for near neighbours. Of course, this is often scientifically natural. However, fields in which all sites are negatively correlated could arise in settings where growth in one region deletes supplies from other regions. In situations like ours in which all sites are negatively correlated, there are constraints on the strength of the correlation. This is related to the well–known fact that n exchangeable random variables have correlations at least $-\frac{1}{n-1}$.

More generally, let $\mathbb{G} = (\mathcal{X}, E)$ be an undirected graph with vertex set $\mathcal{X}$ and edge set E. Suupose that the automorphism group G of $\mathbb{G}$ is such that given two edges $\{x', y'\}$ and $\{x'', y''\}$ there exists an element g of G such that $gx' = x''$ and $gy' = y''$. Let $\mathbf{Y} = \{Y_x\}_{x \in \mathcal{X}}$ be a mean zero Gaussian field that is invariant under the action of G. By renormalising if necessary, we may suppose that the common value of $\mathbb{E}[Y_x^2]$, $x \in \mathcal{X}$, is $d/(2|E|)$, where d is the common degree of the vertices of $\mathbb{G}$. Suppose the $\mathbb{E}[Y_x Y_y] \le 0$ for all $x, y \in \mathcal{X}$. Write ρ for the common value of $\mathbb{E}[Y_x Y_y]$ when $\{x, y\}$ is an edge. We have

$$\begin{aligned}
0 &\le \mathbb{E}\left[\left(\sum_{x \in \mathcal{X}} Y_x\right)^2\right] \\
&= 1 + \sum_{x \neq y} \mathbb{E}[Y_x Y_y] \\
&= 1 + 2|E|\rho + 2 \sum_{\{x,y\} \notin E} \mathbb{E}[Y_x Y_y] \\
&\le 1 + 2|E|\rho.
\end{aligned}$$

Thus $\rho \ge -1/(2|E|)$ with equality if and only if $\mathbb{E}[Y_x Y_y] = 0$ for all $\{x, y\} \notin E$. This extremal case when all the non-edge covariances are zero corresponds to our Markov chain construction with

$$\pi(x) = \frac{d}{2|E|}, \quad x \in \mathcal{X},$$

and

$$Q(x, y) = \begin{cases} -1, & \text{if } x = y, \\ \frac{1}{d}, & \text{if } x \neq y. \end{cases}$$

That is, the chain $\mathbf{X}$ exits from any state at rate 1, and when it exits it jumps to each of the d neighbouring states with equal probability. Examples 1.1 and 1.2 fit into this framework. The exchangeable case also fits into this framework, with the graph $\mathbb{G}$ being the complete graph.

3. Prediction and conditional distributions

The following result relates the dependence structure of the Gaussian field $\mathbf{Z} = \{Z_x\}_{x\in\mathcal{X}}$ to the properties of the occupation times of the original Markov chain $\mathbf{X}$. For B a proper subset of $\mathcal{X}$, suppose the field is observed at $x \in B$ and we want to predict it at $y \notin B$. The mean square optimal prediction is a linear combination of the observed values $\mathbf{Z}_B = \{Z_x\}_{x\in B}$. In order to describe the associated weights in terms of quantities for the chain $\mathbf{X}$, let

$$L_t(C) := \int_0^t \mathbf{1}\{X_s \in C\}\, ds, \quad t \geq 0,\ C \subseteq \mathcal{X},$$

denote the *occupation time field* for the chain $\mathbf{X}$. Write

$$\tau_t^C := \inf\{s \geq 0 : L_s(C) = t\}$$

and let $\mathbf{X}^C$ be the chain $\mathbf{X}$ *time–changed* according to $L_t(C)$: that is, $\mathbf{X}^C$ is a Markov chain with state space C such that the law of $\mathbf{X}^C$ starting at $c \in C$ is the same as that of $\{X_{\tau_t^C} : t \geq 0\}$ starting at c. Denote by

$$S := \inf\{t \geq 0 : X_t \neq X_0\}$$

the first time that the chain $\mathbf{X}$ leaves its initial state and by

$$R_D := \inf\{t \geq S : X_t \in D\}, \quad D \subseteq \mathcal{X},$$

the first time after leaving its initial state that the chain $\mathbf{X}$ enters the subset of states D.

Proposition 3.1. *Let $\mathbf{Z} = \{Z_x\}_{x\in\mathcal{X}}$ be a mean zero Gaussian field with covariance Σ given by (1.2). For a proper subset $B \subset \mathcal{X}$ and $A = \mathcal{X}\backslash B$, the conditional distribution of $\mathbf{Z}_A = \{Z_x : x \in A\}$ given $\mathbf{Z}_B = \{Z_x : x \in B\}$ is Gaussian with mean $\mathbb{E}[\mathbf{Z}_A \mid \mathbf{Z}_B] = M\mathbf{Z}_B$, where*

$$M(a,b) = \frac{\pi(a)Q(a,a)}{\pi(b)}\mathbb{E}_a\left[L_{R_A}(\{b\}),\ X_S \in B\right],$$

and covariance given by

$$-\pi(a')Q^A(a',a''), \quad a',a'' \in A,$$

where Q^A is the infinitesimal generator of the time–changed chain $\mathbf{X}^A$.

Proof. Classical theory gives that the conditional distribution of $\mathbf{Z}_A$ given $\mathbf{Z}_B$ is Gaussian with mean

$$\mathbb{E}[\mathbf{Z}_A \mid \mathbf{Z}_B] = \Sigma_{AB}\Sigma_{BB}^{-1}\mathbf{Z}_B.$$

Now $\Sigma = -\Pi Q$ where $\Pi := \operatorname{diag}(\pi(x))$, and thus

$$\Sigma_{AB}\Sigma_{BB}^{-1} = \Pi_{AA}Q_{AB}Q_{BB}^{-1}\Pi_{BB}^{-1}.$$

By direct expansion,

$$-(Q_{BB}^{-1})_{b'b''} = \mathbb{E}_{b'}\left[L_{T_A}(\{b''\})\right], \quad b',b'' \in B.$$

Moreover,

$$Q_{a,b} = -Q(a,a)\mathbb{P}_a\left[X_S = b\right], \quad a \in A,\ b \in B.$$

Therefore,

$$(\Sigma_{AB}\Sigma_{BB}^{-1})_{ab} = \frac{\pi(a)Q(a,a)}{\pi(b)}\mathbb{E}_a\left[L_{R_A}(\{b\}),\ X_S \in B\right].$$

Classical theory also gives that the covariance of the conditional distribution of $\mathbf{Z}_A$ given $\mathbf{Z}_B$ is

$$\Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA} = -\left(\Pi_{AA}Q_{AA} - \Pi_{AA}Q_{AB}Q_{BB}^{-1}Q_{BA}\right),$$

and it is straightforward to see that

$$Q^A = Q_{AA} + \Pi_{AA}Q_{AB}(-Q_{BB}^{-1})Q_{BA}.$$

□

The following result is immediate from Proposition 3.1.

Corollary 3.2. *In the notation of Proposition 3.1, construct a graph with vertex set $\mathcal{X}$ by placing an (undirected) edge between two vertices, $x \neq y$ if $\Sigma(x,y) < 0$. Fix a point $a \in A$. Say that a point $b \in B$ is* shielded from a *if every path from a to b passes through a point of $A \backslash \{a\}$. Write B_a for the set of points in B that are shielded from a. Then Z_a is conditionally independent of $\mathbf{Z}_B$ given $\mathbf{Z}_{B\backslash B_a}$ (equivalently, Z_a is conditionally independent of $\mathbf{Z}_{B_a}$ given $\mathbf{Z}_{B\backslash B_a}$). Moreover, if $B\backslash B_a \nsubseteq \tilde{B} \subseteq B$, then Z_a is not conditionally independent of $\mathbf{Z}_B$ given $\mathbf{Z}_{\tilde{B}}$*

Remark 3.3.
i) For large state spaces, Ylvisaker [Ylv87] suggested using simulation of the Markov chain as an aid to computing regression coefficients via Dynkin's construction. Proposition 3.1 can be used similarly.
ii) Note from the assumption of reversibility that
$$\frac{\pi(a)}{\pi(b)} = \frac{Q(b,a)}{Q(a,b)}, \quad a \in A,\, b \in B,$$
if the numerator and denominator on the right–hand side are positive. In this case
$$M(a,b) = \frac{Q(b,a)Q(a,a)}{Q(a,b)}\mathbb{E}_a\left[L_{R_A}(\{b\}),\ X_S \in B\right],$$
In any case, π only needs to be determined up to a constant.
iii) The coefficients $M(a,b)$ are always non-positive.
iv) Note the parallel with the form of the coefficients in Dynkin's construction described in the Introduction.

Example 3.4. Consider our running example of simple random walk on $\mathbb{Z}_n$ (Example 1.2). Choose a partition A, B of $\mathbb{Z}_n$ into two non-empty subsets. Fix a point $a \subset A$. If we have $a+1, a+2, \ldots, a+r \in B$ and $a+r+1 \in A$, then set $B_+ = \{a+1, a+2, \ldots, a+r\}$. Similarly, if we have $a-1, a-2, \ldots, a-\ell \in B$ and $a-(\ell+1) \in A$, then set $B_- = \{a-1, a-2, \ldots, a-\ell\}$. Of course, B_+ or B_- may be empty. Then, in the notation of Corollary 3.2, $B_a = B\backslash(B_+ \cup B_-)$.

More generally, standard probability calculations can be used to calculate the matrix M of Proposition 3.1 for various configurations. For example, suppose that $a = 0$ and $B_+ = \{1, 2, \ldots, r\}$. The probability that the random walk gets to $1 \le b \le r$ before returning to a or hitting another point of A is, by the classical Gambler's Ruin problem, $\frac{1}{2}\frac{1}{b}$. Moreover, the probability that the walk returns to b before hitting A is, again by Gambler's Ruin,

$$\frac{1}{2}\left(1 - \frac{1}{r+1-b}\right) + \frac{1}{2}\left(1 - \frac{1}{b}\right),$$

and so the distribution of the number of visits to b given that b is reached at all is the same as the number of trials up to and including the first success in Bernoulli trials with this return probability as the failure probability. It follows after a little algebra that

$$M(a,b) = -\left(1 - \frac{b}{r+1}\right).$$

4. Minimizing variances subject to constraints

In this section we will study the problem of minimizing the variance

$$\mathbb{E}\left[\left(\sum_x f(x) Z_x\right)^2\right]$$

of a linear combination $\sum_x f(x) Z_x$ under certain constraints on the coefficients. Here $\mathbf{Z} = \{Z_x\}_{x \in \mathcal{X}}$ is a mean zero Gaussian field with covariance Σ that has positive diagonal entries, non-positive off-diagonal entries, and non-negative row sums. We will also assume that Σ is irreducible in the sense that for $x \neq y$ we can find a sequence $x = z_0 \neq z_1 \neq \ldots \neq z_k = y$ such that $\Sigma(z_i, z_{i+1}) \neq 0$ for $0 \leq i \leq k-1$.

Proposition 4.1. *Suppose that Σ has at least one row sum positive. Given a proper subset $B \subset \mathcal{X}$, consider the problem of minimizing $\mathbb{E}[(\sum_x f(x) Z_x)^2]$ subject to $f(x) = 1$, $x \in B$. The minimum is achieved by*

$$f_B(x) := \mathbb{P}_x\{T_B < \infty\}$$

where $T_B = \inf\{t \geq 0 : X_t \in B\}$ for $\mathbf{X}$ a Markov chain with infinitesimal generator $Q(x,y) = -\pi(x)^{-1}\Sigma(x,y)$ for any probability vector π with positive entries.

Proof. This follows immediately from Theorem 4.3.3 of [FOT94] once we note that the condition that at least one row sum of Σ is positive is equivalent to the chain $\mathbf{X}$ being transient. □

Proposition 4.2. *Suppose that Σ has all row sums zero. Given a proper subset $B \subset \mathcal{X}$ and a probability vector π with positive entries, consider the problem of minimizing $\mathbb{E}[(\sum_x f(x) Z_x)^2]$ subject to $f(x) = 1$, $x \in B$, and $\sum_x f(x)\pi(x) = 0$. The minimum is given by*

$$(\mathbb{E}_\pi [T_B])^{-1},$$

where $T_B = \inf\{t \geq 0 : X_t \in B\}$ for $\mathbf{X}$ the Markov chain with infinitesimal generator $Q(x,y) = -\pi(x)^{-1}\Sigma(x,y)$. Moreover, the minimum is achieved by

$$f_B(x) = \frac{-\mathbb{E}_x[T_B] + \mathbb{E}_\pi[T_B]}{\mathbb{E}_\pi[T_B]}.$$

Proof. We first recall some standard facts. For $\alpha > 0$, let $\mathcal{E}_\alpha$ denote the inner product $\mathcal{E} + \alpha\langle \cdot \mid \cdot \rangle$. Write Cap_α for the corresponding capacity. Then

$$\mathrm{Cap}_\alpha(B) = \inf\{\mathcal{E}_\alpha(f,f) : f = 1\} = \mathcal{E}_\alpha(p_B^\alpha, p_B^\alpha), \tag{4.1}$$

where $p_B^\alpha(x) := \mathbb{P}^x[\exp(-\alpha T_B)]$ (see Theorem 4.2.5 of [FOT94] for the case $\alpha = 1$, the proof for general α involves just obvious changes). By Theorem 4.3.1 of [FOT94], p_B^α and $1 - p_B^\alpha$ are orthogonal with respect to $\mathcal{E}_\alpha$ and so

$$0 = \mathcal{E}_\alpha(p_B^\alpha, 1 - p_B^\alpha) = -\mathrm{Cap}_\alpha(B) + \alpha\langle p_B^\alpha \mid 1\rangle = -\mathrm{Cap}_\alpha(B) + \alpha\mathbb{E}_\pi[\exp(-\alpha T_B)],$$

where we have used the fact that $\mathcal{E}(1,1) = 0$. Thus

$$c_B^\alpha := \sum_x p_B^\alpha(x)\pi(x) = \mathbb{E}_\pi[\exp(-\alpha T_B)] = \alpha^{-1}\mathrm{Cap}_\alpha(B). \tag{4.2}$$

We have

$$\begin{aligned} c_B^\alpha &= \alpha^{-1}\inf\{\mathcal{E}_\alpha(f,f) : f = 1 \text{ on } B\} \\ &= \alpha^{-1}\inf\{\mathcal{E}_\alpha(f,f) : f = 1 \text{ on } B, \sum_x f(x)\pi(x) = c_B^\alpha\}. \end{aligned}$$

Thus, if we put

$$f_B^\alpha(x) := \frac{p_B^\alpha(x) - c_B^\alpha}{1 - c_B^\alpha},$$

then

$$\inf\{\mathcal{E}_\alpha(f,f) : f = 1 \text{ on } B, \sum_x f(x)\pi(x) = 0\} = \mathcal{E}_\alpha(f_B^\alpha, f_B^\alpha) = \frac{\alpha c_B^\alpha}{1 - c_B^\alpha} \tag{4.3}$$

after a little algebra. By the assumption of irreducibility, $\mathbb{P}_\pi\{T_B = \infty\} = 0$, and so

$$\lim_{\alpha\downarrow 0}\alpha^{-1}(1 - c_B^\alpha) = \lim_{\alpha\downarrow 0}\mathbb{E}_\pi\left[\int_0^{T_B} e^{-\alpha s}\,ds\right] = \mathbb{E}_\pi[T_B] \tag{4.4}$$

and the last term in (4.3) converges to $(\mathbb{E}_\pi[T_B])^{-1}$ as $\alpha \downarrow 0$.

Note, by the same argument that gave (4.4), that

$$\lim_{\alpha\downarrow 0}\alpha^{-1}(1 - p_B^\alpha(x)) = \mathbb{E}_x[T_B].$$

Therefore, in order to establish the claim of the proposition, it suffices to observe that

$$\begin{aligned} &\lim_{\alpha\downarrow 0}\inf\{\mathcal{E}_\alpha(f,f) : f = 1 \text{ on } B, \sum_x f(x)\pi(x) = 0\} \\ &\quad = \inf\{\mathcal{E}(f,f) : f = 1 \text{ on } B, \sum_x f(x)\pi(x) = 0\} \end{aligned}$$

and that

$$\lim_{\alpha\downarrow 0}\mathcal{E}_\alpha(f_B^\alpha, f_B^\alpha) = \mathcal{E}(f_B, f_B).$$

□

Acknowledgements. We thank Marc Coram for producing the graphics and Don Ylvisaker for a number of helpful conversations.

References

[BD58] A. Beurling and J. Deny. Espaces de Dirichlet. I. Le cas élémentaire. *Acta Math.*, 99:203–224, 1958.

[Bes74] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B*, 36:192–236, 1974. With discussion by D. R. Cox, A. G. Hawkes, P. Clifford, P. Whittle, K. Ord, R. Mead, J. M. Hammersley, and M. S. Bartlett and with a reply by the author.

[BFS82] D. Brydges, J. Fröhlich, and T. Spencer. The random walk representation of classical spin systems and correlation inequalities. *Comm. Math. Phys.*, 83(1):123–150, 1982.

[BG93] J. Besag and P. J. Green. Spatial statistics and Bayesian computation. *J. Roy. Statist. Soc. Ser. B*, 55(1):25–37, 1993.

[BGHM95] J. Besag, P. Green, D. Higdon, and K. Mengersen. Bayesian computation and stochastic systems. *Statist. Sci.*, 10(1):3–66, 1995. With comments and a reply by the authors.

[BH99] J. Besag and D. Higdon. Bayesian analysis of agricultural field experiments. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 61(4):691–746, 1999. With discussion and a reply by the authors.

[Bha82] R. N. Bhattacharya. On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrsch. Verw. Gebiete*, 60(2):185–201, 1982.

[BK95] J. Besag and C. Kooperberg. On conditional and intrinsic autoregressions. *Biometrika*, 82(4):733–746, 1995.

[BM74] K. Borre and P. Meissl. Strength analysis of leveling-type networks. An application of random walk theory. *Geodaet. Inst. Medd.*, (50):80, 1974.

[Bol01] E. Bolthausen. Random walk representations and entropic repulsion for gradient models. Preprint, 2001.

[Bor79] K. Borre. Error propagation in absolute geodetic networks—a continuous approach. In *Optimization of design and computation of control networks (Proc. Internat. Sympos., Sopron, 1977)*, pages 459–472. Akad. Kiadó, Budapest, 1979.

[Bor01] K. Borre. *Plane networks and their applications.* Birkhäuser Boston Inc., Boston, MA, 2001.

[DE01] P. Diaconis and S. N. Evans. Linear functionals of eigenvalues of random matrices. *Trans. Amer. Math. Soc.*, 353:2615–2633, 2001.

[DF81] P. Diaconis and D. Freedman. On the statistics of vision: the Julesz conjecture. *J. Math. Psych.*, 24(2):112–138, 1981.

[Doz81] M. Dozzi. Two-parameter harnesses and the Wiener process. *Z. Wahrsch. Verw. Gebiete*, 56(4):507–514, 1981.

[Doz89] M. Dozzi. *Stochastic processes with a multidimensional parameter.* Longman Scientific & Technical, Harlow, 1989.

[DS94] P. Diaconis and M. Shahshahani. On the eigenvalues of random matrices. *J. Appl. Probab.*, 31A:49–62, 1994.

[Dyn80] E. B. Dynkin. Markov processes and random fields. *Bull. Amer. Math. Soc. (N.S.)*, 3(3):975–999, 1980.

[Dyn83] E. B. Dynkin. Markov processes as a tool in field theory. *J. Funct. Anal.*, 50(2):167–187, 1983.

[Dyn84a] E. B. Dynkin. Gaussian and non-Gaussian random fields associated with Markov processes. *J. Funct. Anal.*, 55(3):344–376, 1984.

[Dyn84b] E. B. Dynkin. Polynomials of the occupation field and related random fields. *J. Funct. Anal.*, 58(1):20–52, 1984.

[Eis95] N. Eisenbaum. Une version sans conditionnement du théorème d'isomorphisms de Dynkin. In *Séminaire de Probabilités, XXIX*, number 1613 in Lecture Notes in Math., pages 266–289. Springer, Berlin, 1995.

[FOT94] M. Fukushima, Y. Oshima, and M. Takeda. *Dirichlet Forms and Symmetric Markov Processes.* Walter de Gruyter, Berlin, 1994.

[Gro99] L. Gross. Hypercontractivity over complex manifolds. *Acta Math.*, 182(2):159–206, 1999.

[GS89] J. Goodman and A. Sokal. Multigrid Monte–Carlo method: conceptual foundations. *Phys. Rev. D*, 40:2035–2071, 1989.

[Ham67] J. M. Hammersley. Harnesses. In *Proc. Fifth Berkeley Sympos. Mathematical Statistics and Probability (Berkeley, Calif., 1965/66), Vol. III: Physical Sciences*, pages 89–117. Univ. California Press, Berkeley, Calif., 1967.

[Joh97] K. Johansson. On random matrices from the compact classical groups. *Ann. of Math. (2)*, 145:519–545, 1997.

[Kin85] J. F. C. Kingman. Random variables with unsymmetrical linear regressions. *Math. Proc. Cambridge Philos. Soc.*, 98(2):355–365, 1985.

[Kin86] J. F. C. Kingman. The construction of infinite collections of random variables with linear regressions. *Adv. in Appl. Probab.*, (suppl.):73–85, 1986.

[Meh91] M. L. Mehta. *Random matrices.* Academic Press Inc., Boston, MA, second edition, 1991.

[MR92a] M. B. Marcus and J. Rosen. Moduli of continuity of local times of strongly symmetric Markov processes via Gaussian processes. *J. Theoret. Probab.*, 5(4):791–825, 1992.

[MR92b] M. B. Marcus and J. Rosen. Moment generating functions for local times of symmetric Markov processes and random walks. In *Probability in Banach spaces, 8 (Brunswick, ME, 1991)*, pages 364–376. Birkhäuser Boston, Boston, MA, 1992.

[MR92c] M. B. Marcus and J. Rosen. p-variation of the local times of symmetric stable processes and of Gaussian processes with stationary increments. *Ann. Probab.*, 20(4):1685–1713, 1992.

[MR92d] M. B. Marcus and J. Rosen. Sample path properties of the local times of strongly symmetric Markov processes via Gaussian processes. *Ann. Probab.*, 20(4):1603–1684, 1992.

[MR93] M. B. Marcus and J. Rosen. ϕ-variation of the local times of symmetric Lévy processes and stationary Gaussian processes. In *Seminar on Stochastic Processes, 1992 (Seattle, WA, 1992)*, pages 209–220. Birkhäuser Boston, Boston, MA, 1993.

[She85] P. Sheppard. On the Ray-Knight Markov property of local times. *J. London Math. Soc. (2)*, 31(2):377–384, 1985.

[ST87] H.-J. Schmeisser and H. Triebel. *Topics in Fourier Analysis and Function Spaces*. Wiley, Chichester, 1987.

[Sym69] K. Symanzik. Euclidean quantum field theory. In R. Jost, editor, *Local Quantum Theory*. Academic, New York, 1969.

[Wil73] D. Williams. Some basic theorems on harnesses. In *Stochastic analysis (a tribute to the memory of Rollo Davidson)*, pages 349–363. Wiley, London, 1973.

[Ylv87] D. Ylvisaker. Prediction and design. *Ann. Statist.*, 15(1):1–19, 1987.

E-mail address: `diaconis@math.Stanford.EDU`

DEPARTMENT OF MATHEMATICS, STANFORD UNIVERSITY, BUILDING 380, MC 2125, STANFORD, CA 94305, U.S.A.

E-mail address: `evans@stat.Berkeley.EDU`

DEPARTMENT OF STATISTICS #3860, UNIVERSITY OF CALIFORNIA AT BERKELEY, 367 EVANS HALL, BERKELEY, CA 94720-3860, U.S.A