Combinatorics, Probability and Computing (2014) 23, 140–160.
doi:10.1017/S0963548313000345

Concentration of Lipschitz Functionals of Determinantal and Other Strong Rayleigh Measures

ROBIN PEMANTLE[1]† and YUVAL PERES[2]

[1]Department of Mathematics, University of Pennsylvania, 209 South 33rd Street, Philadelphia, PA 19104, USA
(e-mail: pemantle@math.upenn.edu)
[2]Microsoft Research, 1 Microsoft Way, Redmond, WA 98052, USA
(e-mail: peres@microsoft.com)

Received 4 August 2011; revised 10 May 2013; first published online 19 September 2013

Let $\{X_1, \ldots, X_n\}$ be a collection of binary-valued random variables and let $f : \{0,1\}^n \to \mathbb{R}$ be a Lipschitz function. Under a negative dependence hypothesis known as the *strong Rayleigh* condition, we show that $f - \mathbb{E}f$ satisfies a concentration inequality. The class of strong Rayleigh measures includes determinantal measures, weighted uniform matroids and exclusion measures; some familiar examples from these classes are generalized negative binomials and spanning tree measures. For instance, any Lipschitz-1 function of the edges of a uniform spanning tree on vertex set V (*e.g.*, the number of leaves) satisfies the Gaussian concentration inequality

$$\mathbb{P}(f - \mathbb{E}f \geqslant a) \leqslant \exp\left(-\frac{a^2}{8\,|V|}\right).$$

We also prove a continuous version for concentration of Lipschitz functionals of a determinantal point process.
2010 *Mathematics subject classification*: 60G55

1. Introduction

Our goal in this paper is to prove concentration inequalities for Lipschitz functions of certain collections of negatively dependent binary-valued random variables. To illustrate our general methods we state our main result in a special case that was motivated by a question of E. Mossel (personal communication).

Theorem 1.1. *Let $G = (V, E)$ be a finite connected graph, let $\mathbb{P}$ be the uniform measure on the spanning trees of G, and for $e \in E$ let X_e be the indicator function of the event that e is in the chosen spanning tree. Let $f : \{0,1\}^E \to \mathbb{R}$ be any function with Lipschitz constant* 1.

† Research supported in part by NSF grant DMS-0905937

Then

$$\mathbb{P}(f - \mathbb{E}f \geqslant a) \leqslant \exp\left(-\frac{a^2}{8\,|V|}\right).$$

For example, we might take f to be one half the number of vertices whose degree in the random tree is odd. This result is a consequence of more general results stated in Sections 3 and 5.

1.1. Classical concentration inequalities

Let $\{X_n : n \geqslant 1\}$ be independent Bernoulli random variables with respective means $\{p_n\}$. Let $S_n := \sum_{k=1}^n X_k$ denote the partial sums, $\mu_n := \mathbb{E}S_n = \sum_{k=1}^n p_k$ denote the means and $V_n := \sum_{k=1}^n p_k(1-p_k)$ denote the variance of S_n. The simple and well-known one-sided tail estimate for S_n is the classical Gaussian bound

$$\mathbb{P}(S_n - \mu_n \geqslant a) \leqslant \exp\left(-\frac{2a^2}{n}\right). \tag{1.1}$$

Replacing X_n with $1 - X_n$ gives the two-sided bound

$$\mathbb{P}(|S_n - \mu_n| \geqslant a) \leqslant 2\exp\left(-\frac{2a^2}{n}\right). \tag{1.2}$$

The bound (1.1) may be found, among other places, in [22, Corollary 5.2]. The references given there include [13, (2.3)] as well as [4], which proves the result for identically distributed variables.

When p_n and $1 - p_n$ are bounded away from zero, the variance of S_n is of order n and this kind of bound is the best one can expect. However, when $n \gg \mu_n$, one might hope for uniformity in n via bounds in which the exponent depends on μ_n and not on n. For example, if $\max_{j\leqslant n} p_j$ is small then S_n is well approximated by a Poisson variable with mean μ_n. The upper tail of a Poisson is not as thin as a Gaussian, being $\exp[-\Theta(a\log(a/\mu))]$ rather than $\exp[-\Theta(a^2/\mu)]$. The bound

$$\mathbb{P}(S_n \geqslant a + \mu) \leqslant e^a \left(\frac{\mu}{a+\mu}\right)^{a+\mu} \leqslant \exp\left[-\frac{a^2}{2(a+\mu)}\right] \tag{1.3}$$

is proved in [13, Theorem 1] and asymptotically matches the Poissonian upper tail.

1.2. Generalizations

Our aim is to generalize (1.1) or its Poissonian version (1.3) in two ways. Instead of S_n we consider arbitrary Lipschitz functions of $X_1, \ldots, X_n$, and instead of independent Bernoullis we consider a more general negatively dependent collection of binary random variables. We will give a number of applications, but before this, we briefly discuss what is known about each of the two generalizations separately.

For the first generalization, let $\mathcal{B}_n$ denote the rank-n Boolean lattice $\{0,1\}^n$ and let $f : \mathcal{B}_n \to \mathbb{R}$ be Lipschitz with respect to the Hamming distance. Replacing f by f/c if necessary, we will lose no generality in assuming our Lipschitz functions to have Lipschitz constant 1, and we do so hereafter; thus $|f(x) - f(x')| \leqslant 1$ whenever x and x'

https://doi.org/10.1017/S0963548313000345 Published online by Cambridge University Press

are two strings differing in only one position. When $\mathbb{P}$ is a product measure, a well-known generalization of (1.1) [22] is

$$\mathbb{P}(f - \mathbb{E}f \geqslant a) \leqslant e^{-2a^2/n}. \tag{1.4}$$

For the second generalization, we say that a collection of random variables $\{X_j\}$ in $\{0,1\}$ is *negatively cylinder-dependent* if

$$\mathbb{P}(X_j = 1 \text{ for all } j \in S) \leqslant \prod_{j \in S} p_j \tag{1.5}$$

and

$$\mathbb{P}(X_j = 0 \text{ for all } j \in S) \leqslant \prod_{j \in S} (1 - p_j). \tag{1.6}$$

Negative cylinder dependence implies the inequalities (1.1)–(1.2) (see, *e.g.*, [23, Theorem 3.4] with $\lambda = 1$). Lyons [21, Section 6] lists extensions and applications including one to balls in bins [5] and one to determinantal measures [26].

It is not known whether these two generalizations can be combined. The random variables $\{X_j : 1 \leqslant j \leqslant n\}$ are said to be *negatively associated* if $\mathbb{E}fg \leqslant (\mathbb{E}f)(\mathbb{E}g)$ for every pair f, g of increasing functions on $\{0,1\}^n$ such that $f(X_1, \ldots, X_n)$ depends only on the values $\{X_i : i \in S\}$ and $g(X_1, \ldots, X_n)$ depends only on the values $\{X_i : i \notin S\}$, for some subset $S \subseteq \{1, \ldots, n\}$. By induction, this implies the weaker property of negative cylinder dependence. E. Mossel (personal communication, 2009) asked us whether the following holds.

Conjecture 1.2. *Let $X_1, \ldots, X_n$ be negatively associated binary-valued random variables. Let $f : \{0,1\}^n \to \mathbb{R}$ be Lipschitz-1 and denote $f_n := f(X_1, \ldots, X_n)$. Then (1.4) holds with the bound $\exp(-2a^2/n)$ replaced by $c_0 \exp(-c\, a^2/n)$ for some positive constants c_0 and c.*

To see why the exponent must be weakened, consider the example of Bernoulli random variables $X_1, \ldots, X_n$ with n even, $\{X_1, \ldots, X_{n/2}\}$ independent with mean $1/2$, and $X_{n/2+j} = 1 - X_j$ for $1 \leqslant j \leqslant n/2$. These are negatively associated, and yet the Lipschitz-1 function

$$f := \sum_{j=1}^{n/2} X_j - \sum_{j=n/2+1}^{n} X_j$$

has tail probabilities on the order of $e^{-a^2/n}$. It is possible that this is the worst example and that the conjecture holds with $c = 1$, but a resolution of the conjecture would be interesting even without the optimal value of c. A recent paper [7] appears to settle this conjecture and more, but the relevant result in that paper, Theorem 2, is not correct, the proof therein failing at equation (6).

Recent investigations of negative dependence properties indicate that negative association may not be sufficiently robust to use as a hypothesis in this context. The problem was posed in [24] to find a more useful negative dependence property; this was answered in [2], who showed that the *strong Rayleigh property* implies negative association and many other desirable consequences, and is stable under probabilistic operations such as conditioning, symmetrizing and reweighting.

Our main result implies that Conjecture 1.2 holds with $c = 1/8$ if one assumes the strong Rayleigh property rather than just negative association. The strong Rayleigh property is known to hold for most standard examples in which negative association is known to hold, so this gives up little generality, and moreover the strong Rayleigh property is usually easier to check than is negative association. Indeed, for some of the measures described below, the only way we know they are negatively associated is by establishing the strong Rayleigh property. Several classes of measures satisfying the strong Rayleigh property are:

- determinantal measures and point processes,
- Bernoullis conditioned on the sum,
- measures obtained by running exclusion dynamics from a deterministic starting state (or more generally, exclusion with birth and death).

An overview of the rest of the paper is as follows. In the next section we introduce the strong Rayleigh property and discuss its consequences. One important consequence for us will be the *stochastic covering property*, which is all we use to derive our basic concentration inequality. In Section 3 we state our results, and these are proved in Section 4. Section 5 contains a number of applications.

2. Strong Rayleigh property, stochastic covering property, and other negative dependence conditions

Let $[n]$ denote $\{1,\ldots,n\}$ and let $\mathcal{B}_n := \{0,1\}^n$ denote the Boolean lattice of rank n, with coordinatewise partial order. The function $N : \mathcal{B}_n \to \mathbb{Z}^+$ will be used throughout to denote the counting function defined by $N(\omega) := \sum_{j=1}^n \omega_j$. A measure ν on $\mathcal{B}_n$ is said to be *k-homogeneous* if ν is supported on the set of $\{\omega : N(\omega) = k\}$. The probability measure ν on $\mathcal{B}_n$ is said to be *negatively associated* if $\int fg\,d\nu \leqslant (\int f\,d\nu)(\int g\,d\nu)$ for every pair of non-negative monotone functions f and g such that, for some set $S \subseteq [n]$, the function f depends only on coordinates $\{\omega_j : j \in S\}$ while the function g depends only on coordinates $\{\omega_j : j \in S^c\}$.

The *strong Rayleigh condition* is said to hold for a measure $\mathbb{P}$ on $\mathcal{B}_n$ if the generating function

$$\sum_{\omega\in\mathcal{B}_n} \mathbb{P}(\omega) \prod_{j=1}^n z_j^{\omega_j}$$

has no roots $(z_1,\ldots,z_n)$ all of whose coordinates lie in the (strict) upper half-plane. This and many consequences are given in [2], including (implicitly) the *stochastic covering property* (see Proposition 2.1), which was conjectured [24, Conjecture 9] to follow from something a little weaker. Some of the relevant implications are summarized in Figure 1.

The definition of the stochastic covering property requires a few preliminary definitions. Recall that a measure ν on a partially ordered set is said to *stochastically dominate* a measure ρ, denoted $\nu \succeq \rho$, if $\nu(A) \geqslant \rho(A)$ for every upwardly closed set A. An equivalent

https://doi.org/10.1017/S0963548313000345 Published online by Cambridge University Press

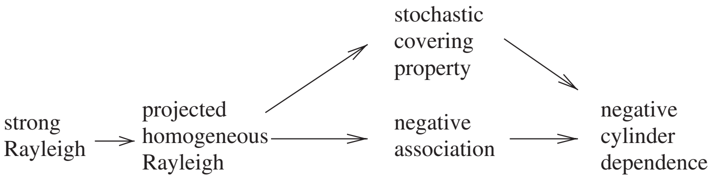

Figure 1. Relations among negative dependence properties.

condition is that there exists a coupling, that is a measure Q on $\mathcal{B}_n \times \mathcal{B}_n$ with respective marginals ν and ρ, supported on the set $\{(x, y) : x \geqslant y\}$. If $\mathbb{P}$ is a measure on $\mathcal{B}_n$ making the coordinate variables $\{X_i : 1 \leqslant i \leqslant n\}$ negatively associated, an immediate consequence of negative association is that the conditional measure $(\mathbb{P} \mid X_n = 0)$ on $\mathcal{B}_{n-1}$ stochastically dominates the conditional measure $(\mathbb{P} \mid X_n = 1)$.

We say that the probability measure ν on $\mathcal{B}_n$ *stochastically covers* another probability measure ρ if there is a measure on $\mathcal{B}_n^2$ with first marginal ν and second marginal ρ (in other words, a coupling) supported on the set of pairs (x, y) for which $x = y$ or x covers y in the coordinatewise partial order; here x is said to cover y when $x > y$ but there is no z such that $x > z > y$. We denote the covering relation in $\mathcal{B}_n$ by $x \gtrdot y$, and one measure covering another by $\nu \rhd \rho$. Stochastic covering is strictly stronger than stochastic domination, and may be thought of as 'stochastic domination, but by at most 1'.

Stochastic covering combines stochastic ordering with closeness in the so-called L^∞-transportation metric, defined on probability measures on a given metric space as follows: $d_\infty(\mu, \nu)$ is the least ρ such that there is a coupling of μ and ν supported on the set $\{(x, y) : |x - y| \leqslant \rho\}$. Thus $\mu \rhd \nu$ implies $d_\infty(\mu, \nu) \leqslant 1$. This is useful because if $||f||_{\mathrm{Lip}}$ denotes the Lipschitz norm on Lipschitz functions, then

$$\left| \int f \, d\mu - \int f \, d\nu \right| \leqslant ||f||_{\mathrm{Lip}} \, d_\infty(\mu, \nu). \tag{2.1}$$

Suppose that $x \geqslant y$ and we compare the conditional laws $\mathbb{P}_x := (\mathbb{P} \mid X_j = x_j, j \in S)$ and $\mathbb{P}_y := (\mathbb{P} \mid X_j = y_j, j \in S)$ on the remaining coordinates, that is, as laws on $\{0, 1\}^{S^c}$. If $\mathbb{P}$ and all its conditionalizations are negatively associated, it follows that $\mathbb{P}_x \preceq \mathbb{P}_y$.

Definition (stochastic covering property). We say that a probability measure ν on $\mathcal{B}_n$ has the stochastic covering property if, for every $S \subseteq \{1, \ldots, n\}$ and for every $x, y \in \{0, 1\}^S$ with $x \gtrdot y$, the conditional law $(\nu \mid X_j = x_j, j \in S)$ is covered by the conditional law $(\nu \mid X_j = y_j, j \in S)$.

In [2, Theorem 4.2] it was shown that strong Rayleigh property implies the *projected homogeneous Rayleigh* property (PHR), meaning that the measure can be embedded as the first n coordinates of a homogeneous measure ν' on $\mathcal{B}_m$, for some $m \geqslant n$, that has the ordinary Rayleigh property; the ordinary Rayleigh property is that the partial derivatives of the generating function $F(z_1, \ldots, z_n) := \mathbb{E} \prod_{j=1}^n z_j^{X_j}$ satisfy $F_i F_j \geqslant F_{ij} F$ at any point with positive real coordinates. We record two further consequences.

https://doi.org/10.1017/S0963548313000345 Published online by Cambridge University Press

Proposition 2.1. *PHR (and hence strong Rayleigh) implies the stochastic covering property.*

Proof. PHR implies negative association of all conditionalizations (CNA) [2, Theorem 4.10]; the homogeneous extension ν' witnessing the PHR property is also PHR hence also CNA. By negative association, if $x > y$ then $(\nu' \mid X_j = y_j, j \in S) \succeq (\nu' \mid X_j = x_j, j \in S)$, when viewed as measures on the coordinates in $[m] \setminus S$. Because ν' is homogeneous, the coupling that witnesses this $\succeq$ relation in fact witnesses the relation $\triangleright$. Restricting to $[n] \setminus S$ we see that $(\nu \mid X_j = y_j, j \in S) \triangleright (\nu \mid X_j = x_j, j \in S)$. $\square$

Proposition 2.2 ([2, Theorem 4.19]). *Let $\mathbb{P}$ be a strong Rayleigh measure on $\mathcal{B}_n$. Let $\mathbb{P}_k$ denote $\mathbb{P}$ conditioned on $N = k$. Then for every $0 \leqslant k \leqslant n-1$ such that $\mathbb{P}(N = k)$ and $\mathbb{P}(N = k+1)$ are both non-zero, we have the covering relation $\mathbb{P}_{k+1} \triangleright \mathbb{P}_k$.*

3. Results

The chief consequence of the strong Rayleigh property that we use to prove concentration inequalities is the stochastic covering property. Although all of our examples so far of measures with the SCP are in fact strong Rayleigh, we note that this may not be the case in the future, and with this in mind, we state a result that uses only the SCP.

Theorem 3.1 (homogeneity and SCP imply Gaussian concentration). *Let $\mathbb{P}$ be a k-homogeneous probability measure on $\mathcal{B}_n$ satisfying the SCP. Let f be a Lipschitz-1 function on $\mathcal{B}_n$. Then*

$$\mathbb{P}(f - \mathbb{E}f \geqslant a) \leqslant \exp\left(-\frac{a^2}{8k}\right).$$

Remarks. **(i)** Replacing f with $-f$ gives immediate two-sided bounds:

$$\mathbb{P}(|f - \mathbb{E}f| > a) \leqslant 2\exp\left(\frac{-a^2}{8k}\right). \tag{3.1}$$

(ii) Replacing every X_i by $1 - X_i$ if necessary, we may assume without loss of generality that $k \leqslant n/2$, whence

$$\mathbb{P}(f - \mathbb{E}f > a) \leqslant \exp\left(\frac{-a^2}{4n}\right). \tag{3.2}$$

For strong Rayleigh measures that are not necessarily homogeneous, we have the following result.

https://doi.org/10.1017/S0963548313000345 Published online by Cambridge University Press

Theorem 3.2 (Gauss–Poisson bounds for general strong Rayleigh measures). *Let $\mathbb{P}$ be strong Rayleigh with mean $\mu = \mathbb{E}N$. Let $f : \mathcal{B}_n \to \mathbb{R}$ be Lipschitz-1. Then*

$$\mathbb{P}(f - \mathbb{E}f > a) \leqslant 3\exp\left(-\frac{a^2}{16(a+2\mu)}\right),$$
$$\mathbb{P}(|f - \mathbb{E}f| > a) \leqslant 5\exp\left(-\frac{a^2}{16(a+2\mu)}\right).$$

Remark. Because $a, \mu \leqslant n$, the denominator in these inequalities may be replaced by $48n$.

Continuous versions

Continuous versions of these results may be stated in terms of point processes, which we now briefly review. Formally, a point process on a space S is a random counting measure on S. In other words, a point process is a map $\mathcal{Z}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in the space of counting measures on S, a counting measure being one that takes only integer values or $+\infty$. Intuitively, one envisions the sample counting measure $\mathcal{Z}(\omega)$ as a set of points such that the sum of delta functions at these points is the sample counting measure.

If the number k of points in the support of $\mathcal{Z}$ is deterministic, we may dispense with much of the formalism by ordering the points in the support of $\mathcal{Z}$ uniformly at random and identifying the process $\mathcal{Z}$ with the resulting exchangeable probability law on sequences of length k in S. Notationally, if $\mathcal{Z}$ is a k-homogeneous point process on $\mathbb{R}^d$ with law $\mathbb{P}$, we denote by $\mathbb{P}_\uparrow$ the corresponding exchangeable law on $(\mathbb{R}^d)^k$. For $1 \leqslant j \leqslant k$, we use X_j to denote the 'jth random point', that is, the jth coordinate function on $(\mathbb{R}^d)^k$. The following sampling algorithm for any k-homogeneous point process is almost trivial once one identifies $\mathcal{Z}$ with $\mathbb{P}_\uparrow$, and yet it is a generalization of an algorithm previously proved only in the case of determinantal point process in [15, Proposition 4.4.3].

Lemma 3.3 (sampling in k steps). *Let $\mathcal{Z}$ be a k-homogeneous point process on a standard Borel space S and let $\mathbb{P}_\uparrow$ be the corresponding exchangeable measure on S^k. Then for $0 \leqslant j < k$ there are regular conditional distributions $Q_{x_1,\dots,x_j}$ for the law of X_{j+1} given $X_1 = x_1, \dots, X_j = x_j$ such that the following procedure samples from $\mathbb{P}_\uparrow$:*

> *Sample X_1 from $Q_\emptyset$.*
>
> *Recursively, conditional on $X_1 = x_1, \dots, X_j = x_j$, sample X_{j+1} from $Q_{x_1,\dots,x_j}$.*

When S is finite, let R denote the random set $\{X_1, \dots, X_n\}$. Then the law $Q_{x_1,\dots,x_j}$ is equal to $1/(k-j)$ times the conditional intensity measure of $R \setminus \{x_1, \dots, x_j\}$ given $x_1, \dots, x_j \in R$.

Proof. Any standard Borel space admits regular conditional distributions [6, Theorem 4.1.6]. The sampling algorithm essentially restates the definition of regular conditional probabilities for sequential sampling. Because $\mathbb{P}_\uparrow$ is exchangeable, conditioning on $X_{i_1} = x_1, \dots, X_{i_j} = x_j$ gives the same exchangeable measure on the sequence of remaining elements of R for any $i_1, \dots, i_j$. Thus, for any x other than $x_1, \dots, x_j$, we have

$$\mathbb{P}(x \in R \mid x_1 \in R, \dots, x_j \in R) = (k-j)\mathbb{P}(X_{j+1} = x \mid X_1 = x_1, \dots, X_j = x_j)$$

which is the final conclusion. □

https://doi.org/10.1017/S0963548313000345 Published online by Cambridge University Press

Remark. In the case of a measure on a finite set of size n, the main point of this sampling scheme is to sample in k steps rather than n steps, so as better to control the Azuma martingale. But also, sequential conditioning on $x \in R$ can be easy to compute. For example, conditioning on an edge being in a spanning tree replaces the original graph by a contraction along that edge.

Intuitively, there is a stochastic covering property for point processes defined to hold when conditioning on the presence of a point depresses the process everywhere else but by at most one point. To make this definition precise, start by extending the notion of one measure stochastically covering another to point processes. We say the point process $\mathcal{Z}$ stochastically covers the point process $\mathcal{W}$ if there is a coupling of these two laws on counting measures supported on pairs (μ, ν) such that $\mu = \nu$ or $\mu = \nu + \delta_x$ for some x. Metrizing the space of finite counting measures on S by the total variation distance, we see as before that if $\mathcal{Z} \rhd \mathcal{W}$ then $d_\infty(\mathcal{L}(\mathcal{Z}), \mathcal{L}(\mathcal{W})) \leqslant 1$.

Next, given a k-homogeneous point process $\mathcal{Z}$ on a space S, we let $\mathcal{Z}_{x_1,\dots,x_j}$ denote the $(k-j)$-homogeneous point process whose law is the law of $\{X_{j+1}, \dots, X_k\}$ when sampling according to the procedure in Lemma 3.3 conditional on $X_1 = x_1, \dots, X_j = x_j$. The k-homogeneous point process $\mathcal{Z}$ is said to have the *stochastic covering property* if

$$\mathcal{Z}_{x_1,\dots,x_j} \rhd \mathcal{Z}_{x_1,\dots,x_{j+1}}$$

for all choices of $x_1, \dots, x_{j+1}$. Note that the left-hand side is $(k-j)$-homogeneous while the right-hand side is $(k-j-1)$-homogeneous.

Theorem 3.4. *Let $\mathcal{Z}$ be a k-homogeneous point process on a standard Borel space S and let f be a Lipschitz-1 function (with respect to the total variation distance) on counting measures with total mass k on $\mathbb{R}^d$. If $\mathcal{Z}$ has the SCP, then*

$$\mathbb{P}(f - \mathbb{E}f \geqslant a) \leqslant \exp\left(-\frac{a^2}{8k}\right).$$

For point processes that are not homogeneous, as in the discrete case, we require more than the SCP. Rather than defining a notion of strong Rayleigh here, we will stick to the case of determinantal point processes, this being where all of our examples arise; see Section 6 for definitions.

Theorem 3.5. *Let $\mathcal{Z}$ be a determinantal point process with $\mathbb{E}N = \mu < \infty$. Let f be a Lipschitz-1 function on finite counting measures. Then*

$$\mathbb{P}(f - \mathbb{E}f \geqslant a) \leqslant 3\exp\left(-\frac{a^2}{16(a+2\mu)}\right),$$
$$\mathbb{P}(|f - \mathbb{E}f \geqslant a) \leqslant 5\exp\left(-\frac{a^2}{16(a+2\mu)}\right).$$

4. Proofs

4.1. The classical proofs

To prove bounds such as (1.1), one obtains an upper bound for $\mathbb{E}e^{\lambda S_n}$ and then applies Markov's inequality, choosing λ optimally. Underlying the bounds on $\mathbb{E}e^{\lambda S_n}$ are corresponding bounds for compensated increments. Let Δ denote a variable with mean zero. Three classical exponential bounds are as follows:

$$|\Delta| \leqslant 1 \Rightarrow \mathbb{E}e^{\lambda\Delta} \leqslant e^{\lambda^2/2}, \tag{4.1}$$

$$\Delta \in [r, s] \Rightarrow \mathbb{E}e^{\lambda\Delta} \leqslant e^{\lambda^2(s-r)^2/8}, \tag{4.2}$$

$$\Delta \in [r, s] \Rightarrow \mathbb{E}e^{\lambda\Delta} \leqslant \exp\big[(e^\lambda - 1 - \lambda)\,|rs|\big], \tag{4.3}$$

when $|r - s| \leqslant 1$. These are used together with the following two special cases of Markov's inequality:

$$\mathbb{E}e^{\lambda X} \leqslant e^{c\lambda^2/2} \Longrightarrow \mathbb{P}(X \geqslant a) \leqslant e^{-a^2/(2c)}, \tag{4.4}$$

$$\mathbb{E}e^{\lambda X} \leqslant e^{b\,(e^\lambda - \lambda - 1)} \Longrightarrow \mathbb{P}(X > a) \leqslant e^a\left(\frac{b}{a+b}\right)^{a+b} \leqslant \exp\left[-\frac{a^2}{2(a+b)}\right]. \tag{4.5}$$

These inequalities have appeared many times in the literature. Inequalities (4.1) and (4.4) constitute the classical Azuma–Hoeffding inequality and imply

$$\mathbb{E}e^{\lambda(S_n - \mu_n)} \leqslant e^{\lambda^2 n/2}, \tag{4.6}$$

$$\mathbb{P}(S_n - \mu_n \geqslant a) \leqslant e^{-a^2/(2n)}. \tag{4.7}$$

This is valid for any martingale with differences bounded by 1; an exposition can be found in [1, Theorem 7.2.1]. The improvement to (4.2) is present already in [13], though the exposition in [22] is clearer (see Lemma 5.8 therein). When the increments of $S_n - \mu_n$ are compensated Bernoullis, one may take $b - a = 1$ rather than 2, resulting in an improvement by a factor of 4 in the exponent,

$$\mathbb{E}e^{\lambda(S_n - \mu_n)} \leqslant e^{\lambda^2 n/8}, \tag{4.8}$$

which together with (4.4) yields (1.1). Finally, (4.3) and induction yield

$$\mathbb{E}e^{\lambda(S_n - \mu_n)} \leqslant e^{(e^\lambda - \lambda - 1)\,V_n} \leqslant e^{(e^\lambda - \lambda - 1)\,\mu_n}, \tag{4.9}$$

where $V_n := \sum_{k=1}^n p_k(1 - p_k)$ is the variance of S_n; together with (4.5) this implies (1.3). These results appear in [9, (1.3)–(1.6)], for example.

To prove the generalization to Lipschitz functions, let

$$M_k := \mathbb{E}\big(f(X_1, \ldots, X_n)\,|\,X_1, \ldots, X_k\big) - \mathbb{E}f(X_1, \ldots, X_n).$$

It is immediate that $\{M_k\}$ is a martingale and that, conditional on $X_1, \ldots, X_{k-1}$, the two possible values of M_k differ by at most 1. Hence, conditional on $X_1, \ldots, X_{k-1}$, the increment $\Delta_k := M_k - M_{k-1}$ is constrained to an interval of length at most 1. Applying (4.2) then yields (1.4).

The extension of inequalities (1.1)–(1.2) to negatively cylinder-dependent random variables is established by examining the power series for $e^{\lambda S_n}$. This may be expanded

https://doi.org/10.1017/S0963548313000345 Published online by Cambridge University Press

into positive sums of expectations of products of powers of the variables $\{X_j : 1 \leqslant j \leqslant n\}$. Negative cylinder dependence implies that these are bounded from above by the corresponding products of expectations. Therefore, (4.8) and (4.9) hold when the assumption of independence is replaced by negative cylinder dependence, whence the probability inequalities (1.1) and (1.3) hold as well. This and more is shown in [23, Theorem 3.4], specializing their more general negative cylinder property to $\lambda = 1$. We remark that only the first inequality (1.5) in the definition of negative cylinder dependence is used to obtain bounds on $\mathbb{E}e^{\lambda S_n}$ for $\lambda > 0$, which suffices for the upper tail bounds. Lower tail bounds require these inequalities for $\lambda < 0$, for which the second inequality (1.6) is required.

4.2. Proof of Theorems 3.1 and 3.4

Theorem 3.1 is a special case of Theorem 3.4. This is because any probability measure μ on $\mathcal{B}_n$ may be viewed as the law of a point process on the n element set $[n]$, where the random counting measure $\mathcal{Z}(\omega)$ is defined by $\mathcal{Z}(\omega)(S) = \sum_{j \in S} \omega_j$. Informally, the points of the process are the coordinates of the ones in the sample ω. With this interpretation, the SCP on $\mathcal{B}_n$ is inherited from the SCP for the point process $\mathcal{Z}$, whence Theorem 3.4 with $S = \mathbb{R}$ (or any other standard Borel space containing $[n]$) implies Theorem 3.1. It remains to prove Theorem 3.4.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which is constructed the generalized sampling scheme described in Lemma 3.3. Let $\mathcal{F}_j := \sigma(X_1, \ldots, X_j)$ and let

$$M_j := \mathbb{E}(f \mid \mathcal{F}_j) - \mathbb{E}f \tag{4.10}$$

denote the martingale of sequential revelation. Applying the method of bounded differences is now mostly a matter of bookkeeping. At a sample point where $X_i = x_i, 1 \leqslant i \leqslant k$, the quantity M_j may be written as the integral of f against the law of the point process

$$\mathcal{Z}_{x_1,\ldots,x_j} + \sum_{i=1}^{j} \delta_{x_i}.$$

By the SCP, we have $\mathcal{Z}_{x_1,\ldots,x_j} \rhd \mathcal{Z}_{x_1,\ldots,x_{j+1}}$, whence

$$d_\infty\left(\mathcal{Z}_{x_1,\ldots,x_j} + \sum_{i=1}^{j} \delta_{x_i} \ , \ \mathcal{Z}_{x_1,\ldots,x_{j+1}} + \sum_{i=1}^{j+1} \delta_{x_i} \right) \leqslant 2.$$

By the Lipschitz assumption on f, it follows that $|M_{j+1} - M_j| \leqslant 2$. We now apply the basic Azuma–Hoeffding inequality (4.7) to $\{M_j/2\}_{1 \leqslant j \leqslant k}$, yielding

$$\mathbb{P}(f - \mathbb{E}f \geqslant a) = \mathbb{P}\left(\frac{M_k}{2} > \frac{a}{2} \right) \leqslant \exp\left(-\frac{a^2}{8k} \right). \qquad \square$$

4.3. Proof of Theorem 3.2

In this section we assume $\mathbb{P}$ is the law of a strong Rayleigh measure on $\mathcal{B}_n$ with finite mean $\mathbb{E}N = \mu$. We also let f denote an arbitrary but fixed Lipschitz-1 function on configurations,

https://doi.org/10.1017/S0963548313000345 Published online by Cambridge University Press

and define a function ϕ on $\mathbb{Z}^+$ by

$$\phi(k) := \mathbb{E}(f \mid N = k).$$

Lemma 4.1. *The variable N is distributed as the sum of independent Bernoullis.*

Proof. In the definition of the strong Rayleigh property, setting the variables $z_1, \ldots, z_n$ equal produces a univariate polynomial with no roots in the upper half-plane. As pointed out at the beginning of Section 3 of [2], such a polynomial with real coefficients must have all its roots real. Since the coefficients are non-negative, this implies that it is the generating function for a convolution of Bernoullis. □

Lemma 4.2. *The variable N satisfies*

$$\mathbb{E}e^{\lambda(N-\mu)} \leqslant \exp\left[\mu(e^\lambda - 1 - \lambda)\right]$$

and consequently, for any $a > 0$,

$$\mathbb{P}\left(N \geqslant \mu + \frac{a}{2}\right) \leqslant \exp\left(-\frac{(a/2)^2}{2(\mu + a/2)}\right).$$

Proof. By Lemma 4.1 N is distributed as the sum of independent Bernoullis, which implies the first inequality; this implies the second inequality by (4.5). □

Lemma 4.3. *The function ϕ is Lipschitz*-1.

Proof. By Proposition 2.2 in the case of strong Rayleigh measures on $\mathcal{B}_n$, we know that $\mathbb{P}_{k+1} \rhd \mathbb{P}_k$. By definition of the stochastic covering relation, $(\phi(k+1), \phi(k))$ may be written as $\mathbb{E}(f(\eta), f(\xi))$ where $d(\eta, \xi) = 1$ almost surely. The conclusion then follows from the fact that f is Lipschitz-1. □

Lemma 4.4. *The random variable $\phi(N)$ satisfies the concentration inequality*

$$\mathbb{E}e^{\lambda(\phi(N) - \mathbb{E}\phi(N))} \leqslant e^{\mu(e^\lambda - 1 - \lambda)}.$$

Consequently, the upper tails of $\phi(N)$ obey the bound

$$\mathbb{P}(\phi(N) - \mathbb{E}\phi(N) > t) \leqslant e^{-\frac{t^2}{2(t+\mu)}}.$$

Proof. Pursuant to Lemma 4.1, let $\{Y_j\}$ be a finite or countably infinite collection of independent Bernoulli variables whose sum has the same law as N; we may therefore prove the statements with N replaced by $\sum_j Y_j$. Write

$$\phi\left(\sum_j Y_j\right) - \mathbb{E}\phi\left(\sum_j Y_j\right)$$

https://doi.org/10.1017/S0963548313000345 Published online by Cambridge University Press

as the final term of a martingale $\{M_\ell\}$, where

$$M_\ell := \mathbb{E}\Big(\phi\Big(\sum_j Y_j\Big) \,|\, \mathcal{F}_\ell\Big) - \mathbb{E}\phi \quad \text{and} \quad \mathcal{F}_\ell := \sigma(Y_1, \dots, Y_\ell).$$

If the number of Bernoullis is infinite, the final term is a limit almost surely and in L^2. The martingale $\{M_\ell\}$ is a binary martingale, meaning that conditional on $\mathcal{F}_\ell$, the distribution of $M_{\ell+1}$ is concentrated on two values. In other words,

$$(M_{\ell+1} \,|\, \mathcal{F}_\ell) = p\delta_r + (1-p)\delta_s,$$

where p is the mean of the Bernoulli variable $Y_{\ell+1}$. More importantly, $r = \int f\, d\mu$ and $s = \int f\, d\nu$, where μ is the conditional law of $\sum_j Y_j$ given the values of $Y_1, \dots, Y_\ell$ (which are measurable with respect to $\mathcal{F}_\ell$) and given $Y_{\ell+1} = 1$, and ν is the conditional law of $\sum_j Y_j$ given the values of $Y_1, \dots, Y_\ell$ and given $Y_{\ell+1} = 0$. Clearly μ and ν are probability measures on $\mathbb{Z}^+$ satisfying $d_\infty(\mu, \nu) \leqslant 1$, whence, because ϕ is Lipschitz-1, we see that $|r - s| \leqslant 1$. From (4.3) we then obtain

$$\mathbb{E}\big(e^{\lambda(M_{\ell+1} - M_\ell)} \,|\, \mathcal{F}_\ell\big) \leqslant \exp\big(p(1-p)(e^\lambda - 1 - \lambda)\big).$$

The lemma follows by induction. □

Proof of Theorem 3.2. The event $\{f - \mathbb{E}f > a\}$ is contained in the union of three events:

$$\Big\{N > \mu + \frac{a}{2}\Big\} \cup \Big\{\phi(N) - \mathbb{E}f > \frac{a}{2}\Big\} \cup \Big\{f - \phi(N) > \frac{a}{2},\, N \leqslant \mu + \frac{a}{2}\Big\}.$$

Thus $\mathbb{P}(f - \mathbb{E}f > a)$ is bounded above by the sum of the corresponding probabilities. Each of the first two pieces is bounded above by $\exp[-a^2/(4(a + 2\mu))]$: the first follows from Lemma 4.2 and the second uses Lemma 4.4, noting that $\mathbb{E}\phi = \mathbb{E}\,\mathbb{E}(f \,|\, N) = \mathbb{E}f$. The last piece is bounded above by $\exp[-a^2/(16(a + 2\mu))]$. To see this, observe that the measures $\mathbb{P}_k$ are all strong Rayleigh (this is [2, Corollary 4.18]). For any $k \leqslant \mu + a/2$, we can apply Theorem 3.1 to the homogeneous measure $\mathbb{P}_k$, obtaining

$$\mathbb{P}\Big(f - \phi(N) > \frac{a}{2} \,|\, N = k\Big) \leqslant \exp\Big(-\frac{(a/2)^2}{8k}\Big) \leqslant \exp\Big(-\frac{a^2}{16(a + 2\mu)}\Big).$$

Reassembling these gives the upper bound

$$\mathbb{P}\Big(f - \phi(N) > \frac{a}{2},\, N \leqslant \mu + \frac{a}{2}\Big) \leqslant \exp\Big(-\frac{a^2}{16(a + 2\mu)}\Big).$$

This last piece has the worst bound; using it for all three pieces gives the first inequality of the theorem; we remark that the better upper bound of

$$2\exp[-a^2/(4(a + 2\mu))] + \exp[-a^2/(16(a + 2\mu))]$$

is in fact valid.

For the two-sided bound, we need to consider two more events in addition to the three already considered, namely the events

$$\{\phi(N) - \mathbb{E}\phi(N) < -a/2\} \quad \text{and} \quad \{f - \phi(N) < -a/2, N \leqslant \mu + a/2\}.$$

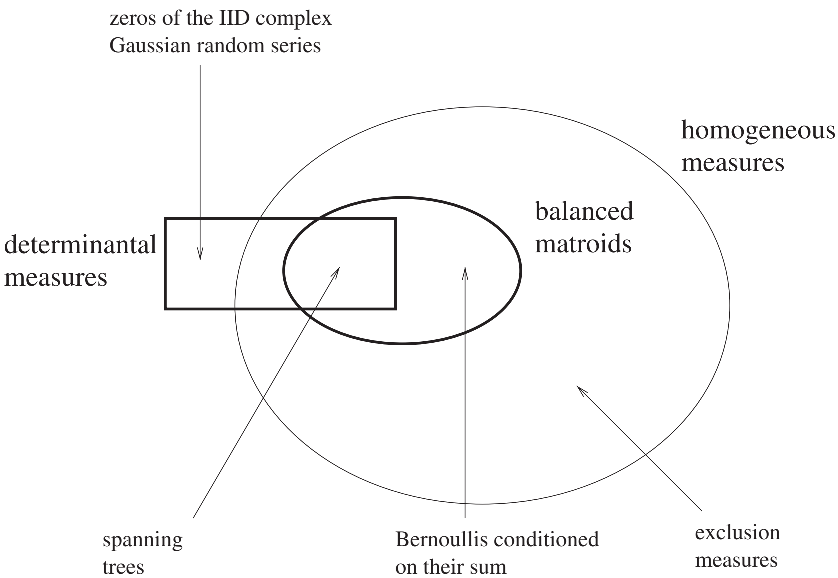

Figure 2. Some classes of strong Rayleigh measures.

The arguments for these two extra events are exactly analogous to two of the three arguments we have already seen, leading to a bound of $\exp[-a^2/(16(a+2\mu))]$ for each of the two new summands and establishing the two-sided bounds. □

5. Applications

In this section we discuss some classes of measures known to satisfy the hypotheses of our concentration results. The Venn diagram in Figure 2 gives a sense of how these classes intersect each other.

5.1. Matroids

A collection $\mathcal{C}$ of subsets of a finite set E, all of a given cardinality, k, is said to be the set of bases of a matroid if it satisfies the base exchange axiom (see, *e.g.*, [27]): if A and B are distinct members of $\mathcal{C}$ and $a \in A \setminus B$, then there exists $b \in B \setminus A$ such that $A \cup \{b\} \setminus \{a\} \in \mathcal{C}$. Given a matroid, it is natural to consider the uniform measure on $\mathcal{C}$. More generally, the *weighted random base* is chosen from the probability measure

$$\nu_w(B) := C \prod_{e \in B} w(e),$$

where $\{w(e) : e \in E\}$ is a collection of non-negative real numbers (weights) and C is a normalizing constant. Identifying E with the set $\{1, \ldots, |E|\}$, the measure ν_w and the random variables $X_e := 1_{e \in B}$ can be thought of as living on $\mathcal{B}_{|E|}$.

For general matroids, $\mathbb{E} X_e X_f$ may be greater than $(\mathbb{E} X_e)(\mathbb{E} X_f)$. Some speculation has been given to the most natural class of matroids for which negative correlation or negative association must hold. Feder and Mihail [8] define a balanced matroid to be a matroid

https://doi.org/10.1017/S0963548313000345 Published online by Cambridge University Press

all of whose minors satisfy pairwise negative correlation. Their proof of the following fact was the basis for the original proof of negative association for determinantal processes [21, Theorem 6.5].

Proposition 5.1 ([8, Theorem 3.2]). *The law ν_w of a random base of a balanced matroid, multiplicatively weighted by the weighting function w, has the SCP.*

Because measures supported on the bases of a matroid are homogeneous, there is nothing gained by improving the SCP to the strong Rayleigh property, and we have the following immediate corollary.

Corollary 5.2. *Let f be a Lipschitz-1 function with respect to Hamming distance on the bases of a balanced matroid of rank k on n elements. Then*

$$\mathbb{P}(f - \mathbb{E}f > a) \leqslant \exp\left(-\frac{a^2}{8k}\right).$$

Example 5.3 (spanning trees). One of the most important examples of a matroid is the set of spanning trees of a finite, connected, undirected graph. To spell this out, a spanning tree for a finite graph $G = (V, E)$ is a subset $E' \subseteq E$ such that (V, E') is a connected and acyclic. The set of spanning trees is a matroid on E. The weighted random spanning tree was shown to be a balanced matroid by [3, Theorem 1]. In fact they showed that it is determinantal (see also [21, Example 1.1] and [15, Example 4.3.2]), though at the time consequences of being determinantal, such as the strong Rayleigh property, had not been developed. Spanning trees are the only well-known class of matroid whose uniform (or weighted) measure is determinantal.

Let $f_0 : \{0,1\}^E \to \mathbb{Z}$ count the number of vertices of odd degree in the graph defined by any subset of the edges. Deleting or adding an edge changes f_0 by at most 2. Let f be the random variable resulting from applying $(1/2)f_0$ to a the weighted random spanning tree on a graph G. Thus f is a Lipschitz-1 function that counts half the number of vertices that have odd degree in the random tree. Random variables that count local properties such as this are of natural graph-theoretic interest. Parity counting variables similar to f play a role, for example, in the randomized TSP approximation algorithm of [11]. The number of edges in any spanning tree is $|V| - 1$. An application of Corollary 5.2 immediately gives the concentration inequality in Theorem 1.1; note that $|V|$, rather than $|E|$, appears in the denominator of the exponent.

Example 5.4 (conditioned Bernoullis, weighted matroids). Let $\lambda_1, \ldots, \lambda_n > 0$ be real numbers and let μ be the measure on the subsets of $[n]$ of cardinality k given by

$$\mu(x_1, \ldots, x_n) = \frac{\prod_{j=1}^n \lambda_j^{x_j}}{\sum_{N(\mathbf{y})=k} \prod_{j=1}^n \lambda_j^{y_j}}.$$

We may think of μ in two ways. The first is as a special case of the weighted random base, specialized to $M(n,k)$, the matroid whose bases are all the subsets of $[n]$ that have cardinality k. The second is that it is the law of independent Bernoulli variables X_j with $\mathbb{E}X_j = \lambda_j/(1+\lambda_j)$, conditioned on $\sum_{j=1}^n X_j = k$. We see from Proposition 5.1 that μ is strong Rayleigh. Alternatively, we may deduce this from the strong Rayleigh property for product measures along with closure under conditioning on the sum [2, Corollary 4.18]. Restricting to $[m]$, for $m < n$, gives a joint distribution on $\mathcal{B}_m$ which may be thought of as a multivariate generalization of the hypergeometric distribution. Because the strong Rayleigh property is inherited, this restriction is strong Rayleigh as well. The resulting concentration properties of these measures have been exploited in [11] in connection with TSP approximation. We remark that more general conditioning, such as conditioning $N := \sum_j X_j$ to lie in an interval of more than two points, does not preserve the strong Rayleigh property.

5.2. Exclusion measures

The symmetric group $\mathcal{S}_n$ acts on $\mathcal{B}_n$ by permuting the coordinates. Suppose a non-negative rate $r(\tau)$ is given for each transposition $\tau \in \mathcal{S}_n$. Define a random evolution on $\mathcal{B}_n$ by letting each pair of coordinates (i, j) transpose independently at rate $r(\tau_{ij})$. In other words, we have a continuous time chain on $\mathcal{B}_n$ which jumps from x to $\tau(x)$ at rate $r(\tau)$ for each transposition τ. This process is known as the symmetric exclusion process.

Borcea, Branden and Liggett [2, Proposition 5.1] prove that the strong Rayleigh property is preserved under this evolution. In particular, because the point mass at a single state is always strong Rayleigh, it follows that the time t distribution of a symmetric exclusion process started from a deterministic state is strong Rayleigh. The stochastic covering property follows, as do PHR and negative association. Interestingly, before the publication of [2], all that was known about this model was negative cylinder dependence [20, Lemma 2.3.4]).

Recently, it was shown by [28] that one can add birth and death to the exclusion dynamics and still preserve the strong Rayleigh property. More specifically, let $\{\alpha_i, \beta_i : 1 \leqslant i \leqslant n\}$ be positive real numbers and let ω_i change to one at rate α_i and to zero at rate β_i, along with the exclusion dynamics. Then the evolution preserves the strong Rayleigh property and in particular, if the starting state is deterministic, all time t marginals are strong Rayleigh.

Corollary 5.5. *Let $\mathbb{P}$ be the law on $\mathcal{B}_n$ resulting from running an exclusion process for a fixed time, starting from a deterministic state with k sites occupied. Then*

$$\mathbb{P}(f - \mathbb{E}f \geqslant a) \leqslant e^{-a^2/(8k)}.$$

Example 5.6. Let $n > 0$ be even and populate an $n \times n$ square of the integer lattice in $\mathbb{Z}^2$ (with torus boundary conditions) by filling all sites in the left half and leaving empty all sites in the right half. Run the symmetric exclusion process for time t with rate 1 on each edge. Let $f_t(\omega)$ denote the number of edges at time t with exactly one endpoint occupied. The mean of f_t starts at n at time 0 and approaches its limiting value of $n^2 - O(1)$ as

https://doi.org/10.1017/S0963548313000345 Published online by Cambridge University Press

$t \to \infty$. Once $t = \Theta(n^2)$, the variance of f_t becomes $\Theta(n^2)$ and the concentration inequality

$$\mathbb{P}(f - \mathbb{E}f \geqslant a) \leqslant e^{-a^2/(4n^2)},$$

which holds for all t, becomes a meaningful Gaussian tail bound (here $k = n^2/2$).

5.3. Determinantal measures on a finite Boolean lattice

We say that a probability measure $\mathbb{P}$ on $\mathcal{B}_n$ is *determinantal* (in the general sense) if there is an $n \times n$ real or complex matrix K such that, for every $S \subseteq \{1, \dots, n\}$,

$$\mathbb{E} \prod_{j \in S} X_j = \det K_S, \tag{4.1}$$

where K_S is the submatrix of K obtained by choosing only those rows and columns whose index is in S. In this definition, the phrase 'general sense' refers to the lack of further assumptions on K. An important subclass is the *Hermitian* determinantal measures, for which the matrix K is Hermitian. In this paper we will be interested only in the Hermitian case and will use the term *determinantal* hereafter to refer only to the case where K is Hermitian. Determinantal measures are known to be negatively associated [21, Theorem 6.5]. In fact they are strong Rayleigh [2, proof of Theorem 3.4] and therefore satisfy the stochastic covering property.

Example 5.7 (uniform or weighted spanning tree). As previously remarked, the uniform or weighted random spanning tree is a determinantal measure.

In the next section we will extend the notion of a determinantal measure to the continuous setting. The extension to a countably infinite set of variables is more straightforward: the kernel K is now indexed by a countably infinite set, but (4.1) may be interpreted as holding for all finite sets S. The following example of a determinantal process on $\mathbb{Z}$ appeared first in [16].

Example 5.8 (positions of non-colliding RWs). Let $\{Y^{(k)} : 1 \leqslant k \leqslant n\}$ be n independent time-homogeneous nearest-neighbour random walks on $\mathbb{Z}$. Start the walks at locations $y_1, \dots, y_n$, and assume that the event that the walks are all at their starting positions at time $2n$ and have not intersected has positive probability. Conditional on this event, the positions at time n form a determinantal measure. That is, the indicator functions $\{X_j\}$ have a determinantal law, where $X_j = 1$ if some $Y^{(k)}$ is at position j at time n, and zero otherwise.

Remark. The positions of non-colliding random walks are given by a determinant under more general conditions (see [18]). The present situation is arranged so as to make the kernel Hermitian.

6. Determinantal point processes

We consider here only simple point processes and often assume $\mathbb{E}N < \infty$ too. If $\rho_k : (\mathbb{R}^d)^k \to \mathbb{R}^+$ are measurable functions, then the simple point process $\mathcal{Z}$ is said to have

https://doi.org/10.1017/S0963548313000345 Published online by Cambridge University Press

joint intensities $\{\rho_k\}$ if, for any k and any family $D_1, \ldots, D_k$ of disjoint Borel subsets of $\mathbb{R}^d$,

$$\mathbb{E}\left[\prod_{j=1}^{k} \mathcal{Z}(D_j)\right] = \int_{\prod_j D_j} \rho_k(x_1, \ldots, x_k)\, dx_1 \cdots dx_k.$$

In particular,

$$\mathbb{E}N = \int_{\mathbb{R}^d} \rho_1(x)\, dx,$$

so under the assumption $\mathbb{E}N < \infty$, we see that $\rho_1(x)\, dx$ is a finite measure on $\mathbb{R}^d$. If ρ_1 is not finite, we will assume it is σ-finite. In any case, ρ_1 is called the *first intensity measure*: see [15, Sections 1.2 and 4.2] for further discussion of joint intensities and determinantal measures.

Definition (determinantal point process). A point process $\mathcal{Z}$ is said to be determinantal if it has joint intensities $\{\rho_k\}$ and there is a measurable kernel $K : (\mathbb{R}^d)^2 \to \mathcal{C}$ such that

$$\rho_k(x_1, \ldots, x_k) = \det\big(K(x_i, x_j)\big)_{1 \leqslant i,j \leqslant k}. \tag{6.1}$$

If $K(y, x) = \overline{K(x, y)}$ for every x, y, then the process is said to be Hermitian. When discussing determinantal processes below, we will always assume they are Hermitian.

Stochastic covering carries over to the continuous case. To state the relevant results we invoke the notion of the *Palm process*. This is a version of the process conditioned on the (measure zero) event of a point at a specified location, x. It may be obtained by conditioning on there being a point within distance ϵ of a given location x, then taking a weak limit. A more complete treatment may be found in [17]. The following proposition is proved in [12].

Proposition 6.1 ([12]). *Suppose $\mathcal{Z}$ is a determinantal point process with continuous kernel K and finite trace. Fix x and let $\mathcal{Z}_x$ denote the Palm process that conditions on a point at x. Let $\mathcal{Z}_x'$ denote the result of removing the point at x from $\mathcal{Z}_x$. Then:*

(i) *whenever $K - L$ is positive semidefinite, the process with kernel K stochastically dominates the process with kernel L (this is Theorem 3 in* [12]*),*
(ii) *$\mathcal{Z}_x'$ is determinantal with kernel L such that $K - L$ is positive semidefinite,*
(iii) *consequently, $\mathcal{Z} \succeq \mathcal{Z}_x'$.*

The continuous analogue of Proposition 2.2 is as follows.

Proposition 6.2. *Let $\mathcal{Z}$ be a determinantal point process with finite mean $\mathbb{E}N = \mu < \infty$. Then for any k for which $\mathbb{P}(N = k+1)$ and $\mathbb{P}(N = k)$ are both non-zero, the conditional distributions of $\mathcal{Z}$ given N satisfy*

$$(\mathcal{Z} \mid N = k+1) \rhd (\mathcal{Z} \mid N = k).$$

https://doi.org/10.1017/S0963548313000345 Published online by Cambridge University Press

Proof. The following facts may be found in [14, Theorem 7]. A determinantal point process $\mathcal{Z}$ with mean $\mu < \infty$ has a kernel K whose spectrum is countable, contained in $[0,1]$, and sums to μ. Furthermore, $\mathcal{Z}$ may be represented as a mixture of homogeneous determinantal processes as follows. Let $\{\lambda_i : i \geqslant 1\}$ enumerate the eigenvalues with multiplicities and let $\{\phi_i\}$ be a corresponding eigenbasis. For each i, flip an independent coin with success probability λ_i. Let I denote the set of i for which the coin-flip was successful. Let K_I be the (random) projection operator onto the subspace spanned by the eigenvectors ϕ_i for which the coin-flip was successful. Then K_I is almost surely a projection of finite dimension $|I|$ and is the kernel of a $|I|$-homogeneous determinantal point process. Choosing K_I at random and then sampling from the corresponding process recovers the law of $\mathcal{Z}$.

Several consequences are apparent. First, conditioning on $N = k$ is the same as conditioning on exactly k successes among the Bernoulli trials. Secondly, the conditional law of I given $|I| = k+1$ stochastically dominates the conditional law of I given $|I| = k$. When the number of Bernoullis is finite, this follows from the strong Rayleigh property for independent Bernoullis; an easy limit argument extends the conclusion to the infinite case. This fact about stochastic domination is equivalent to saying that the conditional law of the random subspace K_I given $|I| = k+1$ stochastically dominates the conditional law of the random subspace K_I given $|I| = k$, in the sense that the two laws can be coupled as (K, K') so that $K' \subseteq K$. When $K' \subseteq K$, the operator $\pi_K - \pi_{K'}$ is positive semidefinite. By Proposition 6.1(ii), we conclude that $(\mathcal{Z} \mid N = k+1) \succeq (\mathcal{Z} \mid N = k)$, which is equivalent to stochastic covering in this case. □

Proof of Theorem 3.5. With f as in the statement of the theorem, and I the collection of indices described in the previous proposition, define $\psi(I)$ to be the expectation of f applied to a configuration chosen from the determinantal process with kernel K_I. Recall the notation $N = |I|$.

The event $\{f - \mathbb{E}f > a\}$ is contained in the union of three events:

$$\left\{N > \mu + \frac{a}{2}\right\} \cup \left\{\psi(I) - \mathbb{E}f > \frac{a}{2}\right\} \cup \left\{f - \psi(I) > \frac{a}{2},\, N \leqslant \mu + \frac{a}{2}\right\}.$$

Thus $\mathbb{P}(f - \mathbb{E}f > a)$ is bounded above by the sum of the corresponding probabilities. Each of the first two pieces is bounded above by $\exp[-a^2/(4(a+2\mu))]$: the first follows from Lemma 4.2 and the second follows from the proof of Lemma 4.4 because ψ is Lipschitz in the Bernoulli variables $Y_i := \mathbf{1}_{i \in I}$. The last piece is bounded above by $\exp[-a^2/(16(a+2\mu))]$. To see this, apply Theorem 3.4 to the homogeneous determinantal processes $\mathbb{P}_I$ with kernels K_I, obtaining, when $|I| = k$, that

$$\mathbb{P}_I\left(f - \psi(I) > \frac{a}{2}\right) \leqslant \exp\left(-\frac{(a/2)^2}{8k}\right) \leqslant \exp\left(-\frac{a^2}{16(a+2\mu)}\right).$$

Reassembling these gives the upper bound

$$\mathbb{P}\left(f - \psi(I) > \frac{a}{2},\, N \leqslant \mu + \frac{a}{2}\right) \leqslant \exp\left(-\frac{a^2}{16(a+2\mu)}\right).$$

The rest of the argument is identical to the conclusion of the proof of Theorem 3.2. □

https://doi.org/10.1017/S0963548313000345 Published online by Cambridge University Press

Example 6.3 (Ginibre's translation-invariant process). Ginibre [10] considers the distribution of eigenvalues of an $k \times k$ matrix with independent complex Gaussian entries. In the limit as $k \to \infty$, the density becomes constant over the whole plane. The limiting process $\mathcal{Z}$ turns out to be a (Hermitian) determinantal point process with kernel

$$K(z_1, z_2) := \frac{1}{\pi} e^{z_1 \overline{z_2}} \exp\left(-\frac{|z_1|^2 + |z_2|^2}{2}\right);$$

see, *e.g.*, [26, (2.16)]. The process $\mathcal{Z}$ is ergodic and invariant under all rigid transformations of the plane. Le Caer and Ho [19] suggested using this process as the set of centres for a random Voronoi tesselation because the mutual repulsion of the points makes the resulting tesselation more realistic than the standard Poisson-Voronoi tesselation for many purposes. Some rigorous results along these lines were obtained in [12].

The mean number of points in any region D is $1/\pi$ times the area $|D|$, so the restriction of $\mathcal{Z}_D$ to such a region of finite area is a determinantal process with finite mean number of points. Fix a finite region, D, and let f count the number of 'lonely' points in D, these being such that no other point of $\mathcal{Z}$ in D is within distance 1. We claim that f is Lipschitz with constant equal to 6. Clearly if a point z is added to the configuration η then f can increase by at most 1. It is well known that the maximum number of points in a unit disk that can be at mutual distance of at least 1 from one another is 6, which implies that the addition of z can result in the loss of at most 6 lonely points. Applying Theorem 3.5 to $f/6$ yields the concentration inequality

$$\mathbb{P}(|f - \mathbb{E}f| \geqslant a) \leqslant 5 \exp\left(-\frac{a^2}{96(a + 12|D|/\pi)}\right).$$

Example 6.4 (zeros of random polynomials). Let $\{X_n\}$ be IID standard complex Gaussian random variables and define the random power series

$$h(z) := \sum_{n=0}^{\infty} X_n z^n.$$

It is easy to see that h is almost surely analytic on the open unit disk and the number of zeros on any disk of radius $\rho < 1$ has finite mean. The remarkable properties of the point process $\mathcal{Z}$ on the unit disk that is the zero set of h are detailed in [25]. It is a determinantal process whose kernel is the Bergman kernel $\pi^{-1}(1 - z\overline{w})^{-2}$. It is invariant under Möbius transformations of the unit disk and has intensity measure $\pi^{-1}/(1 - |z|^2)^2$. Endowing the unit disk with the hyperbolic metric, the Möbius transformations become isometries, whence $\mathcal{Z}$ is hyperbolic isometry invariant.

Fix $\rho < 1$ and $r > 0$ and let f count the number of zeros of the restriction $\mathcal{Z}_\rho$ of $\mathcal{Z}$ to the disk of radius $1 - \rho$ that are 'hyperbolically lonely', meaning that no other point of $\mathcal{Z}_\rho$ is within a hyperbolic distance r. Let c_r denote the maximum number of points at mutual hyperbolic distance r that may be be placed in a disk of hyperbolic radius r. Arguing as in Example 6.3 we see that f is Lipschitz with constant c_r. The mean number of points in $\mathcal{Z}_\rho$ is $\rho^2/(1 - \rho^2)$, which for simplicity we can bound from above by $1/(1 - \rho^2)$. An application of Theorem 3.5 to f/c_r now yields

$$\mathbb{P}(|f - \mathbb{E}f| \geqslant a) \leqslant 5\exp\left(-\frac{a^2}{16c_r a + 32c_r^2(1-\rho^2)^{-1}}\right).$$

Acknowledgements

We thank Subhroshekhar Ghosh and an anonymous referee for detailed comments leading to the correction and improvement of the exposition and results.

References

[1] Alon, N. and Spencer, J. (2008) *The Probabilistic Method*, third edition, Wiley.
[2] Borcea, J., Brändén, P. and Liggett, T. (2009) Negative dependence and the geometry of polynomials. *J. Amer. Math. Soc.* **22** 521–567.
[3] Burton, R. and Pemantle, R. (1993) Local characteristics, entropy and limit theorems for uniform spanning trees and domino tilings via transfer-impedances. *Ann. Probab.* **21** 1329–1371.
[4] Chernoff, H. (1952) A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.
[5] Dubhashi, D. and Ranjan, D. (1998) Balls and bins: A study in negative dependence. *Random Struct. Alg.* **13** 99–124.
[6] Durrett, R. (2004) *Probability: Theory and Examples*, third edition, Duxbury Press.
[7] Farcomeni, A. (2008) Some finite sample properties of negatively dependent random variables. *Theor. Prob. Math. Statist.* **77** 155–163.
[8] Feder, T. and Mihail, M. (1992) Balanced matroids. In *Annual ACM Symposium on Theory of Computing*, pp. 26–38.
[9] Freedman, D. (1975) On tail probabilities for martingales. *Ann. Probab.* **3** 100–118.
[10] Ginibre, J. (1965) Statistical ensembles of complex, quaternion, and real matrices. *J. Math. Phys.* **6** 440–449.
[11] Gharan, S., Saberi, A. and Singh, M. (2011) A randomized rounding approach to the Traveling Salesman Problem. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pp. 550–559.
[12] Goldman, A. (2010) The palm measure and the Voronoi tesselation for the Ginibre process. *Ann. Appl. Probab.* **20** 90–128.
[13] Hoeffding, W. (1963) Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.
[14] Hough, J. B., Krishnapur, M., Peres, Y. and Virág, B. (2006) Determinantal processes and independence. *Probability Surveys* **3** 206–229.
[15] Hough, J. B., Krishnapur, M., Peres, Y. and Virág, B. (2009) *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*, Vol. 51 of *University Lecture Series*, AMS.
[16] Johansson, K. (2003) Discrete polynuclear growth and determinantal processes. *Comm. Math. Phys.* **242** 277–329.
[17] Kallenberg, O. (1986) *Random Measures*, fourth edition, Academic Press.
[18] Karlin, S. and McGregor, J. (1959) Confidence probabilities. *Pacific J. Math.* **9** 1141–1164.
[19] Le Caer, G. and Ho, J. S. (1990) The Voronoi tesselation generated from eigenvalues of complex random matrices. *J. Phys. A* **23** 3279–3295.
[20] Liggett, T. (1977.) The stochastic evolution of an infinite system of interacting particles. In *Ecole de Probabilites de Saint Flour VI, 1976*, Vol. 598 of *Lecture Notes in Mathematics*, Springer, pp. 187–248.
[21] Lyons, R. (2003) Determinantal probability measures. *IHES* **98** 167–212.

[22] McDiarmid, C. (1989) On the method of bounded differences. In *Surveys in Combinatorics* (J. Siemons, ed.), Vol. 41 of *London Mathematical Society Lecture Note Series*, Cambridge University Press, pp. 148–188.

[23] Panconesi, A. and Srinivasan, A. (1997) Randomized distributed edge coloring via an extension of the Chernoff–Hoeffding bounds. *SIAM J. Comput.* **26** 350–368.

[24] Pemantle, R. (2000) Toward a theory of negative dependence. *J. Math. Phys.* **41** 1371–1390.

[25] Peres, Y. and Virág, B. (2005) Zeros of the i.i.d. Gaussian power series: A conformally invariant determinantal process. *Acta Math.* **194** 1–35.

[26] Soshnikov, A. (2000) Determinantal random point fields. *Uspekhi Mat. Nauk* **55** 107–160.

[27] Tutte, W. (1971) *Introduction to the Theory of Matroids*, Vol. 37 of *Modern Analytic and Computational Methods in Science and Mathematics*, American Elsevier.

[28] Wagner, D. G. (2011) Multivariate stable polynomials: Theory and application. *Bull. Amer. Math. Soc.* **48** 53–84.