The Annals of Applied Probability
2010, Vol. 20, No. 2, 696–721
DOI: 10.1214/09-AAP634

NORMAL APPROXIMATION FOR COVERAGE MODELS OVER BINOMIAL POINT PROCESSES

BY LARRY GOLDSTEIN AND MATHEW D. PENROSE[1]

University of Southern California and University of Bath

We give error bounds which demonstrate optimal rates of convergence in the CLT for the total covered volume and the number of isolated shapes, for germ-grain models with fixed grain radius over a binomial point process of n points in a toroidal spatial region of volume n. The proof is based on Stein's method via size-biased couplings.

1. Introduction. Given a collection of n independent uniformly distributed random points in a d-dimensional cube of volume n (the so-called *binomial point process*), let V denote the (random) total volume of the union of interpenetrating balls of fixed radius ρ centered at these points, and let S denote the number of balls of radius $\rho/2$ (centered at the same set of points) which are singletons, that is, do not overlap any other such ball. These variables are fundamental topics of interest in the stochastic geometry of coverage processes and random geometric graphs [9, 10, 13, 18].

As $n \to \infty$ with ρ fixed (the so-called thermodynamic limit), both V and S are known to satisfy a central limit theorem (CLT) [12, 13, 16]. In the present work we provide associated Berry–Esseen type results; that is, we show under periodic boundary conditions that the cumulative distribution functions converge to that of the normal at the same $O(n^{-1/2})$ rate as for a sum of n independent identically distributed variables, and provide bounds on the quality of the normal approximation for finite n.

Were we to consider instead a Poisson-distributed number of points, that is, a Poisson point process instead of a binomial one, both of our variables of interest could be expressed as sums of locally dependent random variables, and thereby Berry–Esseen type bounds could be (and have been) obtained by known methods [1, 8, 15, 17]. But with a nonrandom number of points, the local dependence is lost and the de-Poissonization arguments in [13, 16] do not provide error bounds for the de-Poissonized CLTs. The early work of Moran [11, 12] on V was in response to queries in the statistical physics literature (including the well-known

Received December 2008; revised August 2009.

[1]Supported in part by the Alexander von Humboldt Foundation through a Friedrich Wilhelm Bessel Research Award.

AMS 2000 subject classifications. Primary 60D05; secondary 62E17, 60F05, 05C80.

Key words and phrases. Stochastic geometry, coverage process, Berry–Esseen theorem, size biased coupling, Stein's method.

paper of Widom and Rowlinson [19]) which specifically addressed normal approximation of V for nonrandom n, and in general, it seems worthwhile to study the de-Poissonized setting since in practice one might well observe the actual number of points, in which case the conditional distribution of any test statistic, based on what is observed, will be over a binomial rather than a Poisson point process.

The variables V and S are just two of a large class of variables of interest that can be expressed as a sum, over the n points, of terms that depend only on the configuration of nearby points in some sense. General CLTs have been developed for such variables [14, 16] and general Berry–Esseen type results are available in the Poissonized setting [8, 15, 17], but it remains open to provide a generally applicable Berry–Esseen type result for such sums when n is nonrandom (see however [3], which is discussed further in Section 2). However, there seem to be good prospects of adapting the approach of the present paper (which is new in the geometrical setting) to a wider class of geometrical sums.

Our approach to normal approximation is based on Stein's method via size-biased couplings. Given a nonnegative random variable Y with positive finite mean $\mu = \mathbb{E}Y$, we say Y' has the Y size-biased distribution if $P[Y' \in dy] = (y/\mu)P[Y \in dy]$, or more formally, if

$$\mathbb{E}[Yg(Y)] = \mu\mathbb{E}g(Y') \qquad \text{for bounded continuous functions } f. \tag{1.1}$$

The method of size-biased couplings was introduced by Baldi, Rinott and Stein [2], who used it to develop bounds of order $\sigma^{-1/2}$ to the normal approximation to the number of local maxima Y of a random function on a graph, where $\sigma^2 = \mathrm{Var}(Y)$. Goldstein and Rinott [7] extended the technique to multivariate normal approximations, and improved the rate to σ^{-1} for the expectation of smooth functions of a vector $\mathbf{Y}$ recording the number of edges with certain fixed degrees in a random graph. In [6], the method is used to give bounds of order σ^{-1} for various functions on graphs and permutations.

Here we shall use Lemma 3.1 below, which improves the constant in a more general result from [6]. Loosely speaking, this result says that given any coupling of Y and Y' on a common space, an upper bound on the distance between the distribution of Y and the normal can be found which involves functions of the joint distribution of Y, Y' in terms of (i) the uniform distance between Y and Y', that is, the L^∞ norm of $Y - Y'$, and (ii) the variance of $\mathbb{E}[Y - Y'|Y]$.

In Section 4 we show how to find a coupled realization of Y' that is uniformly close to Y, for those Y under consideration here. To do this we show that here the size-biasing amounts to conditioning the (binomial) number of points falling in a certain (randomly located) ρ-ball to be nonzero, and can be achieved by modifying at most a single point location to obtain Y' from Y, so that $\|Y' - Y\|_\infty$ is bounded. This construction may be of independent interest, along with Lemma 4.1 (a general result on how to size-bias a conditional probability) and Lemma 3.1.

2. Results. Let $d \geq 1$ and $n \geq 4$ be integers. Suppose $U_1, \ldots, U_n$ are independent random d-vectors, uniformly distributed over the cube $C_n := [0, n^{1/d})^d$ (we write U_i rather than $U_{n,i}$ because the value of n should be clear from the context). Write $\mathcal{U}_n$ for the point set $\{U_1, \ldots, U_n\}$. For x, y in the cube C_n, let $D(x, y)$ denote the distance between x and y under the Euclidean toroidal metric on C_n. For $x \in C_n$ and $r > 0$ let $B_r(x)$ denote the ball $\{y \in C_n : D(x, y) \leq r\}$. Let $B_{i,r}$ denote the ball $B_r(U_i)$. Given r, the collection of balls $B_{i,r}$ form a coverage process (also known as a germ-grain model) in C_n; see [9, 18]. Let $\rho > 0$, and define

$$V := \text{Volume}\left(\bigcup_{i=1}^{n} B_{i,\rho}\right); \tag{2.1}$$

$$S := \sum_{i=1}^{n} \mathbf{1}\{\mathcal{U}_n \cap B_{i,\rho} = \{U_i\}\}. \tag{2.2}$$

Then V is the total covered volume for the coverage process with $r = \rho$, while S is the number of singletons (isolated balls) in the case $r = \rho/2$, and may also be viewed as the number of isolated points in the geometric graph on vertex set $\mathcal{U}_n$ with distance parameter ρ [13].

Let Z denote a standard normal random variable. Given a random variable X with $\mathrm{SD}(X) := \sqrt{\mathrm{Var}(X)} \in (0, \infty)$, let D_X denote the Kolmogorov distance between the distribution of X (scaled and centered) and that of Z, that is,

$$D_X := \sup_{t \in \mathbb{R}} \left| P\left[\frac{X - \mathbb{E}X}{\mathrm{SD}(X)} \leq t\right] - P[Z \leq t] \right|.$$

Our main results provide bounds in the normal approximation for V and S; if ρ is fixed then as $n \to \infty$,

$$D_V = O(n^{-1/2}); \qquad D_S = \Theta(n^{-1/2}). \tag{2.3}$$

Recall that $a_n = \Theta(b_n)$ means that $a_n = O(b_n)$ and $b_n = O(a_n)$. We conjecture that the first bound in (2.3) can be improved to $\Theta(n^{-1/2})$.

To state our results more precisely, we need further notation. Set π_d to be the volume of the unit ball in d dimensions, that is, $\pi_d := \pi^{d/2}/\Gamma(1 + d/2)$, and $\phi := \pi_d \rho^d$. We say two unit balls *touch* if their closures intersect, but their interiors do not. Let κ_d (respectively, κ_d^*) denote the maximum number of closed unit balls in d dimensions that can be packed so they all intersect (respectively, touch) a closed unit ball at the origin, but are disjoint from each other (respectively, have disjoint interiors). Then κ_d^* is the so-called *kissing number* in d dimensions, which has been studied for centuries (see [5, 20]). It is not hard to see κ_d^* is an upper bound for κ_d, and in most dimensions it seems likely that $\kappa_d = \kappa_d^*$, but $\kappa_2 = 5$ whereas $\kappa_2^* = 6$. It is known that $\kappa_3 = \kappa_3^* = 12$. Set $\kappa_d^+ := 1 + \kappa_d$.

Set $\mu_V := \mathbb{E}[V]$, $\mu_S := \mathbb{E}[S]$, $\sigma_V := \mathrm{SD}(V)$, and $\sigma_S := \mathrm{SD}(S)$. It is straightforward to write down formulae for μ_V, μ_S, σ_V^2 and σ_S^2; see (7.1), (7.2) and (7.3).

Our first two main results provide nonasymptotic upper bounds on the Kolmogorov distance.

THEOREM 2.1. *If $n > 6^d\phi$, then*

$$D_V \leq \frac{\mu_V}{5\sigma_V^2}\left(\sqrt{\frac{11\phi^2}{\sigma_V} + \frac{5\sqrt{\eta_V(n,\rho)}}{\sqrt{n}}} + \frac{2\phi}{\sqrt{\sigma_V}}\right)^2$$

with

$$\begin{aligned}\eta_V(n,\rho) := {}& 2\phi^2\big((3^d+1)\phi+1\big)^2\\ &\times\left(1+(2^d+1)6^d\phi+\left(\frac{2n-6^d\phi}{n-6^d\phi}\right)6^{2d}\phi^2\right)\\ &+2\phi^4\left(3(4^d+2^d)\phi+3(4^d)\phi^2\left(\frac{2n-3(2^d)\phi}{n-3(2^d)\phi}\right)+4+\frac{2}{n}\right).\end{aligned}\tag{2.4}$$

THEOREM 2.2. *If $n > \max(3^d, 2^{d+1}+1)\phi$, then*

$$D_S \leq \frac{n-\mu_S}{5\sigma_S^2}\left(\sqrt{\frac{11(\kappa_d^+)^2}{\sigma_S} + \frac{5\sqrt{\eta_S(n,\rho)}}{\sqrt{n}}} + \frac{2\kappa_d^+}{\sqrt{\sigma_S}}\right)^2$$

with

$$\begin{aligned}\eta_S(n,\rho) := {}& 2(1+2\kappa_d)^2\left(1+(2^d+1)3^d\phi+\left(\frac{2n-3^d\phi}{n-3^d\phi}\right)9^d\phi^2\right)\\ &+\frac{(\kappa_d^+)^2}{2}\Bigg((2^d+2(3^d)+3)\phi\\ &\qquad+(2^{d+1}+1)\left(\frac{2n-(2^{d+1}+1)\phi}{n-(2^{d+1}+1)\phi}\right)\phi^2+\frac{4n-2}{n-1}\Bigg).\end{aligned}\tag{2.5}$$

By using the inequality $(x+y)^2 \leq 2(x^2+y^2)$, the bounds in Theorems 2.1 and 2.2 can replaced by bounds which are simpler, though less sharp.

The next result confirms that for large n, all of μ_V, σ_V^2, μ_S and σ_S^2 are $\Theta(n)$, so that (2.3) follows from Theorems 2.1 and 2.2. To provide details we require further notation.

For $0 \leq r \leq 2$, write $\omega_d(r)$ for the volume of the union of two unit balls in $\mathbb{R}^d$ with centers distant r apart (see (7.5) for a formula). Define the integral

$$J_{r,d}(\rho) := d\pi_d \int_0^r \exp(-\rho^d\omega_d(t))t^{d-1}\,dt \tag{2.6}$$

and the functions

$$g_V(\rho) := \rho^d J_{2,d}(\rho) - (2^d\phi+\phi^2)e^{-2\phi}; \tag{2.7}$$

$$g_S(\rho) := e^{-\phi} - \big(1+(2^d-2)\phi+\phi^2\big)e^{-2\phi} + \rho^d\big(J_{2,d}(\rho)-J_{1,d}(\rho)\big). \tag{2.8}$$

Also, define $\eta_V(\rho) := \lim_{n\to\infty}\eta_V(n,\rho)$ and $\eta_S(\rho) := \lim_{n\to\infty}\eta_S(n,\rho)$. Formulae for these limits are immediate from the definitions (2.4) and (2.5).

THEOREM 2.3. *If ρ is fixed then as $n \to \infty$,*

$$\lim_{n\to\infty}(1 - n^{-1}\mu_V(\rho)) = \lim_{n\to\infty}(n^{-1}\mu_S(\rho)) = e^{-\phi}; \tag{2.9}$$

$$\lim_{n\to\infty}(n^{-1}\sigma_V^2) = g_V(\rho) > 0; \tag{2.10}$$

$$\lim_{n\to\infty}(n^{-1}\sigma_S^2) = g_S(\rho) > 0 \tag{2.11}$$

and

$$\limsup_{n\to\infty}(n^{1/2}D_V) \le \frac{1-e^{-\phi}}{5g_V(\rho)}\left(\sqrt{\frac{11\phi^2}{g_V(\rho)^{1/2}} + 5\eta_V^{1/2}} + \frac{2\phi}{g_V(\rho)^{1/4}}\right)^2; \tag{2.12}$$

$$\limsup_{n\to\infty}(n^{1/2}D_S) \le \frac{1-e^{-\phi}}{5g_S(\rho)}\left(\sqrt{\frac{11(\kappa_d^+)^2}{g_S(\rho)^{1/2}} + 5\eta_S^{1/2}} + \frac{2\kappa_d^+}{g_S(\rho)^{1/4}}\right)^2; \tag{2.13}$$

$$\liminf_{n\to\infty}(n^{1/2}D_S) \ge (8\pi g_S(\rho))^{-1/2}. \tag{2.14}$$

Theorems 2.1 and 2.2 are proved in Sections 5 and 6, respectively. Theorem 2.3 is proved in Section 7, where we also derive numerical values for the asymptotic upper bounds in Theorem 2.3, for some particular cases.

Remarks. The limiting variances in (2.10), respectively (2.11), are consistent with those given by Moran [11, 12], respectively, Penrose ([13], Theorem 4.14). Moran and Penrose do not explicitly rule out the possibility that these limiting variances might be zero, as we do here.

Clearly (2.12) and (2.13) imply central limit theorems whereby both $(V - \mu_V)/\sigma_V$ and $(S - \mu_S)/\sigma_S$ converge in distribution to the standard normal, thereby providing an alternative to existing proofs of these central limit theorems [12, 13, 16]. In the Poissonized setting, nonasymptotic bounds analogous to those in Theorems 2.1 and 2.2 are given in [15] and imply $O(n^{-1/2})$ bounds analogous to (2.12) and (2.13). In the de-Poissonized setting considered here, Chatterjee [3] provides bounds similar to those in (2.12) and (2.13), which hold for general metric spaces, but using the Kantorovich–Wasserstein distance, rather than the Kolmogorov distance considered here, and without providing any explicit constants. As stated in [3], "obtaining optimal rates for the Kolmogorov distance requires extra work and new ideas."

Generalizations of our results should be possible in many directions. These include:

More general germ-grain models. Replace the balls of fixed radius in the description of V and S by (independent identically distributed) balls of random radius, or more generally, random shapes.

Random measures. Consider the random measure associated with V (the Lebesgue measure on the covered region) or with S (a sum of Dirac measures at the isolated points), and look at normal approximation for the random variable given by the integral of a test function f on C_n with respect to that measure.

Euclidean distance. Suppose in the definition of V and S, that the periodic boundary conditions on C_n are dropped, that is, the toroidal distance D is replaced by the ordinary Euclidean distance.

Nonuniform points. Consider a sequence of independent random points $(\mathbf{X}_n)_{n\geq 1}$ with a common density function $\nu : \mathbb{R}^d \to \mathbb{R}$. Placing balls of radius r_n around each point of $\mathcal{X}_n := \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$, for some specified sequence r_n tending to zero, one may define quantities analogous to V and S. When $r_n \propto n^{-1/d}$ this is a rescaling of our model but allows for nonuniform ν. Our approach might also provide information about other asymptotic regimes.

k-nearest neighbors. Let $k \in \mathbb{N}$ and consider the number of points U_i whose kth nearest neighbor in the point set $\mathcal{U}_n \setminus \{U_i\}$ lies at a distance greater than ρ. The case $k = 1$ reduces to S.

These extensions generally seem to be nontrivial, and worthy of further study.

3. Lemmas. The proof of (2.12) and (2.13) is based on the following result. This result improves the constant which would be obtained by applying the more general Theorem 1.2 of [6] to the particular case of Kolmogorov distance.

LEMMA 3.1. *Let $Y \geq 0$ be a random variable with mean μ and variance $\sigma^2 \in (0, \infty)$, and let Y^s be defined on the same probability space, with the Y-size biased distribution. If $P[|Y^s - Y| \leq B] = 1$ for some constant $B > 0$, then*

$$D_Y \leq \frac{\mu}{5\sigma^2}\left(\sqrt{\frac{11B^2}{\sigma} + 5\Delta} + \frac{2B}{\sqrt{\sigma}}\right)^2, \tag{3.1}$$

where $\Delta := \sqrt{\mathrm{Var}(\mathbb{E}[Y^s - Y|Y])}$.

PROOF. Given $z \in \mathbb{R}$ and $\varepsilon > 0$, define real-valued functions h_z and $h_{z,\varepsilon}$ by

$$h_z(x) = \mathbf{1}_{(-\infty,z]}(x), \qquad h_{z,\varepsilon}(x) = \varepsilon^{-1}\int_0^\varepsilon h_z(x-t)\,dt, \qquad z \in \mathbb{R}.$$

Then with $W := (Y - \mu)/\sigma$ and Z denoting a standard normal, by definition

$$D_Y = \sup\{|\mathbb{E}h_z(W) - \mathbb{E}h_z(Z)| : z \in \mathbb{R}\}. \tag{3.2}$$

For $\varepsilon > 0$, set

$$D_Y^\varepsilon := \sup\{|\mathbb{E}h_{z,\varepsilon}(W) - \mathbb{E}h_{z,\varepsilon}(Z)| : z \in \mathbb{R}\}. \tag{3.3}$$

Fix z and ε, and let f be the unique bounded solution of the Stein equation

$$f'(w) - wf(w) = h_{z,\varepsilon}(w) - \mathbb{E}h_{z,\varepsilon}(Z)$$

for $h_{z,\varepsilon}$; see [4]. With some abuse of notation, let $W^s = (Y^s - \mu)/\sigma$. Then

$$\begin{aligned}
&\mathbb{E}[h_{z,\varepsilon}(W) - \mathbb{E}h_{z,\varepsilon}(Z)] \\
&\qquad = \mathbb{E}[f'(W) - Wf(W)] \\
&\qquad = \mathbb{E}\Big[f'(W) - \frac{\mu}{\sigma}\big(f(W^s) - f(W)\big)\Big] \qquad (3.4)\\
&\qquad = \mathbb{E}\Big[f'(W)\Big(1 - \frac{\mu}{\sigma}(W^s - W)\Big) \\
&\qquad\qquad - \frac{\mu}{\sigma}\int_0^{W^s - W}\big(f'(W+t) - f'(W)\big)\,dt\Big].
\end{aligned}$$

The following bounds on the solution f can be found in [4]:

$$|f'(w)| \le 1 \tag{3.5}$$

and

$$|f'(w+t) - f'(w)| \le (|w|+1)|t| + \varepsilon^{-1}\int_{t\wedge 0}^{t\vee 0}\mathbf{1}_{[z,z+\varepsilon]}(w+u)\,du. \tag{3.6}$$

Noting that $\mathbb{E}Y^s = \mathbb{E}Y^2/\mu$ by (1.1) with $g(y) = y$, we find that

$$\frac{\mu}{\sigma}\mathbb{E}[W^s - W] = \frac{\mu}{\sigma^2}\Big(\frac{\mathbb{E}Y^2}{\mu} - \mu\Big) = 1,$$

and therefore, taking expectation by conditioning, and then using (3.5), we have

$$\Big|\mathbb{E}\Big\{f'(W)\mathbb{E}\Big[1 - \frac{\mu}{\sigma}(W^s - W)\Big|W\Big]\Big\}\Big| \le \frac{\mu}{\sigma}\sqrt{\operatorname{Var}(\mathbb{E}[W^s - W|W])} = \frac{\mu}{\sigma^2}\Delta.$$

Now, using (3.4) and (3.6) yields

$$\begin{aligned}
&|\mathbb{E}[h_{z,\varepsilon}(W) - \mathbb{E}h_{z,\varepsilon}(Z)]| \\
&\quad \le \frac{\mu}{\sigma^2}\Delta + \frac{\mu}{\sigma}\mathbb{E}\Big[\int_{(W^s-W)\wedge 0}^{(W^s-W)\vee 0}(|W|+1)|t|\,dt \\
&\qquad\qquad + \int_{-B/\sigma}^{B/\sigma}\varepsilon^{-1}\int_{t\wedge 0}^{t\vee 0}\mathbf{1}_{[z,z+\varepsilon]}(W+u)\,du\,dt\Big] \qquad (3.7)\\
&\quad \le \frac{\mu}{\sigma^2}\Delta + \frac{\mu B^2}{2\sigma^3}(\mathbb{E}|W|+1) + \frac{\mu}{\sigma}\varepsilon^{-1}\int_{-B/\sigma}^{B/\sigma}\int_{t\wedge 0}^{t\vee 0}(0.4\varepsilon + 2D_Y)\,du\,dt \\
&\quad \le \frac{\mu}{\sigma^2}\Delta + 1.4\frac{\mu}{\sigma^3}B^2 + \frac{2\mu}{\sigma^3}B^2\varepsilon^{-1}D_Y,
\end{aligned}$$

where in the second-to-last inequality above we have used the fact that

$$P[\alpha \le W \le \beta] \le (\beta - \alpha)/\sqrt{2\pi} + 2D_Y,$$

and in the last, the fact that $\mathbb{E}|W| \le 1$. By (3.2) we see that D_Y^{ε} is bounded by (3.7), and since (3.2) and (3.3) imply $D_Y \le 0.4\varepsilon + D_Y^{\varepsilon}$, substitution yields

$$D_Y \le \gamma(\varepsilon) := \frac{a\varepsilon + b}{1 - c/\varepsilon},$$

where

$$a := \frac{2}{5}, \qquad b := \frac{\mu}{\sigma^2}\Delta + \frac{7}{5}\frac{\mu}{\sigma^3}B^2 \quad \text{and} \quad c := \frac{2\mu B^2}{\sigma^3}.$$

The optimum bound on D_Y is at the positive root of $\gamma'(\varepsilon) = 0$, namely $\varepsilon = c + r$ where $r := \sqrt{c^2 + cb/a}$.

We wish to calculate $\gamma(c+r)$. The denominator equals

$$1 - \frac{c}{c+r} = 1 - \frac{c(c-r)}{c^2 - r^2} = 1 + \frac{a(c-r)}{b} = \frac{b + a(c-r)}{b},$$

and therefore

$$\begin{aligned}
\gamma(c+r) &= b\left(\frac{a(c+r)+b}{a(c-r)+b}\right) = b\left(\frac{c+b/a+r}{c+b/a-r}\right) \\
&= b\left(\frac{(c+b/a+r)^2}{(c+b/a)^2 - r^2}\right) = b\left(\frac{(c+b/a+r)^2}{cb/a + (b/a)^2}\right) \\
&= a\left(\frac{(c+b/a+r)^2}{c+(b/a)}\right) = a\left(\frac{(c+b/a+\sqrt{c}\sqrt{c+b/a})^2}{c+(b/a)}\right) \\
&= a\left(\sqrt{c+b/a} + \sqrt{c}\right)^2 = \frac{2}{5}\left(\sqrt{\frac{11}{2}\frac{\mu B^2}{\sigma^3} + \frac{5}{2}\frac{\mu}{\sigma^2}\Delta} + \sqrt{\frac{2\mu B^2}{\sigma^3}}\right)^2,
\end{aligned}$$

and this bound on D_Y yields (3.1). □

Let $\operatorname{Bin}(n, p)$ denote the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$. Our next two lemmas are concerned with binomial and conditioned binomial distributions. Lemma 3.2 is used to prove Lemma 3.3.

LEMMA 3.2. *Let $m \in \mathbb{N}$ and $p \in (0, 1)$. Suppose $N \sim \operatorname{Bin}(m, p)$, and $\mathcal{L}(N') = \mathcal{L}(N|N > 0)$, $N'' - 1 \sim \operatorname{Bin}(m - 1, p)$. Then for all $k \in \mathbb{N}$,*

$$P[N \ge k] \le P[N' \ge k] \le P[N'' \ge k]. \tag{3.8}$$

PROOF. The first inequality in (3.8) is easy since for $k \ge 1$, by definition $P[N' \ge k] = P[N \ge k]/P[N \ge 1]$. It remains to prove the second inequality. Suppose $\xi_1, \xi_2, \ldots$ are independent Bernoulli random variables with parameter p. Let

$M = \min\{i : \xi_i = 1\}$ and $\tilde{N}'' := \sum_{i=M}^{M+m-1} \xi_i$. Then M and $\tilde{N}''$ are independent and $\mathcal{L}(\tilde{N}'') = \mathcal{L}(N'')$.

Define the random variables

$$J := \left\lceil \frac{M}{m} \right\rceil \quad \text{and} \quad \tilde{N}' := \sum_{i=m(J-1)+1}^{mJ} \xi_i .$$

In other words, split the sequence of Bernoulli trials into disjoint intervals of length m, and let $\tilde{N}'$ denote the number of successful Bernoulli trials in the first such interval that contains at least one successful trial.

Then $\tilde{N}'$ has the distribution of N', and by construction $\tilde{N}' \le \tilde{N}''$ almost surely. Since $\tilde{N}''$ has the distribution of N'', this shows that N' is stochastically dominated by N'', that is, the second inequality in (3.8) holds. □

Our next lemma demonstrates the existence of a "uniformly close coupling" of random variables with a binomial distribution, and with the same distribution conditioned to be nonzero (denoted, respectively, N and M in the lemma). This result will be used in Section 4 to provide a uniformly close coupling of V [given by (2.2)] and its size biased version, and likewise for S (in fact, for $n - S$).

LEMMA 3.3. *Let $m \in \mathbb{N}$ and $p \in (0, 1)$. Suppose $N \sim \mathrm{Bin}(m, p)$, with $N = \sum_{i=1}^{m} \xi_i$ where ξ_i are independent Bernoulli variables with parameter p. Defining π_k by*

$$\pi_k := \begin{cases} \dfrac{P[N > k|N > 0] - P[N > k]}{P[N = k](1 - (k/m))}, & \text{if } 0 \le k \le m - 1, \\ 0, & \text{if } k = m, \end{cases} \tag{3.9}$$

we then have $0 \le \pi_k \le 1$ for all $k \in \{0, \dots, m\}$.

Suppose also that $\mathcal{B}$ is a further Bernoulli variable with $P(\mathcal{B} = 1|\xi_1, \dots, \xi_m) = \pi_N$, and suppose I is an independent discrete uniform random variable over $\{1, 2, \dots, m\}$. Set $M := N + (1 - \xi_I)\mathcal{B}$, that is, let M be given by the same sum as N except that if $\mathcal{B} = 1$ the Ith term is set to 1. Then

$$\mathcal{L}(M) = \mathcal{L}(N|N > 0). \tag{3.10}$$

PROOF. Lemma 3.2 shows $\pi_k \ge 0$. For the upper bound, set $N'' = 1 + \sum_{i=2}^{m} \xi_i$. Then $N'' - 1 \sim \mathrm{Bin}(m - 1, p)$ and N'' is equal either to N or to $N + 1$, with $P[N'' = k + 1|N = k] = 1 - (k/m)$ for $0 \le k \le m$. Hence for all k, by Lemma 3.2,

$$P[N > k] + P[N = k](1 - k/m) = P[N'' > k] \ge P[N > k|N > 0]$$

so $\pi_k \le 1$. Also, assertion (3.10) follows by (3.9) and the fact that

$$\{M > k\} = \{N > k\} \cup \{N = k, \mathcal{B} = 1, \xi_I = 0\}.$$ □

Our next result refers to measurable real-valued functions ψ defined on all pairs $(x, \mathcal{X})$ such that $\mathcal{X}$ is a finite subset of C_n and $x \in \mathcal{X}$. We say that such a functional ψ is *translation-invariant* if $\psi(x, \mathcal{X}) = \psi(y+x, y+\mathcal{X})$ for all $x, \mathcal{X}$ and all $y \in C_n$ (here addition is in the torus C_n, and $y + \mathcal{X} := \{y + w : w \in \mathcal{X}\}$). For $r > 0$, we say that ψ has *radius* r if $\psi(x, \mathcal{X})$ is unaffected by the addition of points to, or removal of points from, the point set $\mathcal{X}$ at a distance more than r from x, that is, if for all $(x, \mathcal{X})$ we have $\psi(x, \mathcal{X}) = \psi(x, \mathcal{X} \cap B_r(x))$. The notion of radius is the same as that of *range of interaction* used in [15]; see also the notion of *radius of stabilization*, in [15, 17] and elsewhere. We also define

$$\|\psi\| := \operatorname*{ess\,sup}_{x,\mathcal{X}}\{|\psi(x, \mathcal{X})|\};$$

$$\operatorname{rng}(\psi) := \operatorname*{ess\,sup}_{x,\mathcal{X}}\{\psi(x, \mathcal{X})\} - \operatorname*{ess\,inf}_{x,\mathcal{X}}\{\psi(x, \mathcal{X})\}.$$

Recall that $\mathcal{U}_n := \{U_1, \dots, U_n\}$ denotes a collection of n independent uniformly distributed points in C_n, and π_d is the volume of the unit d-ball.

LEMMA 3.4. *Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$ with $2 \le k \le n$. Suppose that for $i = 1, \dots, k$, ψ_i is a measurable real-valued function defined on all pairs $(x, \mathcal{X})$ with $\mathcal{X}$ a finite set in C_n and $x \in \mathcal{X}$. Suppose for each i that ψ_i is translation-invariant and has radius r_i for some $r_i \in (0, \infty)$, and that $\|\psi_i\| < \infty$, and $\mathbb{E}[\psi_1(U_1, \mathcal{U}_n)] = 0$. With $\phi_i := \pi_d r_i^d$, suppose also that $\phi_2 + \cdots + \phi_k < n$. Then*

$$\begin{aligned}\left|\mathbb{E}\left[\prod_{i=1}^k \psi_i(U_i, \mathcal{U}_n)\right]\right| \le{}& \left(n^{-1} \prod_{i=2}^k \|\psi_i\|\right) \operatorname{rng}(\psi_1) \\ &\times \left(\pi_d\left(\sum_{i=2}^k (r_1 + r_i)^d\right)\right. \\ &\left. + \phi_1\left(k - 1 + \left(\sum_{i=2}^k \phi_i\right)\left(\frac{2n - \sum_{i=2}^k \phi_i}{n - \sum_{i=2}^k \phi_i}\right)\right)\right).\end{aligned}$$

PROOF. Given $\mathbf{x} = (x_1, \dots, x_k) \in C_n^k$, define the set of points

$$\mathcal{U}_n^{\mathbf{x}} := \{x_1, \dots, x_k, U_{k+1}, \dots, U_n\}.$$

Let F_n be the set of $\mathbf{x} = (x_1, \dots, x_k) \in C_n^k$ such that $D(x_1, x_i) > r_1 + r_i$ for $i \in \{2, \dots, k\}$, and let $F_n^c := C_n^k \setminus F_n$. Then by the law of total probability,

$$\begin{aligned}\mathbb{E}\left[\prod_{i=1}^k \psi_i(U_i, \mathcal{U}_n)\right] ={}& n^{-k} \int_{F_n} \mathbb{E} \prod_{i=1}^k \psi_i(x_i, \mathcal{U}_n^{\mathbf{x}})\, d\mathbf{x} \\ &+ n^{-k} \int_{F_n^c} \mathbb{E} \prod_{i=1}^k \psi_i(x_i, \mathcal{U}_n^{\mathbf{x}})\, d\mathbf{x}.\end{aligned}$$

Since $\mathbb{E}[\psi_1(U_1, \mathcal{U}_n)] = 0$ it follows that $\|\psi_1\| \le \mathrm{rng}(\psi_1)$, so that

$$\left| n^{-k} \int_{F_n^c} \mathbb{E} \prod_{i=1}^k \psi_i(x_i, \mathcal{U}_n^{\mathbf{x}})\, d\mathbf{x} \right| \le \left(\prod_{i=1}^k \|\psi_i\| \right) P[(U_1, \ldots, U_k) \in F_n^c] \tag{3.11}$$
$$\le \mathrm{rng}(\psi_1) \left(\prod_{i=2}^k \|\psi_i\| \right) \sum_{i=2}^k \pi_d (r_1 + r_i)^d / n.$$

Fix $\mathbf{x} = (x_1, \ldots, x_k) \in F_n$. For $m \in \mathbb{Z}_+$, let $h_1(m) := \mathbb{E}\psi_1(x_1, \{x_1\} \cup \mathcal{Y}_m)$, where $\mathcal{Y}_m$ denotes a collection of m uniformly distributed points in $B_{r_1}(x_1)$. Let $h_2(m) := \mathbb{E} \prod_{i=2}^k \psi_i(x_i, \{x_2, \ldots, x_k\} \cup \mathcal{Y}'_m)$, where $\mathcal{Y}'_m$ denotes a collection of m uniformly distributed points in $\bigcup_{i=2}^k B_{r_i}(x_i)$.

If N_1 and N_2 denote the number of points of $\{U_{k+1}, \ldots, U_n\}$ in $B_{r_1}(x_1)$ and in $\bigcup_{i=2}^k B_{r_i}(x_i)$, respectively, then the values of $\psi_1(x_1, \mathcal{U}_n^{\mathbf{x}})$ and of $\prod_{i=2}^k \psi_i(x_i, \mathcal{U}_n^{\mathbf{x}})$ are conditionally independent, given (N_1, N_2), because the regions $B_{r_1}(x_1)$ and $\bigcup_{i=2}^k B_{r_i}(x_i)$ are disjoint since we assume $\mathbf{x} \in F_n$. Hence, we assert that

$$\mathbb{E}\left[\prod_{i=1}^k \psi_i(x_i, \mathcal{U}_n^{\mathbf{x}}) \right] = \mathbb{E}[h_1(N_1) h_2(N_2)], \tag{3.12}$$

where (N_1, N_2, N_3) have the multinomial distribution

$$(N_1, N_2, N_3) \sim \mathrm{Mult}\left(n - k; \frac{a_1}{n}, \frac{a_2}{n}, \frac{a_3}{n} \right) \tag{3.13}$$

with a_1 denoting the volume of a ball of radius r_1 in C_n [so that $a_1 \le \phi_1 = \pi_d r_1^d$ with equality if $r_1 \le (1/2) n^{1/d}$] while a_2 is the volume of $\bigcup_{i=2}^k B_{r_i}(x_i)$ in C_n and $a_3 := n - a_1 - a_2$. To verify (3.12), use the law of total probability to decompose the left-hand side as a sum over possible values of (N_1, N_2).

Also, if $\tilde{N}_1 \sim \mathrm{Bin}(n-1, \frac{a_1}{n})$ then $\mathbb{E}[h_1(\tilde{N}_1)] = 0$, because of the assumption that $\mathbb{E}[\psi_1(U_1, \mathcal{U}_n)] = 0$, along with translation invariance; the value of $\mathbb{E}[h_1(\tilde{N}_1)]$ does not depend on x_1.

We give a coupling of N_1 to another random variable N'_1 with the same distribution as $\tilde{N}_1$ that is independent of N_2, for which we can give a useful bound on $P[N_1 \ne N'_1]$.

Consider throwing a series of colored balls so each ball can land in one of three urns, where the probability of landing in urn i is a_i/n for $1 \le i \le 3$. First, throw $n-k$ white balls and let N_1^*, N_2, N_3^* denote the number of white balls in urn i for $i = 1, 2, 3$, respectively, that is, let (N_1^*, N_2, N_3^*) have the $\mathrm{Mult}(n-k; \frac{a_1}{n}, \frac{a_2}{n}, \frac{a_3}{n})$ distribution. Now pick out the $n - k - N_2$ balls in urns 1 and 3, paint them red, and throw them again; that is, given the values of N_1^*, N_2, N_3^* let N_1^r, N_2^r, N_3^r count the number of red balls in urns 1, 2, 3, respectively, and so be nonnegative integer valued variables such that

$$\mathcal{L}((N_1^r, N_2^r, N_3^r) | N_1^*, N_2) = \mathrm{Mult}\left(n - k - N_2; \frac{a_1}{n}, \frac{a_2}{n}, \frac{a_3}{n} \right).$$

Now take the N_2^r red balls in urn 2, paint them blue, and throw them again but condition them to land in urns 1 and 3 (or equivalently, throw each blue ball again and again until it avoids urn 2), so that

$$\mathcal{L}((N_1^b, N_3^b)|N_1^*, N_2, N_1^r, N_2^r) = \mathrm{Mult}\left(N_2^r; \frac{a_1}{a_1 + a_3}, \frac{a_3}{a_1 + a_3}\right).$$

Finally, throw $k - 1 + N_2$ green balls, making the total number of green, red and blue balls $n - 1$, and record how many land in urn 1, so

$$\mathcal{L}(N_1^g|N_1^*, N_2, N_1^r, N_2^r, N_1^b) = \mathrm{Bin}\left(k - 1 + N_2; \frac{a_1}{n}\right).$$

Now set

$$N_1 = N_1^r + N_1^b, \qquad N_3 = N_3^r + N_3^b \quad \text{and} \quad N_1' = N_1^r + N_1^g.$$

Then (N_1, N_2, N_3) have the multinomial distribution given by (3.13). Also, $N_1' \sim \mathrm{Bin}(n - 1, \frac{a_1}{n})$ and N_1' is independent of N_2.

Since $N_1' = N_1 - N_1^b + N_1^g$, we have that

$$\begin{aligned} P[N_1 \neq N_1'] &\leq \mathbb{E}[N_1^g] + \mathbb{E}[N_1^b] \leq \frac{a_1}{n}(k - 1 + \mathbb{E}N_2) + \left(\frac{a_1}{a_1 + a_3}\right)\mathbb{E}[N_2^r] \\ &\leq \frac{a_1}{n}(k - 1 + a_2) + \left(\frac{a_1}{n - a_2}\right)a_2 \end{aligned}$$

so that

$$\left|\mathbb{E}\big[h_2(N_2)\big(h_1(N_1) - h_1(N_1')\big)\big]\right| \leq \frac{a_1}{n}\left(k - 1 + a_2 + \left(\frac{na_2}{n - a_2}\right)\right)\mathrm{rng}(\psi_1)\prod_{i=2}^{k}\|\psi_i\|$$

and since N_1' is independent of N_2 with $N_1' \sim \mathrm{Bin}(n - 1, \frac{a_1}{n})$,

$$\mathbb{E}[h_1(N_1')h_2(N_2)] = 0,$$

so by (3.12) and the fact that $a_1 \leq \phi_1$ and $a_2 \leq \sum_{i=2}^k \phi_i$ and the assumption that $\sum_{i=2}^k \phi_i < n$,

$$\begin{aligned} &\left|\mathbb{E}\left[\prod_{i=1}^{k}\psi_i(x_i, \mathcal{U}_n^{\mathbf{x}})\right]\right| \\ &\quad \leq \frac{a_1}{n}\left(k - 1 + a_2\left(\frac{2n - a_2}{n - a_2}\right)\right)\mathrm{rng}(\psi_1)\prod_{i=2}^{k}\|\psi_i\| \\ &\quad \leq \frac{\phi_1}{n}\left(k - 1 + \left(\sum_{i=2}^{k}\phi_i\right)\left(\frac{2n - \sum_{i=2}^k \phi_i}{n - \sum_{i=2}^k \phi_i}\right)\right)\mathrm{rng}(\psi_1)\prod_{i=2}^{k}\|\psi_i\|. \end{aligned}$$

The preceding bound holds uniformly over all possible values of $\mathbf{x} = (x_1, \dots, x_k) \in F_n$. Combined with (3.11), this shows that the asserted bound holds. □

LEMMA 3.5. *Suppose ψ_1 is as defined in Lemma 3.4. Then with notation from that result, if $\phi_1 < n$ then*

$$\operatorname{Var}\left[\frac{1}{n}\sum_{i=1}^{n}\psi_1(U_i,\mathcal{U}_n)\right] \\ \leq \frac{\|\psi_1\|^2}{n}(1+2^d\phi_1) + \frac{\|\psi_1\|}{n}\left(\phi_1+\phi_1^2\left(\frac{2n-\phi_1}{n-\phi_1}\right)\right)\operatorname{rng}(\psi_1).$$

PROOF. By the case $k=2$ of Lemma 3.4,

$$\begin{aligned}\operatorname{Cov}&(\psi_1(U_1,\mathcal{U}_n),\psi_1(U_2,\mathcal{U}_n))\\ &=\mathbb{E}[\psi_1(U_1,\mathcal{U}_n)\psi_1(U_2,\mathcal{U}_n)]\\ &\leq \frac{\|\psi_1\|}{n}\left(2^d\phi_1\|\psi_1\|+\phi_1\left(1+\phi_1\left(\frac{2n-\phi_1}{n-\phi_1}\right)\right)\operatorname{rng}(\psi_1)\right)\end{aligned}$$

and since

$$\operatorname{Var}\left[\frac{1}{n}\sum_{i=1}^{n}\psi_1(U_i,\mathcal{U}_n)\right]=n^{-1}\operatorname{Var}[\psi_1(U_1,\mathcal{U}_n)] \\ +\frac{n-1}{n}\operatorname{Cov}(\psi_1(U_1,\mathcal{U}_n),\psi_1(U_2,\mathcal{U}_n)),$$

the result follows. □

4. Size-biased coupling constructions. We now give a simple lemma which shows how to size-bias a random variable that can be expressed as a conditional probability of an event arising from some further randomization.

LEMMA 4.1. *Suppose Y is a random variable given by $Y=aP[A|\mathcal{F}]$, where $\mathcal{F}$ is some σ-algebra, $a>0$ is a constant, and A is an event with $0<P[A]<1$. Then Y' has the Y size biased distribution if*

$$\mathcal{L}(Y')=\mathcal{L}(Y|A). \tag{4.1}$$

PROOF. With $\mathcal{L}(Y')$ defined by (4.1), we must show for all bounded and continuous $g:\mathbb{R}\to\mathbb{R}$, that $\mathbb{E}[g(Y')]=\mathbb{E}[Yg(Y)]/\mathbb{E}[Y]$ [see (1.1)]. But

$$\begin{aligned}\mathbb{E}[g(Y')]&=\mathbb{E}[g(Y)|A]=\mathbb{E}[g(Y)\mathbf{1}_A]/P[A]\\ &=\mathbb{E}[g(Y)P[A|\mathcal{F}]]/P[A],\end{aligned}$$

where the last equality follows because $g(Y)$ is $\mathcal{F}$-measurable. The last expression equals $\mathbb{E}[Yg(Y)]/\mathbb{E}[Y]$, as required. □

Let V and S be given by (2.1), (2.2), respectively. Set $W = n - S$ (the number of nonsingletons). We assert that either V or W can be expressed as n times the conditional probability of some event A, given the locations of the points of $\mathcal{U}_n$, so that Lemma 4.1 is applicable. For V, take $A = A_V$ to be the event that an additional uniformly distributed random point U_0 in C_n lies in the covered region $\bigcup_{i=1}^n B_{i,\rho}$. For W, take $A = A_W$ to be the event that an element of $\mathcal{U}_n$, selected uniformly at random, is nonisolated.

Event A_V can be written as the event that $N_V > 0$, where N_V denotes the number of points of $\mathcal{U}_n$ in $B_\rho(U_0)$, and $N_V \sim \mathrm{Bin}(n, \phi/n)$ (recall $\phi := \pi_d \rho^d$ and C_n has volume n). A point set (denoted $\mathcal{U}_V$) with the conditional distribution of $\mathcal{U}_n$ given N_V can be obtained as follows:

I. Sample a uniform random point in C_n, denoted U_0.
II. Set $m = n$. Sample $N = N_V$ independent uniform random points in $B_\rho(U_0)$, and $m - N$ independent uniform random points in $C_n \setminus B_\rho(U_0)$.
III. Let $\mathcal{U}_V$ be the union of the two samples of uniform points.

Therefore, coupled realizations of $\mathcal{U}_V$ and $\mathcal{U}'_V$ (having, respectively, the distribution of $\mathcal{U}_n$ and the conditional distribution of $\mathcal{U}_n$ given $N_V > 0$), and hence coupled realizations of V and V', can be obtained as follows.

1. Set $m = n$.
2. Sample U_0 uniformly at random over C_n.
3. Sample m random d-vectors independently and uniformly over C_n, and denote this point set by $\mathcal{U}_{m,1}$.
4. Let N denote the number of points of $\mathcal{U}_{m,1}$ in $B_\rho(U_0)$.
5. Sample a Bernoulli random variable $\mathcal{B}$ with $P[\mathcal{B} = 1] = \pi_N$, where $(\pi_k, k \ge 0)$ is given by (3.9).
6. Sample a random d-vector U which is uniform over $B_\rho(U_0)$.
7. If $\mathcal{B} = 1$, then select one of the points of $\mathcal{U}_{m,1}$ uniformly at random, and move it to U. Denote the resulting modification of $\mathcal{U}_{m,1}$ by $\mathcal{U}_{m,2}$. If $\mathcal{B} = 0$ then set $\mathcal{U}_{m,2} := \mathcal{U}_{m,1}$.
8. Set $\mathcal{U}_V := \mathcal{U}_{m,1}$ and $\mathcal{U}'_V := \mathcal{U}_{m,2}$. Set $V := g_V(\mathcal{U}_V)$ and $V' := g_V(\mathcal{U}'_V)$, where $g_V(\mathcal{U}) := \mathrm{Vol}(\bigcup_{x \in \mathcal{U}} B_\rho(x))$.

By Lemma 3.3, the number of points of $\mathcal{U}_{m,2}$ in $\mathcal{B}_\rho(U_0)$ has the distribution $\mathcal{L}(N_V | N_V > 0)$, and hence $\mathcal{L}(\mathcal{U}'_V) = \mathcal{L}(\mathcal{U}_V | N_V > 0)$. So by Lemma 4.1, V' has the V size biased distribution.

In the case of W, A_W is the event that $N_W > 0$, where N_W denotes the number of points of $\mathcal{U}_n \setminus \{U_0\}$ in $B_\rho(U_0)$, and now U_0 denotes a point of $\mathcal{U}_n$ selected uniformly at random. So $N_W \sim \mathrm{Bin}(n - 1, \phi/n)$. We can obtain a point set (denoted $\mathcal{U}_W$) with the conditional distribution of $\mathcal{U}_n$ given N_W by the same steps as for $\mathcal{U}_V$ except that now in Step II we put $m = n - 1$ and $N = N_W$, and in Step III, $\mathcal{U}_W$ is the union of the two samples of uniform random points with an added point at U_0. Hence, we can obtain coupled realizations of W and W' by the same sequence of steps as described above for (V, V'), except that the following steps are modified:

- In Step 1, set $m = n - 1$ (this affects Steps 3 and 5.)
- In Step 8, set $\mathcal{U}_W := \mathcal{U}_{m,1} \cup \{U_0\}$, and $\mathcal{U}'_W := \mathcal{U}_{m,2} \cup \{U_0\}$. Set $W := g_W(\mathcal{U}_W)$ and $W' := g_W(\mathcal{U}'_W)$ with $g_W(\mathcal{U}) := \sum_{x \in \mathcal{U}} \mathbf{1}\{\mathcal{U} \cap B_\rho(x) \neq \{x\}\}$.

By a similar argument to the V case, W' has the W size biased distribution.

5. Proof of Theorem 2.1. We couple V' to V as described in Section 4. Since V' differs from V through the moving of at most a single point, clearly $|V' - V| \leq \pi_d \rho^d := \phi$. Hence, by Lemma 3.1 with $B = \phi$, to prove Theorem 2.1 it suffices to prove the following.

PROPOSITION 5.1. *Under the assumptions of Theorem* 2.1, $\mathrm{Var}(\mathbb{E}[V' - V|V]) \leq n^{-1}\eta_V(n, \rho)$, *where* $\eta_V(n, \rho)$ *is given by* (2.4).

PROOF. Let $\mathcal{G}$ be the σ-algebra generated by the point set $\mathcal{U}_V$. List the points of $\mathcal{U}_V$, in an order chosen uniformly at random, as $U_1, \dots, U_n$, and set $\mathbf{U} := (U_1, \dots, U_n)$. Then V is $\mathcal{G}$-measurable. The conditional variance formula, with $X = \mathbb{E}[V' - V|\mathcal{G}]$, yields

$$\mathrm{Var}(\mathbb{E}[V' - V|V]) = \mathrm{Var}(\mathbb{E}[X|V]) \leq \mathrm{Var}(X),$$

so it suffices to prove

$$\mathrm{Var}(\mathbb{E}[V' - V|\mathcal{G}]) \leq n^{-1}\eta_V(n, \rho). \tag{5.1}$$

For $x \in C_n$, let ξ_x denote the probability that $\mathcal{B} = 1$, given $\mathcal{U}_n$ and given that $U_0 = x$, that is, $\xi_x = \pi_{N_x}$, where N_x denotes the number of points of $\mathcal{U}_V$ in $B_\rho(x)$. Let R_{xj} denote the expectation (over U) of the increment in the covered volume if U_j is moved to a uniform randomly selected location U in $B_\rho(x)$. Note that for x and j fixed, R_{xj} is determined by $\mathbf{U}$. Then, since both U_0 and I are independent of $\mathcal{G}$,

$$\mathbb{E}[V' - V|\mathcal{G}] = \frac{1}{n} \int_{C_n} \xi_x \left(\frac{1}{n} \sum_{j=1}^{n} R_{xj} \right) dx,$$

where the first factor of $1/n$ comes from the probability density of U_0, and the second arises as the probability that I takes the value j.

Let H_x be the expectation (over U) of the increment in the covered volume when a point is inserted into $\mathcal{U}_V$ at a uniform random location $U \in B_\rho(x)$, and let T_j be the increment in the covered volume when point U_j is removed from $\mathcal{U}$ (for fixed x and j, both H_x and T_j are determined by $\mathbf{U}$). If U_j is far distant from x then $R_{xj} = H_x + T_j$. Set $Q_{xj} := R_{xj} - H_x - T_j$, which is in fact the expectation (over U) of the total volume of the otherwise uncovered regions lying within distance ρ

both of U and of U_j (such regions contribute to T_j but not to H_x or R_{xj}). Then

$$\begin{aligned}\mathbb{E}[V' - V|\mathcal{G}] &= \frac{1}{n^2}\int_{C_n}\sum_{j=1}^{n}\xi_x(H_x + T_j + Q_{xj})\,dx \\ &= \frac{1}{n}\int_{C_n}\xi_x\left(H_x + \frac{1}{n}\sum_{j=1}^{n}Q_{xj}\right)dx + \frac{1}{n^2}\int_{C_n}\sum_{j=1}^{n}\xi_x T_j\,dx.\end{aligned} \tag{5.2}$$

Set $\phi := \pi_d \rho^d$. We have that $0 \le H_x \le \phi$, $0 \ge T_j \ge -\phi$ and if $D(x, U_j) > 3\rho$ then $Q_{xj} = 0$. Moreover, if $D(x, U_j) > 3\rho$ for all $j \in \{1, \ldots, n\}$, then $H_x = \phi$ and if $D(x, U_j) > \rho$ for all $j \in \{1, \ldots, n\}$, then $\xi_x = 1$. Finally, $Q_{xj} \ge 0$ and

$$0 \le H_x + \frac{1}{n}\sum_{j=1}^{n}Q_{xj} \le H_x + \sum_{j=1}^{n}Q_{xj} \le \phi.$$

Hence setting

$$\tau_x := \xi_x\left(H_x + \frac{1}{n}\sum_{j=1}^{n}Q_{xj}\right) - \phi,$$

we have that $-\phi \le \tau_x \le 0$, and τ_x is determined by the collection of points of $\mathcal{U}_n$ within distance 3ρ of x, and $\tau_x = 0$ if there are no such points of $\mathcal{U}_n$. We can rewrite (5.2) as

$$\mathbb{E}[V' - V|\mathcal{G}] = \phi + \frac{1}{n}\int_{C_n}\tau_x\,dx + \frac{1}{n}\left(\sum_{j=1}^{n}T_j\right) + \frac{1}{n^2}\int_{C_n}\sum_{j=1}^{n}(\xi_x - 1)T_j\,dx.$$

Recalling that $B_r(x) := \{y \in C_n : D(x, y) \le r\}$, let $\Gamma_{i,r}$ be the set of points $y \in B_r(U_i)$ such that $D(y, U_i) < D(y, U_j)$ for all $j \in \{1, \ldots, n\} \setminus \{i\}$ (i.e., the intersection of the r-ball around U_i and the Voronoi cell of U_i relative to $\mathcal{U}_n$). Set

$$S_i' := \int_{\Gamma_{i,3\rho}}\tau_x\,dx, \qquad S_i'' := \int_{\Gamma_{i,\rho}}(\xi_x - 1)\,dx.$$

Then

$$\begin{aligned}\mathbb{E}[V' - V|\mathcal{G}] &= \phi + \left(\frac{1}{n}\sum_{i=1}^{n}S_i'\right) + \left(\frac{1}{n}\sum_{j=1}^{n}T_j\right) + \left(\frac{1}{n^2}\sum_{i=1}^{n}S_i''T_i\right) \\ &\quad + \left(\frac{1}{n^2}\sum_{(i,j):i\ne j}S_i''T_j\right),\end{aligned} \tag{5.3}$$

and if we put $b = \mathbb{E}T_i$ (which does not depend on i), we have

$$\frac{1}{n^2}\sum_{(i,j):i\ne j}S_i''T_j = \frac{1}{n^2}\left(\sum_{(i,j):i\ne j}S_i''(T_j - b)\right) + \frac{b(n-1)}{n^2}\left(\sum_{i=1}^{n}S_i''\right),$$

so by (5.3),

$$\begin{aligned}
&\mathbb{E}[V' - V|\mathcal{G}] \\
&\quad = \phi + \frac{1}{n^2}\biggl(\sum_{(i,j):i\neq j} S_i''(T_j - b)\biggr) \\
&\qquad + \frac{1}{n}\sum_{i=1}^{n}\bigl(S_i' + T_i + \bigl(n^{-1}T_i + (1-n^{-1})b\bigr)S_i''\bigr).
\end{aligned}$$

Since $(x+y)^2 \le 2(x^2+y^2)$ for any real x, y,

$$\begin{aligned}
&\mathrm{Var}(\mathbb{E}[V' - V|\mathcal{G}]) \\
&\quad \le 2\,\mathrm{Var}\biggl(\frac{1}{n}\sum_{i=1}^{n}\bigl(S_i' + T_i + \bigl(n^{-1}T_i + (1-n^{-1})b\bigr)S_i''\bigr)\biggr) \\
&\qquad + 2\,\mathrm{Var}\biggl(\frac{1}{n^2}\biggl(\sum_{(i,j):i\neq j} S_i''(T_j - b)\biggr)\biggr).
\end{aligned} \tag{5.4}$$

Table 1 summarizes upper and lower bounds and the radius of the relevant variables, where the *radius* of a variable indexed by i is the smallest distance from U_i one needs to look to establish its value (as with the functionals considered in Lemma 3.4).

Hence, the variable

$$S_i' + T_i + \bigl(n^{-1}T_i + (1-n^{-1})b\bigr)S_i''$$

has radius 6ρ relative to U_i and lies between $-\phi - 3^d\phi^2$ and ϕ^2, so that its centered value is bounded in absolute value by $(3^d+1)\phi^2 + \phi$, and this also bounds its range of possible values. So by Lemma 3.5 and the assumption that $6^d\phi < n$,

$$\begin{aligned}
&\mathrm{Var}\biggl(\frac{1}{n}\sum_{i=1}^{n}(S_i' + T_i + (n^{-1}T_i + (1-n^{-1})b)S_i'')\biggr) \\
&\quad \le \frac{\phi^2((3^d+1)\phi+1)^2}{n}\biggl(1 + (2^d+1)6^d\phi + \biggl(\frac{2n - 6^d\phi}{n - 6^d\phi}\biggr)6^{2d}\phi^2\biggr).
\end{aligned} \tag{5.5}$$

TABLE 1
Radii and bounds for covered volume

Variable	τ_x	T_i	S_i''	S_i'	$(n^{-1}T_i + (1-n^{-1})b)S_i''$
Radius	3ρ	2ρ	2ρ	6ρ	2ρ
Lower bound	$-\phi$	$-\phi$	$-\phi$	$-3^d\phi^2$	0
Upper bound	0	0	0	0	ϕ^2

Note: The last two columns are deduced from the previous columns.

Now consider the last term in the right-hand side of (5.4). Set $\bar{T}_j := T_j - b$. Then

$$
\begin{aligned}
&\operatorname{Var}\Big(\sum_{(i,j):i\neq j} S_i''\bar{T}_j\Big)\\
&\quad = n(n-1)(n-2)(n-3)\operatorname{Cov}(S_1''\bar{T}_2, S_3''\bar{T}_4)\\
&\qquad + n(n-1)(n-2)\big(\operatorname{Cov}(S_1''\bar{T}_2, S_1''\bar{T}_3)\\
&\qquad\qquad + \operatorname{Cov}(S_2''\bar{T}_1, S_3''\bar{T}_1) + 2\operatorname{Cov}(S_1''\bar{T}_2, S_3''\bar{T}_1)\big)\\
&\qquad + n(n-1)\big(\operatorname{Var}(S_1''\bar{T}_2) + \operatorname{Cov}(S_1''\bar{T}_2, S_2''\bar{T}_1)\big).
\end{aligned}
\tag{5.6}
$$

It follows from the case $k=4$ of Lemma 3.4 and the assumption $6^d\phi < n$ (which implies $3(2^d\phi) < n$) that

$$
\begin{aligned}
\operatorname{Cov}(S_1''\bar{T}_2, S_3''\bar{T}_4) &= \mathbb{E}[S_1''\bar{T}_2S_3''\bar{T}_4] - (\mathbb{E}[S_1''\bar{T}_2])^2 \le \mathbb{E}[S_1''\bar{T}_2S_3''\bar{T}_4]\\
&\le \frac{\phi^4}{n}\bigg(3\pi_d(4\rho)^d + 2^d\phi\bigg(3 + 3(2^d)\phi\bigg(\frac{2n-3(2^d)\phi}{n-3(2^d)\phi}\bigg)\bigg)\bigg)\\
&= \frac{3\phi^4}{n}\bigg((4^d+2^d)\phi + 4^d\phi^2\bigg(\frac{2n-3(2^d)\phi}{n-3(2^d)\phi}\bigg)\bigg).
\end{aligned}
$$

Since we can always bound $\operatorname{Cov}(S_i''\bar{T}_j, S_{i'}''\bar{T}_{j'})$ above by ϕ^4, we have from (5.6) that

$$
\begin{aligned}
&\operatorname{Var}\bigg(\frac{1}{n^2}\bigg(\sum_{(i,j):i\neq j} S_i''(T_j - b)\bigg)\bigg)\\
&\quad\le \frac{\phi^4}{n}\bigg(3(4^d+2^d)\phi + 3(4^d)\phi^2\bigg(\frac{2n-3(2^d)\phi}{n-3(2^d)\phi}\bigg) + 4 + \frac{2}{n}\bigg).
\end{aligned}
\tag{5.7}
$$

By (5.4), (5.5) and (5.7) we have that

$$
\begin{aligned}
&(n/2)\operatorname{Var}(\mathbb{E}[V'-V|\mathcal{G}])\\
&\quad\le \phi^2((3^d+1)\phi+1)^2\bigg(1 + (2^d+1)6^d\phi + \bigg(\frac{2n-6^d\phi}{n-6^d\phi}\bigg)6^{2d}\phi^2\bigg)\\
&\qquad + \phi^4\bigg(3(4^d+2^d)\phi + 3(4^d)\phi^2\bigg(\frac{2n-3(2^d)\phi}{n-3(2^d)\phi}\bigg) + 4 + \frac{2}{n}\bigg).
\end{aligned}
$$

This completes the proof of Proposition 6.1, and hence of Theorem 2.1. $\square$

6. Proof of Theorem 2.2. We couple W' to W as described in Section 4. Thus $W = g_W(\mathcal{U}_W)$ and $W' = g_W(\mathcal{U}_W')$, where $\mathcal{U}_W'$ is obtained from $\mathcal{U}_W$ by moving at most a single randomly selected point of $\mathcal{U}_W \setminus \{U_0\}$ to a (uniform random) location in $B_\rho(U_0)$, if $\mathcal{B}=1$, and leaving $\mathcal{U}_W$ unchanged if $\mathcal{B}=0$.

The number of points that can be made isolated by removing a single point from $\mathcal{U}_W$ is almost surely bounded by κ_d. Moreover, the number of points that can be made nonisolated by inserting a point (including the inserted point itself) is almost surely bounded by $\kappa_d + 1$. Hence $|W - W'| \le \kappa_d + 1$, so we may take $B = \kappa_d + 1$.

By the symmetry of the normal distribution, $D_{-S} = D_S$ and hence $D_W = D_S$. Thus, Theorem 2.2 follows from Lemma 3.1 along with the following:

PROPOSITION 6.1. *Under the assumptions of Theorem* 2.2, $\mathrm{Var}(\mathbb{E}[W' - W|W]) \le n^{-1}\eta_S(n,\rho)$, *where* $\eta_S(n,\rho)$ *is given by* (2.5).

PROOF. Here we let $\mathcal{G}$ denote the σ-algebra generated by the unlabelled point set $\mathcal{U} := \mathcal{U}_W$. Then W' is $\mathcal{G}$-measurable, and by the conditional variance formula (as in the proof of Proposition 5.1), it suffices to prove that

$$\mathrm{Var}(\mathbb{E}[W' - W|\mathcal{G}]) \le n^{-1}\eta_S(n,\rho). \tag{6.1}$$

Label the points of $\mathcal{U}$, in an order chosen uniformly at random, as $U_1, \dots, U_n$, and set $\mathbf{U} := (U_1, \dots, U_n)$. $\xi_i = \pi_{N_i}$, where N_i denotes the number of points of $\mathcal{U} \setminus \{U_i\}$ in $B_\rho(U_i)$. Let R_{ij} denote the expectation (over U) of the increment in the number of nonisolated points when U_j is moved to a uniform randomly selected location U in $B_\rho(U_i)$. Then

$$\mathbb{E}[W' - W|\mathcal{G}] = \frac{1}{n(n-1)} \sum_{(i,j)\,:\, i\neq j} \xi_i R_{ij},$$

where $\sum_{(i,j)\,:\, i\neq j}$ denotes summation over pairs of distinct integers i, j in $[1,n]$.

Now let H_i be the expectation (over U) of the increment in the number of isolated points when a point is inserted into $\mathcal{U}$ at a uniform random location $U \in B_\rho(U_i)$, and let T_j be the increment in the number of isolated points when point U_j is removed from $\mathcal{U}$ (both H_i and T_j are determined by $\mathbf{U}$). If U_j is far distant from U_i then $R_{ij} = -H_i - T_j$. In fact, setting $Q_{ij} := R_{ij} + H_i + T_j$, we have that Q_{ij} is the expectation (over U) of the number of otherwise isolated points of $\mathcal{U}$ within distance ρ both of U and of U_j (such points contribute to T_j but not to H_i or R_{ij}). Then

$$\begin{aligned}\mathbb{E}[W' - W|\mathcal{G}] &= \frac{1}{n(n-1)} \sum_{(i,j)\,:\, i\neq j} \xi_i(-H_i - T_j + Q_{ij}) \\ &= \frac{1}{n}\sum_{i=1}^n \xi_i \tau_i - \frac{1}{n(n-1)} \sum_{(i,j)\,:\, i\neq j} \xi_i T_j,\end{aligned} \tag{6.2}$$

where we set

$$\tau_i := -H_i + \frac{1}{n-1} \sum_{j\,:\, j\neq i} Q_{ij}. \tag{6.3}$$

Put $a := \mathbb{E}[\xi_i]$ (given n, this expectation does not depend on i) and put $b := (\kappa_d - 1)/2$. Then

$$\frac{1}{n(n-1)} \sum_{(i,j):i\neq j} \xi_i T_j = \frac{1}{n(n-1)}\biggl(\sum_{(i,j):i\neq j} (\xi_i - a)(T_j - b)\biggr) + \frac{a}{n}\Biggl(\sum_{j=1}^{n} T_j\Biggr) + \frac{b}{n}\Biggl(\sum_{i=1}^{n}(\xi_i - a)\Biggr).$$

Hence we can rewrite (6.2) as

$$\mathbb{E}[W' - W|\mathcal{G}] = \frac{1}{n}\sum_{i=1}^{n}\bigl(\xi_i\tau_i - aT_i - b(\xi_i - a)\bigr) - \frac{1}{n(n-1)} \sum_{(i,j):i\neq j} (\xi_i - a)(T_j - b).$$

Since $(x+y)^2 \le 2(x^2+y^2)$ for any real x, y, it follows that

$$\begin{aligned}\operatorname{Var}(\mathbb{E}[W' - W|\mathcal{G}]) &\le 2\operatorname{Var}\Biggl(\frac{1}{n}\sum_{i=1}^{n}\bigl(\xi_i(\tau_i - b) + a(b - T_i)\bigr)\Biggr)\\ &\quad + 2\operatorname{Var}\biggl(\frac{1}{n(n-1)} \sum_{(i,j):i\neq j} (\xi_i - a)(T_j - b)\biggr).\end{aligned} \tag{6.4}$$

We have that $-\kappa_d \le H_i \le 0$, $-1 \le T_j \le \kappa_d$, and $Q_{ij} \ge 0$. Also,

$$0 \le -H_i + \sum_{j:j\neq i} Q_{ij} \le \kappa_d,$$

and if $D(U_i, U_j) > 3\rho$ then $Q_{ij} = 0$. Hence, $0 \le \tau_i \le \kappa_d$, and τ_i is determined by the collection of points of $\mathcal{U}$ within distance 3ρ of U_i. Table 2 summarizes this discussion; recall from Table 1 the notion of radius.

From the last column in this table, we see that after centering, the terms in first sum in the right-hand side of (6.4) have radius 3ρ and absolute values bounded by $1 + 2\kappa_d$. Moreover, even after centering each of these terms has range (i.e., essential supremum minus essential infimum) which is also bounded by $1 + 2\kappa_d$

TABLE 2
Radii and bounds for singletons

Variable	$\boldsymbol{H_i}$	$\boldsymbol{T_i}$	$\boldsymbol{\xi_i}$	$\boldsymbol{\tau_i}$	$\boldsymbol{\xi_i(\tau_i - b)}$	$\boldsymbol{a(b - T_i)}$	$\boldsymbol{\xi_i(\tau_i - b) + a(b - T_i)}$
Radius	3ρ	2ρ	ρ	3ρ	3ρ	2ρ	3ρ
ess inf	$-\kappa_d$	-1	0	0	$(1-\kappa_d)/2$	$-(\kappa_d+1)/2$	$-\kappa_d$
ess sup	0	κ_d	1	κ_d	$(\kappa_d+1)/2$	$(\kappa_d+1)/2$	κ_d+1

Note: The last three columns are deduced from the preceding columns and the definitions of a, b.

(this range is unaffected by the centering). Hence with $\phi := \pi_d \rho^d$, Lemma 3.5, using the assumption $3^d\phi < n$, yields

$$\begin{aligned}
&\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\big(\xi_i(\tau_i - b) + a(b - T_i)\big)\right) \\
&\qquad \le \frac{(1+2\kappa_d)^2}{n}\left(1 + (2^d+1)3^d\phi + \left(\frac{2n - 3^d\phi}{n - 3^d\phi}\right)(3^d\phi)^2\right).
\end{aligned} \tag{6.5}$$

Now consider the second sum in the right-hand side of (6.4). Set $\bar{\xi}_i := \xi_i - a$ and $\bar{T}_j := T_j - b$. Then

$$\begin{aligned}
&\operatorname{Var}\left(\sum_{(i,j):i\neq j}(\xi_i - a)(T_j - b)\right) \\
&\quad = n(n-1)(n-2)(n-3)\operatorname{Cov}(\bar{\xi}_1\bar{T}_2, \bar{\xi}_3\bar{T}_4) \\
&\qquad + n(n-1)(n-2)\big(\operatorname{Cov}(\bar{\xi}_1\bar{T}_2, \bar{\xi}_1\bar{T}_3) \\
&\qquad\qquad + \operatorname{Cov}(\bar{\xi}_2\bar{T}_1, \bar{\xi}_3\bar{T}_1) + 2\operatorname{Cov}(\bar{\xi}_1\bar{T}_2, \bar{\xi}_3\bar{T}_1)\big) \\
&\qquad + n(n-1)\big(\operatorname{Var}(\bar{\xi}_1\bar{T}_2) + \operatorname{Cov}(\bar{\xi}_1\bar{T}_2, \bar{\xi}_2\bar{T}_1)\big).
\end{aligned} \tag{6.6}$$

Note that $\bar{\xi}_i$ has absolute value bounded by 1, and range of possible values also bounded by 1, and mean zero. Also, $\bar{T}_j$ has absolute value almost surely bounded by $(\kappa_d + 1)/2$ (its mean might not be zero). Hence, the case $k = 4$ of Lemma 3.4 [taking $r_1 = r_2 = \rho$ and $r_3 = r_4 = 2\rho$ so that $\phi_2 + \phi_3 + \phi_4 = (2^{d+1} + 1)\phi$] yields

$$\begin{aligned}
&\operatorname{Cov}(\bar{\xi}_1\bar{T}_2, \bar{\xi}_3\bar{T}_4) \\
&\quad = \mathbb{E}[\bar{\xi}_1\bar{T}_2\bar{\xi}_3\bar{T}_4] - (\mathbb{E}[\bar{\xi}_1\bar{T}_2])^2 \le \mathbb{E}[\bar{\xi}_1\bar{T}_2\bar{\xi}_3\bar{T}_4] \\
&\quad \le \frac{(\kappa_d+1)^2}{4n}\left(\phi\big(2(3^d) + 2^d\big) + 3\phi + (2^{d+1}+1)\phi^2\left(\frac{2n - (2^{d+1}+1)\phi}{n - (2^{d+1}+1)\phi}\right)\right),
\end{aligned}$$

where we have also used the assumption that $(2^{d+1} + 1)\phi < n$. Since we can always bound $\operatorname{Cov}(\bar{\xi}_i\bar{T}_j, \bar{\xi}_{i'}\bar{T}_{j'})$ by $((\kappa_d + 1)/2)^2$, we have from (6.6) that

$$\begin{aligned}
&\operatorname{Var}\left(\frac{1}{n(n-1)}\left(\sum_{(i,j):i\neq j}(\xi_i - a)(T_j - b)\right)\right) \\
&\qquad \le \frac{(\kappa_d+1)^2}{4n}\left(\big(2(3^d) + 2^d + 3\big)\phi \right.\\
&\qquad\qquad \left. + (2^{d+1}+1)\phi^2\left(\frac{2n - (2^{d+1}+1)\phi}{n - (2^{d+1}+1)\phi}\right)\right) \\
&\qquad\quad + \left(\frac{\kappa_d+1}{2}\right)^2\left(\frac{4}{n} + \frac{2}{n(n-1)}\right).
\end{aligned} \tag{6.7}$$

By (6.4), (6.5) and (6.7), we have that

$$
\begin{aligned}
&n\,\mathrm{Var}(\mathbb{E}[W'-W|\mathcal{G}]) \\
&\quad \le 2(1+2\kappa_d)^2\biggl(1+(2^d+1)3^d\phi+\biggl(\frac{2n-3^d\phi}{n-3^d\phi}\biggr)9^d\phi^2\biggr) \\
&\qquad + \frac{(\kappa_d+1)^2}{2}\biggl((2(3^d)+2^d+3)\phi+(2^{d+1}+1)\biggl(\frac{2n-(2^{d+1}+1)\phi}{n-(2^{d+1}+1)\phi}\biggr)\phi^2\biggr) \\
&\qquad + \frac{(\kappa_d+1)^2}{2}\biggl(4+\frac{2}{n-1}\biggr).
\end{aligned}
$$

This completes the proof of Proposition 6.1, and hence of Theorem 2.2. □

7. Proof of Theorem 2.3 and numerics. Again set $\phi := \pi_d\rho^d$. It is easy to see that provided $2\rho < n^{1/d}$,

$$
\mathbb{E}[V] = n\bigl(1-(1-\phi/n)^n\bigr); \qquad \mathbb{E}[S] = n(1-\phi/n)^{n-1}, \tag{7.1}
$$

and (2.9) follows from this.

Write $|\cdot|$ for the Euclidean norm and recall that $\omega_d(|x|)$ denotes the volume of the union of unit balls centered at the origin $\mathbf{0}$ and at x. If I_x denotes the indicator of the event that x is not contained in any of the balls $B_{\rho,i}$, then provided $4\rho < n^{1/d}$ we have the exact formula

$$
\begin{aligned}
\mathrm{Var}(V) &= \mathrm{Var}(n-V) \\
&= \mathrm{Var}\int_{C_n} I_x\,dx \\
&= \int_{C_n}\int_{C_n}\mathbb{E}[I_xI_y]\,dx\,dy - \bigl(n(1-\phi/n)^n\bigr)^2 \\
&= n\int_{B_{2\rho}(\mathbf{0})}\biggl(1-\frac{\rho^d\omega_d(|y|/\rho)}{n}\biggr)^n dy \\
&\quad + n(n-2^d\phi)\biggl(1-\frac{2\phi}{n}\biggr)^n - n^2(1-\phi/n)^{2n}.
\end{aligned} \tag{7.2}
$$

PROOF OF (2.10). For asymptotics as $n\to\infty$ with ρ fixed, use the MacLaurin expansion of $\log(1-x)$ to obtain

$$
\begin{aligned}
\biggl(1-\frac{2\phi}{n}\biggr)^n &= e^{-2\phi}\exp\biggl(-\frac{2\phi^2}{n}+O(n^{-2})\biggr); \\
\biggl(1-\frac{\phi}{n}\biggr)^{2n} &= e^{-2\phi}\exp\biggl(-\frac{\phi^2}{n}+O(n^{-2})\biggr)
\end{aligned}
$$

so that

$$
\begin{aligned}
n^{-1}\operatorname{Var}(V) &= \int_{B_{2\rho}(\mathbf{0})}\left(1-\frac{\rho^d\omega_d(|y|/\rho)}{n}\right)^n dy \\
&\quad + ne^{-2\phi}\left(\left(1-\frac{2^d\phi}{n}\right)\exp\left(-\frac{2\phi^2}{n}\right)-\exp\left(-\frac{\phi^2}{n}\right)+O(n^{-2})\right) \\
&\to \left(\int_{B_{2\rho}(\mathbf{0})}\exp(-\rho^d\omega_d(|y|/\rho))\,dy\right)-e^{-2\phi}(2^d\phi+\phi^2)
\end{aligned}
$$

and this limit is equal to $g_V(\rho)$ as defined by (2.7), so the first part of (2.10) is proven.

It remains to show that $g_V(\rho)>0$. This can be done either by using the last part of Theorem 2.1 of [16], or directly. We leave it to the reader to check that the conditions of the last part of Theorem 2.1 are satisfied here, or to look up the direct argument which is in the first version of this paper (arXiv:0812.3084). Thus (2.10) holds in its entirety. □

The computations for S are somewhat similar. With X_i denoting the indicator of the event that U_i is isolated,

$$
\begin{aligned}
\operatorname{Var}(S) &= n\operatorname{Var}(X_1)+n(n-1)\operatorname{Cov}(X_1,X_2) \\
&= n(1-\phi/n)^{n-1}\left(1-(1-\phi/n)^{n-1}\right)+n(n-1)\operatorname{Cov}(X_1,X_2).
\end{aligned}
$$

Since $\operatorname{Cov}(X_1,X_2)=\mathbb{E}[X_1X_2]-\mathbb{E}[X_1]^2$, provided $4\rho<n^{1/d}$ we can write

$$
\begin{aligned}
\operatorname{Var}(S) &= n(1-\phi/n)^{n-1}\left(1-(1-\phi/n)^{n-1}\right) \\
&\quad + (n-1)\int_{B_{2\rho}(\mathbf{0})\setminus B_\rho(\mathbf{0})}\left(1-\frac{\rho^d\omega_d(|y|/\rho)}{n}\right)^{n-2}dy \\
&\quad + n(n-1)\left(\left(1-\frac{2^d\phi}{n}\right)\left(1-\frac{2\phi}{n}\right)^{n-2}-\left(1-\frac{\phi}{n}\right)^{2n-2}\right).
\end{aligned}
\tag{7.3}
$$

PROOF OF (2.11). For asymptotics as $n\to\infty$ with ρ fixed, by again using the MacLaurin expansion of $\log(1-x)$ we obtain

$$
\begin{aligned}
\left(1-\frac{2\phi}{n}\right)^{n-2} &= \exp\left((n-2)\left(-\frac{2\phi}{n}-\frac{2\phi^2}{n^2}+O(n^{-3})\right)\right) \\
&= \exp\left(-2\phi+\frac{4\phi-2\phi^2}{n}+O(n^{-2})\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\left(1-\frac{\phi}{n}\right)^{2n-2} &= \exp\left((2n-2)\left(-\frac{\phi}{n}-\frac{\phi^2}{2n^2}+O(n^{-3})\right)\right) \\
&= \exp\left(-2\phi+\frac{2\phi-\phi^2}{n}+O(n^{-2})\right)
\end{aligned}
$$

and hence the last term in the right-hand side of (7.3) is equal to

$$
\begin{aligned}
&n(n-1)\exp(-2\phi) \\
&\quad\times\left(\left(1-\frac{2^d\phi}{n}\right)\exp\left(\frac{4\phi-2\phi^2}{n}\right)-\exp\left(\frac{2\phi-\phi^2}{n}\right)+O(n^{-2})\right) \\
&\quad= n(n-1)\exp(-2\phi)\left(-\frac{2^d\phi}{n}+\frac{2\phi}{n}-\frac{\phi^2}{n}+O(n^{-2})\right),
\end{aligned}
$$

so that

$$
\begin{aligned}
&\lim_{n\to\infty} n^{-1}\operatorname{Var}(S) \\
&\quad= e^{-\phi}(1-e^{-\phi})-e^{-2\phi}\big((2^d-2)\phi+\phi^2\big)+\int_{B_{2\rho}(\mathbf{0})\setminus B_\rho(\mathbf{0})} e^{-\rho^d\omega_d(|y|/\rho)}\,dy \\
&\quad= e^{-\phi}-\big(1+(2^d-2)\phi+\phi^2\big)e^{-2\phi}+\rho^d\int_{B_2(\mathbf{0})\setminus B_1(\mathbf{0})} e^{-\rho^d\omega_d(|u|)}\,du,
\end{aligned}
$$

and since this limit is equal to $g_S(\rho)$ as defined by (2.8), we have proved the first part of (2.11), namely, convergence to $g_S(\rho)$.

To complete the proof of (2.11), we need to show that $g_S(\rho) > 0$. This can be done by the same arguments as for the proof of (2.10). Hence, (2.11) holds in its entirety. $\square$

PROOF OF THEOREM 2.3. It remains only to prove (2.12), (2.13) and (2.14). By definition $\eta_V(\rho) = \lim_{n\to\infty}\eta_V(n,\rho)$ and $\eta_S(\rho) = \lim_{n\to\infty}\eta_S(n,\rho)$. Then (2.12) follows at once from Theorem 2.1, along with (2.9) and (2.10). Similarly, (2.13) follows at once from Theorem 2.2 along with (2.9), and (2.11).

Finally, we demonstrate the asymptotic lower bound (2.14). For any random variable X, let F_X denote its cumulative distribution function and let f_X denote its probability density function (if it has one). Let $\varepsilon\in(0,1)$. Set

$$
t_1 := \frac{[\mu_S]-\mu_S}{\sigma_S};\qquad t_2 := \frac{[\mu_S]-\mu_S+1-\varepsilon}{\sigma_S}.
$$

Here $[\cdot]$ denotes integer part, so that $|t_i|\le\sigma_S^{-1}$ for $i=1,2$. By the unimodality of the standard normal density,

$$
\begin{aligned}
F_Z(t_2)-F_Z(t_1) &\ge (t_2-t_1)\min(f_Z(t_1), f_Z(t_2)) \\
&\ge (1-\varepsilon)\sigma_S^{-1} f_Z(\sigma_S^{-1}).
\end{aligned}
\tag{7.4}
$$

On the other hand, since S is integer-valued, $F_{(S-\mu_S)/\sigma_S}(t_1)$ is equal to $F_{(S-\mu_S)/\sigma_S}(t_2)$, so that by (7.4)

$$
D_S \ge (1/2)(1-\varepsilon)\sigma_S^{-1} f_Z(\sigma_S^{-1}).
$$

Scaling by $n^{1/2}$, letting $n\to\infty$, using (2.11) and letting $\varepsilon\to 0$ yields (2.14). $\square$

To conclude, we compute some numerical values for the asymptotic upper bounds appearing in (2.12) and (2.13). For this we need to compute $J_{r,d}(\rho)$ defined by (2.6) (for $r = 1$ and $r = 2$), and for this in turn, we need to compute $\omega_d(u)$, the volume of the union of two unit balls in d-space whose centers are at points (x, x' say) distance u apart ($u \le 2$). Clearly, $\omega_1(u) = 2+u$, and generalizing (6) of [11] to arbitrary $d \ge 2$, we have

$$\omega_d(u) = \pi_d + \pi_{d-1} \int_0^u \left(1 - (t/2)^2\right)^{(d-1)/2} dt, \qquad d \ge 2. \tag{7.5}$$

Using the preceding formulae, we have computed numerical values for the asymptotic upper bounds in Theorem 2.3, for the cases with $\rho = 1$ and $d \le 3$. These are as follows to five significant figures, where $\delta_V(\rho)$ denotes the right-hand side of (2.12) and $\delta_S(\rho)$ denotes the right-hand side of (2.13):

$$\delta_V(1) = \begin{cases} 6.4252 \times 10^3, & \text{if } d = 1, \\ 8.6212 \times 10^5, & \text{if } d = 2, \\ 1.4451 \times 10^8, & \text{if } d = 3, \end{cases}$$

$$\delta_S(1) = \begin{cases} 2.1024 \times 10^3, & \text{if } d = 1, \\ 4.6833 \times 10^4, & \text{if } d = 2, \\ 1.0578 \times 10^6, & \text{if } d = 3. \end{cases}$$

Acknowledgments. This work was partly done at meetings in the Institute for Mathematical Sciences at the National University of Singapore, whose support we gratefully acknowledge. We also thank Joseph Yukich for conversations which stimulated our initial interest in this topic. We are indebted to the referee for the idea for Lemma 3.1, and other helpful suggestions.

REFERENCES

[1] AVRAM, F. and BERTSIMAS, D. (1993). On central limit theorems in geometrical probability. *Ann. Appl. Probab.* **3** 1033–1046. MR1241033

[2] BALDI, P., RINOTT, Y. and STEIN, C. (1989). A normal approximation for the number of local maxima of a random function on a graph. In *Probability, Statistics, and Mathematics* 59–81. Academic Press, Boston, MA. MR1031278

[3] CHATTERJEE, S. (2008). A new method of normal approximation. *Ann. Probab.* **36** 1584–1610. MR2435859

[4] CHEN, L. H. Y. and SHAO, Q.-M. (2004). Normal approximation under local dependence. *Ann. Probab.* **32** 1985–2028. MR2073183

[5] CONWAY, J. H. and SLOANE, N. J. A. (1999). *Sphere Packings, Lattices and Groups*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **290**. Springer, New York. MR1662447

[6] GOLDSTEIN, L. (2005). Berry–Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *J. Appl. Probab.* **42** 661–683. MR2157512

[7] GOLDSTEIN, L. and RINOTT, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33** 1–17. MR1371949

[8] GÖTZE, F., HEINRICH, L. and HIPP, C. (1995). m-dependent random fields with analytic cumulant generating function. *Scand. J. Statist.* **22** 183–195. MR1339750

[9] HALL, P. (1988). *Introduction to the Theory of Coverage Processes*. Wiley, New York. MR973404

[10] MOLCHANOV, I. S. (1996). *Statistics of the Boolean Model for Practitioners and Mathematicians*. Wiley, Chichester.

[11] MORAN, P. A. P. (1973). The random volume of interpenetrating spheres in space. *J. Appl. Probab.* **10** 483–490. MR0353412

[12] MORAN, P. A. P. (1973). A central limit theorem for exchangeable variates with geometric applications. *J. Appl. Probab.* **10** 837–846. MR0356183

[13] PENROSE, M. (2003). *Random Geometric Graphs*. *Oxford Studies in Probability* **5**. Oxford Univ. Press, Oxford. MR1986198

[14] PENROSE, M. D. (2007). Gaussian limits for random geometric measures. *Electron. J. Probab.* **12** 989–1035. MR2336596

[15] PENROSE, M. D. and ROSOMAN, T. (2008). Error bounds in stochastic-geometrical normal approximation. In *Fifth Colloquium on Mathematics and Computer Science*. *Discrete Mathematics and Theoretical Computer Science* **AI** 71–94. Assoc. Discrete Math. Theor. Comput. Sci., Nancy. MR2508780

[16] PENROSE, M. D. and YUKICH, J. E. (2001). Central limit theorems for some graphs in computational geometry. *Ann. Appl. Probab.* **11** 1005–1041. MR1878288

[17] PENROSE, M. D. and YUKICH, J. E. (2005). Normal approximation in geometric probability. In *Stein's Method and Applications* (A. D. Barbour and L. H. Y. Chen, eds.) 37–58. World Scientific, Singapore. MR2201885

[18] STOYAN, D., KENDALL, W. S. and MECKE, J. (1987). *Stochastic Geometry and Its Applications*. Wiley, Chichester. MR895588

[19] WIDOM, B. and ROWLINSON, J. S. (1970). New model for the study of liquid-vapor phase transitions. *J. Chem. Phys.* **52** 1670–1684.

[20] ZONG, C. (1999). *Sphere Packings*. Springer, New York. MR1707318

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
LOS ANGELES, CALIFORNIA 90089-2532
USA
E-MAIL: larry@math.usc.edu

DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF BATH
BATH BA2 7AY
UNITED KINGDOM
E-MAIL: m.d.penrose@bath.ac.uk