Gaussian Limits for Generalized Spacing Statistics

Yu. Baryshnikov and J. E. Yukich

Bell Laboratories and Lehigh University

February 16, 2004

Abstract

Nearest neighbor cells in $\mathbb{R}^d$, $d \geq 1$, are used to define coefficients of divergence ($\phi$-divergences) between continuous multivariate samples. For large sample sizes, such statistical distances converge to a normal distribution with a variance depending on the underlying point density. The point measures induced by the coefficients of divergence converge weakly to a generalized Gaussian field with a covariance structure determined explicitly by the point densities. This provides the limiting distribution for a multi-variate goodness-of-fit test and further develops the limit theory for nearest neighbor statistics and $\phi$-divergences over non-uniform point sets. In $d = 1$, this extends upon and generalizes classical central limit theory for sum functions of spacings.

1 Introduction

Let $X_i$, $1 \leq i \leq n$, denote the order statistics drawn from an i.i.d. sample with distribution $F$ on $\mathbb{R}$ and let $G$ be a distribution function. Classical spacing statistics on $\mathbb{R}$ [38] take the form of an empirical $\phi$-divergence

$$\sum_{i=1}^{n} \phi(n[G(X_{i+1}) - G(X_i)])$$

\[ (1.1) \]
where $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}$ is a specified function. When $F$ and $G$ have densities $f$ and $g$, respectively, the statistics (1.1) form a discrete version of the $\phi$-divergence between $f$ and $g$, namely $\int f(x)\phi\left(\frac{g(x)}{f(x)}\right)dx$. $\phi$-divergences, introduced independently by Csiszár [13, 14], and Ali and Silvey [2, 3, 4], are a measure of the dependence between the distributions $F$ and $G$. They are widely used in non-parametric estimation and are well suited for goodness-of-fit tests [40, 18, 10, 12, 25, 37, 47].

Despite their statistical importance and despite decades of research, the limit theory for $\phi$-divergences (1.1) is however not completely understood. With the notable exception of [9] and [7] (Chapter 7.4), the limiting asymptotic distribution of spacing statistics in the multi-dimensional setting is poorly developed. In $d = 1$, the limit theory for (1.1) often assumes that either $G = F$, that $G$ is a small perturbation of $F$, or that $GF^{-1}$ has a step function derivative. Developing the limiting distribution for the empirical $\phi$-divergence (1.1) over continuous multivariate samples is important in goodness-of-fit tests; however the passage from uniform to non-uniform densities requires interchanging limits, which has proved to be an obstacle to rigorous analysis even in dimension $d = 1$.

This paper uses nearest neighbor cells to establish high dimensional analogs of the one dimensional $\phi$-divergences (1.1). The nearest neighbor cells are employed as an adaptive scheme to define statistical distances of continuous samples in $\mathbb{R}^d$. We show that the resulting functionals converge to a normal random variable whenever the distributions $F$ and $G$ have densities bounded away from zero and infinity. Using one and two point correlation functions, together with the fact that nearest neighbor cells have local interactions, we determine limiting variances explicitly in terms of $\phi, F,$ and $G$. These results, which form the heart of the paper, are facilitated by the fact that nearest neighbor cells have local interactions, permitting an efficient local coupling of binomial point sets by Poisson point sets.

The generalized $\phi$-divergences (1.1) generate canonical random point measures, denoted $\mu_{n,\phi}^g$, $n \geq 1$, which, after re-normalization, converge weakly to a generalized Gaussian field. The large scale behavior of these measures, as defined by the limiting covariance kernel, is determined by the local behavior of the underlying density of points. In particular the limiting Gaussian field is distribution free when $F = G$.

We show that in dimension $d = 1$, asymptotic normality of (1.1) holds for a broad choice of $F, G,$ and $\phi$. This generalizes and extends upon CLTs for sum functions of spacings [10, 12, 15, 16, 20, 44, 47]. In particular, asymptotic normality holds for the sum of logarithmic $k$-spacings ($\phi(x) = \log x$) whenever the densities of $F$ and $G$ are bounded away from zero and infinity, thus
resolving an open question of Darling (p. 249 of [16]).

We also develop the limit theory for nearest neighbor statistics related to (1.1); in particular we prove a CLT for the number of pairs of vertices within a fixed distance.

\textbf{Notation.} Let \( \mathcal{X} \subset \mathbb{R}^d \) be locally finite. Given a positive scalar \( a \), let \( a\mathcal{X} := \{ax : x \in \mathcal{X}\} \). Given \( y \in \mathbb{R}^d \) set \( y + \mathcal{X} := \{y + x : x \in \mathcal{X}\} \). For \( x \in \mathbb{R}^d \), let \(|x|\) be its Euclidean modulus and for \( r > 0 \), let \( B_r(x) \) denote the Euclidean ball \( \{y \in \mathbb{R}^d : |y - x| \leq r\} \). Let \( \mathbf{0} \) denote the origin of \( \mathbb{R}^d \).

Throughout let \( A \subset \mathbb{R}^d \) be compact convex subset of \( \mathbb{R}^d \) with non-empty interior. For such subsets, this implies boundary regularity in the sense that \( \lim_{n \to \infty} \frac{\partial_r(n^{1/d}A)}{n} = 0 \) for all \( r > 0 \), where \( \partial_r(n^{1/d}A) \) denotes the volume of the \( r \)-neighborhood of the boundary of \( n^{1/d}A \). Let \( B(A) \) denote the continuous functions on \( A \). If \( B \subset B(\mathbb{R}^d) \), then \( |B| \) denotes its Lebesgue measure.

Given \( f \in B(A) \) and \( \mu \) a Borel measure on \( \mathbb{R}^d \), let \( \langle \mu, f \rangle \) denote the integral of \( f \) with respect to \( \mu \). Let \( \Gamma_k \) denote a gamma random variable with parameters \( k \) and 1. Let \( P_\tau \) denote a homogeneous Poisson point process on \( \mathbb{R}^d \) with intensity \( \tau \). When \( \tau = 1 \), we write \( P \) for \( P_1 \). Let \( \log \) denote the natural logarithm. \( C \) denotes a generic positive constant whose value may change from line to line.

\section{Main Results}

\subsection{\( \phi \)-divergence based on nearest neighbor cells}

Throughout \( X_1, X_2, \ldots \) are independent random variables in \( \mathbb{R}^d \), \( d \geq 1 \), with common continuous probability density \( f : A \to \mathbb{R} \), and \( g : A \to \mathbb{R}^+ \) is continuous. We assume once and for all that \( f \) and \( g \) are bounded away from zero and infinity.

For each \( X_i, 1 \leq i \leq n \), consider the ball \( C_i := C_i(X_1, \ldots, X_n) \) centered at \( X_i \) with radius equal to the distance to the nearest neighbor in the sample \( \{X_1, \ldots, X_n\} \). We will use these cells to define high dimensional spacing statistics analogous to the classical one-dimensional statistics (1.1).

Define for \( 1 \leq i \leq n \) the sample spacings

\[ D_{i,n} := D_i(X_1, \ldots, X_n) := \int_{C_i(X_1, \ldots, X_n)} dx \]

and the transformed spacings

\[ D_{i,n}^g := D_i^g(X_1, \ldots, X_n) := \int_{C_i(X_1, \ldots, X_n)} g(x) dx. \]

We will measure the discrepancy between \( g \) and the sample density \( f \) by comparing the transformed spacings \( \{D_{i,n}^g, 1 \leq i \leq n\} \) with \( \{D_{i,n}^f, 1 \leq i \leq n\} \). Given convex \( \phi : \mathbb{R}^+ \to \mathbb{R} \) consider
the ‘nearest neighbors φ-divergence’ statistics:
\[
I_\phi(\{D_{i,n}^g\}, \{D_{i,n}^f\}) := \sum_{i=1}^n D_{i,n}^f \phi \left( \frac{D_{i,n}^g}{D_{i,n}^f} \right).
\]
(2.1)

\(I_\phi\) is a measure of the ‘distance’ between \(g\) and \(f\) and is a discrete version induced by the balls of the nearest neighbors graph of the \(\phi\)-divergence between \(g\) and \(f\), namely \(\int_A f(x) \phi \left( \frac{g(x)}{f(x)} \right) dx\).

These \(\phi\)-divergences, or ‘coefficients of divergence’, are used heavily in goodness-of-fit tests [40] and are useful in characterizing the amount of information of one distribution contained in another [40, 41]. If \(P\) and \(Q\) are measures with densities \(f\) and \(g\), respectively, then the \(\phi\)-divergence of \(f\) and \(g\) is a measure of the statistical distinguishability of \(P\) and \(Q\). Several choices of \(\phi\) figure prominently in estimation and decision theory. For example, \(\phi_0(x) := \log x\) defines Kullback-Leibler information (also called the modified log-likelihood ratio statistic) and is used in maximum spacing methods. \(\phi_{1/2}(x) := 2(1 - \sqrt{x})^2\) yields the square of the Hellinger distance. \(\phi_1(x) := x \log x\) yields the log-likelihood ratio statistic or I-Divergence of Kullback-Leibler [14]. Finally, \(\phi_2(x) := (x - 1)^2/2\) yields the chi-squared divergence. \(\phi_{1/2}\) and \(\phi_2\) are members of the family \(\phi_\beta(x) = \frac{x^{\beta - 1} - 1}{\beta - 1}\), \(\beta \neq 0, 1\), known as the ‘power divergence’ family of \(\phi\)-divergences [40]. The function \(\phi^{(r)}(x) := x^r\) yields information gain of order \(r\).

If \(f\) is unknown, we can replace \(D_{i,n}^f\) in (2.1) by its empirical estimate \(\hat{D}_{i,n}^f := \frac{1}{n}\), thus extending the statistic (1.1) to all dimensions:
\[
N_{\phi,n}(X_1, \ldots, X_n) := N_{\phi,n,\phi}(X_1, \ldots, X_n) := \sum_{i=1}^n \phi(n \cdot D_{i,n}^g).
\]
(2.2)

We will henceforth call \(N_{\phi,n,\phi}\) the ‘nearest neighbors spacing statistic’, or ‘empirical nearest neighbor \(\phi\)-divergence’. We will usually write \(N_{\phi,n}\) for \(N_{\phi,n,\phi}\) when the meaning is clear. We write \(N_{\phi,n}(X_1, \ldots, X_n)\) when \(g\) is identically equal to the constant \(\beta\) and when \(\beta = 1\), we write \(N_\phi\) instead of \(N_\phi^1\). To show asymptotic normality of the statistics \(N_{\phi,n}(X_1, \ldots, X_n)\), we will need to assume that \(\phi \in \Phi\), where \(\Phi := \{\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+ : \mathbb{E}[\phi^p(\alpha \Gamma_1)] < \infty\text{ for all }\alpha > 0\text{ and some }p > 3\}\). \(\Phi\) includes \(\phi_0, \phi_{1/2}, \phi_1\), and \(\phi_2\), as well as the power functions \(\phi^{(r)}(x)\), \(r > 0\).

Consider the point measures induced by the empirical nearest neighbor \(\phi\)-divergence:
\[
\mu_{n,\phi} := \sum_{i=1}^n \phi(n \cdot D_{i,n}^g) \delta_{X_i},
\]
(2.3)
and let \(\overline{\mu}_{n,\phi} := \mu_{n,\phi} - \mathbb{E}\mu_{n,\phi}\) be their centered versions. We say that random signed measures \(\mu_n\) converge in law to a generalized Gaussian field if the finite-dimensional distributions of the random field \((f, \mu_n), f \in B(A)\) converge in law as \(n \rightarrow \infty\) to those of a mean zero finitely additive Gaussian field.
The goals of this paper are threefold: (i) establish convergence in law of the point measures $n^{-1/2}P_{n,\phi}$ to a Gaussian limit under mild conditions on $f$, $g$, and $\phi$, (ii) establish convergence in distribution of generalized spacings functionals, including those based on nearest neighbor distances and cells of fixed size, and (iii) generalize classical CLT theory for sum functions of spacings in $d = 1$.

Before stating our main results we need two auxiliary propositions. The first is a special case of the upcoming Theorem 2.1 and asserts the existence of a limiting variance for the statistic $N_\phi$ over Poisson point samples.

**Proposition 2.1** (existence of limiting variance) For all $\phi \in \Phi$ and all $\beta > 0$, there exists a positive constant $V_\phi(\beta)$ such that
\[
\lim_{\lambda \to \infty} \frac{\text{Var}[N_\phi^\beta(\mathcal{P} \cap [0,1]^d)]}{\lambda} = V_\phi(\beta). \tag{2.4}
\]

Inserting an extra point into the Poisson point set $\mathcal{P}$ induces only local changes in the nearest neighbors graph [8] and therefore the following proposition holds (the proof is a simple consequence of Lemma 6.1 and Definition 2.1 of [34]).

**Proposition 2.2** (existence of ‘average add-one cost’) For all $\phi \in \Phi$ and all $\beta > 0$, there exists a constant $\Delta_\phi(\beta)$ such that
\[
\mathbb{E} \left[ \lim_{R \to \infty} [N_\phi^\beta(\mathcal{P} \cap B_R(0) \cup 0) - N_\phi^\beta(\mathcal{P} \cap B_R(0))] \right] = \Delta_\phi(\beta).
\]

In most of our examples $\Delta_\phi(\beta)$ will be strictly positive because adding a point tends to increase the value of $N_\phi^\beta$.

### 2.2 A general CLT for $\phi$-divergence statistics

The volume of the nearest neighbors cell around the origin with respect to $\mathcal{P}$ is a $\Gamma_1$ random variable. Given $X_i = x$, $n \cdot D_{i,n}^\phi$ is roughly equal to $g(x) \cdot n \cdot |C_i|$, which is approximated by $g(x)/f(x) \Gamma_1$, since for large $n$, the points locally around $x$ are nearly Poisson with parameter $f(x)$. Thus the summands in (2.2) behave asymptotically like $\phi(\frac{g(x)}{f(x)} \Gamma_1)$. Conditional on $X_i = x$ and neglecting cell dependence, one expects in view of Proposition 2.1 and a simple scaling argument that the limiting variance of $\phi(n \cdot D_{i,n}^\phi)$ should equal $V_\phi(\frac{g(x)}{f(x)})$. Furthermore, in the absence of conditioning, one expects that the limiting variance of $n^{-1/2}P_{n,\phi}$ would behave like a weighted average of $V_\phi(\frac{g(x)}{f(x)})$. This is indeed the case and is the content of Theorem 2.1 below (cf. Remark (iii) as well).
The following general CLT, our main result, establishes convergence in law of the re-normalized point measures $\mu_{n,\phi}^g$, $n \geq 1$, defined in (2.3), to a generalized Gaussian field whose covariance kernel is a weighted average of the variance and add-one cost functions. Related results, corollaries, and examples follow in subsequent sections.

**Theorem 2.1** Let $\phi \in \Phi$ and $h \in B(A)$. Then as $n \to \infty$

$$\frac{\text{Var}[\langle h, \mu_{n,\phi}^g \rangle]}{n} \to \int_A h(x)V_{\phi}(\frac{g(x)}{f(x)})f(x)dx - \left( \int_A h(x)\Delta_{\phi}(\frac{g(x)}{f(x)})f(x)dx \right)^2$$

(2.5)

and $n^{-1/2}\mu_{n,\phi}^g$ converges in law to a Gaussian field with covariance kernel

$$\int_A h_1(x)h_2(x)V_{\phi}(\frac{g(x)}{f(x)})f(x)dx - \int_A h_1(x)\Delta_{\phi}(\frac{g(x)}{f(x)})f(x)dx \int_A h_2(x)\Delta_{\phi}(\frac{g(x)}{f(x)})f(x)dx.$$  (2.6)

For many choices of $\phi$ we may identify $V_{\phi}(\beta)$ and $\Delta_{\phi}(\beta)$ exactly, thus providing explicit limiting distributions of several statistical distances of interest, including information gain and log-likelihood (section 2.4) and sum functions of spacings in $d = 1$ (sections 2.5, 2.6), thus extending classical results involving the ordinary spacings (1.1). Theorem 2.1 extends to statistics based on sums of lengths of nearest neighbors edges (section 2.3) and it also extends to spacings defined by cells of fixed radius (section 2.6). Notice that if we put

$$\sigma_{f,g}^2 := \int_A V_{\phi}(\frac{g(x)}{f(x)})f(x)dx - \left( \int_A \Delta_{\phi}(\frac{g(x)}{f(x)})f(x)dx \right)^2,$$

then Theorem 2.1 yields the following CLT for the statistic $N_{\phi}^g$:

$$\frac{N_{\phi}^g(X_1,\ldots,X_n) - \mathbb{E}N_{\phi}^g(X_1,\ldots,X_n)}{n^{1/2}} \overset{D}{\to} N(0,\sigma_{f,g}^2).$$

**Remarks.**

(i) *(Related work)* Bickel and Breiman [9], and subsequently Schilling [42], consider the statistics $N_{\phi}^g(X_1,\ldots,X_n)$ for the function $\phi(x) := \exp(-x)$. Using the approximation $D_{i,n}^g \approx g(X_i)[C_i(X_1,\ldots,X_n)],$ they establish a CLT for the empirical process of nearest neighbor distances, but do not consider the weak convergence of the associated random measures $\mu_{n,\phi}^g$. Strong limit theorems are available for multi-variate spacings using quite general ‘shapes’ as shown by Deheuvels et al. [17]. Penrose [32] finds a CLT for $k$-nearest neighbor distances and a strong law [31] for the largest nearest-neighbor link. Henze [23] finds the limit distribution for the maxima of weighted nearest neighbor distances.

(ii) *(Independence of limit over disjoint sets)* Theorem 2.1 says that for $h_1,\ldots,h_m$ in $B(A)$ the $m$-vector $(\langle h_1, n^{-1/2}\mu_{n,\phi}^g \rangle, \ldots, \langle h_m, n^{-1/2}\mu_{n,\phi}^g \rangle)$ tends to a Gaussian field with covariance kernel
Knowing the distribution of \( \langle h_1, n^{-1/2} \mathcal{P}^{g}_{n, \phi} \rangle, \ldots, \langle h_m, n^{-1/2} \mathcal{P}^{g}_{n, \phi} \rangle \) over functions with disjoint supports could be useful in situations where the density changes from location to location, as in change point problems.

(iii) (Limits are distribution free) If \( f \) and \( g \) are bounded away from zero and infinity on \( A \), then [28] the empirical nearest neighbors estimator is a.s. consistent whenever there is some \( \gamma > 0 \) such that for all \( \alpha > 0 \),

\[
\mathbb{E} \left[ \phi^4 + \gamma (\alpha \Gamma_1) \right] < \infty:
\]

\[
\lim_{n \to \infty} \frac{N^g_{\phi}(X_1, \ldots, X_n)}{n} = \int_A f(x) \mathbb{E} \left[ \phi \left( \frac{g(x)}{f(x)} \Gamma_1 \right) \right] dx \quad \text{a.s.}
\]

Thus, combining the above with Theorem 2.1, when \( g = f \) the limiting mean, variance, and distribution for \( N^g_{\phi}(X_1, \ldots, X_n) \) and \( \mu^g_{\phi} \) do not depend on \( f \). Therefore the nearest neighbor statistics are asymptotically distribution free under the null hypothesis \( g = f \) and have asymptotic variance \( V_{\phi}(1) - (\Delta_{\phi}(1))^2 \). Therefore a possible goodness-of-fit test would be to take the density \( g \) to be tested, compute the statistic \( N^g_{\phi}(X_1, \ldots, X_n) \), and see whether the cumulative distribution function is close to the cumulative distribution function of a mean normal random variable with variance \( V_{\phi}(1) - (\Delta_{\phi}(1))^2 \).

(iv) (Poisson CLT, rates of convergence) For all \( \lambda > 0 \) let \( \mathcal{P}^f_{\lambda} \) be a Poisson point process on \( A \) with intensity measure \( \lambda f : A \to \mathbb{R}^+ \). The Poissonized version of the \( \phi \)-divergence (2.2) is

\[
N^g_{\phi}(\mathcal{P}^f_{\lambda}) = \sum_{x \in \mathcal{P}^f_{\lambda}} \phi \left( \lambda \cdot \int_{C(x, \mathcal{P}^f_{\lambda})} g(u) du \right)
\]

(2.7)

and the Poisson analog of the point measures (2.3) is

\[
\mu^g_{\phi, \lambda} = \sum_{x \in \mathcal{P}^f_{\lambda}} \phi \left( \lambda \cdot \int_{C(x, \mathcal{P}^f_{\lambda})} g(u) du \right) \delta_x.
\]

(2.8)

Our methods yield the following Poisson version of Theorem 2.1: If \( \phi \in \Phi \), and \( h \in B(A) \), then as \( \lambda \to \infty \)

\[
\frac{\text{Var}[\langle h, \mu^g_{\phi, \lambda} \rangle]}{\lambda} \to \int_A h(x) V_{\phi} \left( \frac{g(x)}{f(x)} \right) f(x) dx
\]

(2.9)

and \( \lambda^{-1/2} \mathcal{P}^g_{\lambda, \phi} \) converges in law to a Gaussian field with covariance kernel

\[
\int_A h_1(x) h_2(x) V_{\phi} \left( \frac{g(x)}{f(x)} \right) f(x) dx.
\]

Moreover, section 3.3 below tells us that the rate of convergence to a normal is \((\log \lambda)^{3d}/\lambda^{1/2}\).

(v) (Voronoi cells) Volumes of nearest neighbor cells are computationally attractive and have exponentially decaying correlations. Defining analogous statistics based on cells generated by any locally defined Euclidean graph (e.g. Voronoi cells) leads to identical CLTs. The only change involves the identity of the functions \( V_{\phi} \) and \( \Delta_{\phi} \).
(vi) (Properties of limiting variance) In most of our examples, $\Delta \phi$ is strictly positive, showing that Poissonization contributes extra randomness which manifests itself in a larger limiting variance. When $V_\phi$ and $\Delta \phi$ are convex, which is the case when $\phi(x) = x^r$, $r \geq 1$, or when $\phi(x) = x \log x$ (see section 2.5.2), then $\int_A V_\phi(\frac{g(x)}{f(x)})f(x)dx$ and $\int_A \Delta \phi(\frac{g(x)}{f(x)})f(x)dx$ are themselves divergences. Thus the limiting variance (2.9) serves as a natural measure of distance. Basic properties of $\phi$-divergences [4, 13] imply that the limiting variance over Poisson samples is minimized when $g = f$ and replacing $g$ and $f$ with transformed densities can only decrease the limiting variance.

(vii) (Scaling) It is a straightforward consequence of the definitions that the existence for all $\beta > 0$ of the limit
\[
\lim_{\lambda \to \infty} \frac{\text{Var}[N_\phi^\beta(\mathcal{P}_\lambda \cap [0,1]^d)]}{\lambda} = V_\phi(\beta) \tag{2.10}
\]
is equivalent to the existence for all $c_1 > 0$, $c_2 > 0$ of the limit
\[
\lim_{\lambda \to \infty} \frac{\text{Var}[N_\phi^{c_1}(\mathcal{P}_\lambda^{c_2} \cap [0,1]^d)]}{\lambda} = c_2 V_\phi(\frac{c_1}{c_2}). \tag{2.11}
\]
Defining $\overline{V}(\beta) := V(\frac{1}{\beta})$, the last statement is equivalent to the existence for all $\beta > 0$ of the limit
\[
\lim_{\lambda \to \infty} \frac{\text{Var}[N_\phi^\beta(\mathcal{P}_\lambda \cap [0,1]^d)]}{\lambda} = \beta \overline{V}_\phi(\beta) \tag{2.12}
\]

2.3 Statistics based on $\phi$-weighted edges in the nearest neighbors graph

Section 2.2 establishes the asymptotic normality of spacings involving nearest neighbor cells. Alternatively, we may define high dimensional spacing statistics based on distances to the $k$th nearest neighbors. This goes as follows. Given a point set $\mathcal{X} \subset \mathbb{R}^d$ and a fixed convex set $\mathcal{C} \subset \mathbb{R}^d$ containing the origin of $\mathbb{R}^d$, define for all $x \in \mathbb{R}^d$ and all $k = 1, 2, \ldots$
\[
d_k^\mathcal{C}(x, \mathcal{X}) := \inf \{ \text{card}(x + t\mathcal{C}) \cap \mathcal{X} \geq k + 1 \}.
\]
Thus $d_k^\mathcal{C}(x, \mathcal{X})$ is the set-induced (asymmetric) distance between $x$ and its $k$th nearest neighbor in $\mathcal{X}$. For every $x \in \mathcal{X}$, consider the cell $C_k^\mathcal{C}(x, \mathcal{X}) := x + d_k^\mathcal{C}(x, \mathcal{X}) \cdot \mathcal{C}$. Given an i.i.d. sample $X_1, X_2, \ldots, X_n$, let $d_{\mathcal{C}, i, n}^k := d_k^\mathcal{C}(X_i, \{X_i\}_{i=1}^n)$.

It is easy to check that Theorem 2.1 and its proof hold if the scaled volumes $nD_{i, n}$ are replaced by the scaled distances $n^{1/d}d_{\mathcal{C}, i, n}^k$. More precisely, given $\phi : \mathbb{R}^+ \to \mathbb{R}$, define the $\phi$-weighted $k$-nearest neighbors edge length statistic
\[
L_\phi(X_1, \ldots, X_n) := \sum_{i=1}^n \phi(n^{1/d}d_{\mathcal{C}, i, n}^k)
\]
as well as the point measures induced by $L_\phi$

$$\mu_{n,\phi}^L := \sum_{i=1}^{n} \phi(n^{1/d}d_{c,i,n}^k)\delta_{X_i}. $$

The analog of Proposition 2.2 holds for the functional $L_\phi$: for all $\phi \in \Phi$ and $\beta > 0$ there is a constant $\Delta_\phi^L(\beta)$ such that

$$\Delta_\phi^L(\beta) = \mathbb{E} \left[ \lim_{R \to \infty} [L_\phi(P_\beta \cap B_R(0)) - L_\phi(P_\beta \cap B_R(0))] \right].$$

Moreover, for all $\beta > 0$, there is a positive constant $V_\phi^L(\beta)$ such that

$$\lim_{\lambda \to \infty} \frac{\operatorname{Var}[L_\phi(P_\beta^\lambda \cap [0,1]^d)]}{\lambda} = \beta V_\phi^L(\beta).$$

It is straightforward to verify that Theorem 2.1 holds if $N_\phi$ is replaced by $L_\phi$ and the measures $\mu_{n,\phi}^L$ are replaced by $\mu_{n,\phi}^L$.

In particular, if $\phi(x) = x^r$, $r > 0$, then $L_\phi$ are homogeneous of degree $r$, that is for all $a > 0$,

$$L_\phi(a(X_1, ..., X_n)) = a^r L_\phi(X_1, ..., X_n).$$

Thus $V_\phi^L(\beta) = \beta^{-2r/d} V_\phi^L(1)$ and $\Delta_\phi^L(\beta) = \beta^{-2r/d} \Delta_\phi^L(1)$, implying the following CLT for the sum of power-weighted edge lengths in the nearest neighbor graph:

**Theorem 2.2** Let $\phi(x) = x^r$, $r > 0$, and $h \in B(A)$. Then as $n \to \infty$

$$\frac{\operatorname{Var}[[h, \mu_{n,\phi}^L]]}{n} \to V_\phi^L(1) \int_A h^2(x)f(x)^{(d-2r)/d}dx - \left( \Delta_\phi^L(1) \int_A h(x)f(x)^{(d-r)/d}dx \right)^2.$$

and $n^{-1/2} \mu_{n,\phi}^L$ converges in law to a Gaussian field with covariance kernel

$$V_\phi^L(1) \int_A h_1(x)h_2(x)f(x)^{(d-2r)/d}dx - (\Delta_\phi^L(1))^2 \int_A h_1(x)f(x)^{(d-r)/d}dx \int_A h_2(x)f(x)^{(d-r)/d}dx.$$

**Remarks.** When $k = 1$ and when $C$ is the unit ball in $\mathbb{R}^d$, then $L_\phi(X_1, ..., X_n)$ is the sum of the edge lengths in the $\phi$-weighted nearest neighbors graph on $X_1, ..., X_n$. In this setting, [35] establishes laws of large numbers for $L_\phi$, and, in the case of uniform $X_i$ and $\phi(x) = x$, [5] establishes a CLT for the functional $L_\phi$. Theorem 2.2 thus generalizes [5] to arbitrary $\phi$ and non-uniform $X_1, ..., X_n$. Byers and Raftery [11] consider functionals related to $L_\phi$, but do not attempt to develop the limit theory.

### 2.4 Information gain and log-likelihood in high dimensions

For differentiable $\phi$ with $\phi(0) = 0$, the limiting variance $V_\phi(\beta)$ in Proposition 2.1 can be easily computed. Therefore the limiting variance in Theorem 2.1 can be described in terms of integrals.
depending explicitly on $f$ and $g$. We will use these representations to develop the limit theory for the information gain and log-likelihood statistic in high dimensions. In dimension $d = 1$ this extends classical results (section 2.6.2, 2.6.3).

For all $s, t, u \in \mathbb{R}^+$, let $I(s, t, u)$ be the volume of the intersection of two balls in $\mathbb{R}^d$, with respective volumes $s$ and $t$, at a distance $u$ apart. Set

$$J_d(s, t) := \int_{\max(s, t)}^\infty \left[ e^{I(s, t; (u/\omega_d)^{1/d})} - 1 \right] du,$$

where $\omega_d$ is the volume of the unit radius ball in $\mathbb{R}^d$. The following is proved in section four.

**Proposition 2.3** (evaluation of the limiting variance $V_\phi(\beta)$) Assume $\phi \in \Phi$ is differentiable and $\phi(0) = 0$. Then for all $\beta > 0$, $V_\phi(\beta) := \lim_{\lambda \to \infty} \frac{\text{Var}[N_\phi^\beta(P_\lambda \cap [0, 1]^d)]}{\lambda} = E[\phi^2(\beta \Gamma_1)] + \beta^2 \int_0^\infty \int_0^\infty \phi'(\beta s)\phi'(\beta t)e^{-(s+t)} [J_d(s, t) - \max(s, t)] ds dt$

provided that the integral exists.

We do not have an explicit representation for $\Delta_\phi(\beta)$ for general $\phi$, but note that its definition easily lends itself to Monte-Carlo approximation in the following way. The increment of the functional $N_\phi$ when the sample size goes from $n$ to $2n$ should be close to $n \Delta_\phi$ with an error of the order $\varepsilon n$, where $\varepsilon n = O(n^{-1/d})$. This, combined with the CLT for $N_\phi$, implies that the increment, when divided by $n$, is close to $\Delta_\phi$, with high probability.

### 2.4.1 Information gain of order $r$

When $\phi(x) = x^r$, $r > 0$, the statistic $N_\phi^\beta$ yields Rényi’s information gain (I-divergence) of order $r$. We write $N_\phi^\beta$ to denote $N_\phi^\beta$. Since $\phi(x) = x^r$ satisfies the conditions of Proposition 2.3, the following is immediate. For all $r > 0$, define the constant

$$C_r := r^2 \int_0^\infty \int_0^\infty s^{r-1}t^{r-1}e^{-(s+t)} [J_d(s, t) - \max(s, t)] ds dt.$$

**Corollary 2.1** For all $\beta > 0$,

$$V_\phi(\beta) = \lim_{\lambda \to \infty} \frac{\text{Var}[N_\phi^\beta(P_\lambda \cap [0, 1]^d)]}{\lambda} = \beta^{2r} \left[ E[(\Gamma_1)^{2r}] + C_r \right].$$

Using Theorem 2.1 we easily deduce asymptotic normality of the information gain statistic as well as the weak convergence of the associated point measures. The special case $d = 1$ is treated in section 2.6.2. Let

$$\sigma_\phi^2(f, g) := \left[ E[(\Gamma_1)^{2r}] + C_r \right] \int_A \left( \frac{f(x)}{g(x)} \right)^{2r} f(x) dx.$$
Theorem 2.3 Let \( \phi(x) = x^r, \ r > 0 \). Then as \( \lambda \to \infty \)
\[
\frac{\text{Var}[N_{\phi}^3(P_{\lambda}^f)]}{\lambda} \to \sigma^2_{\phi}(f, g)
\]
and
\[
\frac{N_{\phi}^3(P_{\lambda}^f) - \mathbb{E} N_{\phi}^3(P_{\lambda}^f)}{\lambda^{1/2}} \xrightarrow{D} N(0, \sigma^2_{\phi}(f, g))
\]
while \( \lambda^{-1/2} \pi_{\lambda, \phi}^g \) converges in law to a Gaussian field with covariance kernel
\[
[\mathbb{E} ((\Gamma_1)^{2r}) + C_r] \int_A h_1(x) h_2(x) \left( \frac{f(x)}{g(x)} \right)^r f(x) dx.
\]

2.4.2 Log likelihood

When \( \phi(x) = x \log x \), the statistic \( N_{\phi}^g \) yields the log-likelihood statistic. To apply Theorem 2.1, we define the constants
\[
I_1 := \int_0^\infty \int_0^\infty (\log s + 1)(\log t + 1)e^{-(s+t)}[J_d(s, t) - \max(s, t)] ds dt,
\]
\[
I_2 := \int_0^\infty \int_0^\infty (\log s + 1)e^{-(s+t)}[J_d(s, t) - \max(s, t)] ds dt,
\]
and
\[
I_3 := \int_0^\infty \int_0^\infty e^{-(s+t)}[J_d(s, t) - \max(s, t)] ds dt.
\]

The following is an easy consequence of Proposition 2.3.

Corollary 2.2 For all \( \beta > 0 \),
\[
V_{\phi}(\beta) = \lim_{\lambda \to \infty} \frac{\text{Var}[N_{\phi}^\beta(P_{\lambda} \cap [0, 1]^d)]]}{\lambda} = 
\beta^2 \left[ 2 + \pi/3 - 6\gamma + 2\gamma^2 + I_1 \right] + \beta^2 \log \beta [6 - 4\gamma + 2I_2] + (\beta \log \beta)^2 [2 + I_3].
\]

Using Theorem 2.1 we may deduce asymptotic normality of the log-likelihood statistic as well as convergence in law of the associated point measures. The special case \( d = 1 \) is treated in section 2.6.3. Let
\[
\sigma^2_{\phi}(f, g) := \int_A \left( \frac{g(x)}{f(x)} \right)^2[C_1 + C_2 \log \left( \frac{g(x)}{f(x)} \right) + C_3 (\log \left( \frac{g(x)}{f(x)} \right)^2] f(x) dx
\]
where \( C := 2 + \frac{\pi}{3} - 6\gamma + 2\gamma^2 + I_1, C_2 := 6 - 4\gamma + 2I_2, \) and \( C_3 := 2 + I_3. \)

Theorem 2.4 Let \( \phi(x) = x \log x \). Then as \( \lambda \to \infty \)
\[
\frac{\text{Var}[N_{\phi}^3(P_{\lambda}^f)]}{\lambda} \to \sigma^2_{\phi}(f, g)
\]
and
\[ \frac{N^2_{\phi}(P_{n,k}^{1}) - \mathbb{E} N^2_{\phi}(P_{n,k}^{1})}{\lambda^{1/2}} \overset{D}{\rightarrow} \mathcal{N}(0, \sigma^2_{\phi}(f, g)) \]
while \( \lambda^{-1/2} \pi_{\phi}^{n} \) converges in law to a Gaussian field with covariance kernel
\[ \int_{A} h_1(x) h_2(x) \left( \frac{g(x)}{f(x)} \right)^2 \left[ C_1 + C_2 \log \left( \frac{g(x)}{f(x)} \right) + C_3 \left( \log \frac{g(x)}{f(x)} \right)^2 \right] f(x) dx. \]

2.5 Asymptotic normality of sum functions of spacings

Recall that \( C^k(x, \mathcal{X}) := x + d_i^k(x, \mathcal{X}) \cdot C \). It is easy to check that the upcoming proof of Theorem 2.2 also holds if the cells \( C_i \) defined there are replaced by cells
\[ C^k_i := C^k(X_i, \{X_1, \ldots, X_n\}), \quad 1 \leq i \leq n, \]
and the transformed spacings \( D_{i,n}^k \) are replaced by \( D_{i,n}^0 := \int_{C^k_i} g(x) dx, \quad 1 \leq i \leq n \).

In particular, if \( d = 1 \) and \( \mathcal{C} = [0, 1] \), then \( C^k_i \) is the difference of the order statistics \( X_{(i+k)} - X_{(i)} \), that is \( C^k_i \) corresponds to \( k \)-spacings on the line (cf. [7]). Notice that
\[ \phi \left( \int_{C^k_i} g(x) dx \right) = \phi(G(X_{(i+k)}) - G(X_{(i)})), \]
where \( G \) is the distribution function for the density \( g \). As in (2.2), this leads to consideration of the empirical \( \phi \)-divergence for overlapping \( k \)-spacings, that is
\[ S_{\phi,k}^2(X_1, \ldots, X_n) = \sum_{i=1}^{n-k} \phi(n \cdot [G(X_{(i+k)}) - G(X_{(i)})]), \]
also known as a \textit{sum function of spacings}.

Let \( \phi \in \Phi_k \), where \( \Phi_k := \{ \phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+: \mathbb{E} [\phi^p(\alpha \Gamma_k)] < \infty \text{ for all } \alpha > 0 \text{ and some } p > 3 \} \). When \( \phi \in \Phi_k \), \( S_{\phi,k}^2 \) is a consistent estimator of the \( \phi \)-divergence [28, 35]:
\[ \lim_{n \rightarrow \infty} \frac{S_{\phi,k}^2(X_1, \ldots, X_n)}{n} = \int_A f(x) \mathbb{E} \left[ \phi \left( \frac{g(x)}{f(x)} \Gamma_k \right) \right] dx \quad a.s. \]

Analogous consistency results hold for non-overlapping \( k \)-spacings [43].

The following corollary of Theorem 2.1 is proved in section five and establishes convergence in law of the measures \( \pi_{n,\phi,k}^0 := \mu_{n,\phi,k}^0 - \mathbb{E} \mu_{n,\phi,k}^0 \) where
\[ \mu_{n,\phi,k}^0 := \sum_{i=1}^{n-k} \phi(n \cdot [G(X_{(i+k)}) - G(X_{(i)})]) \delta X_i. \]

Throughout, let \( \Gamma_k = \sum_{i=1}^k Y_i \), where \( Y_i, \quad i \geq 1 \), are i.i.d. \( \Gamma_1 \) random variables. Put for all \( \beta > 0 \),
\[ V_{\phi,k}^2(\beta) := 2 \sum_{l=1}^{k-1} \text{Cov}[\phi(\beta \Gamma_l), \phi(\beta(\Gamma_{k+l} - \Gamma_i))] + 2k[\mathbb{E} \phi(\beta \Gamma_k) - \mathbb{E} \phi(\beta \Gamma_{k+1})] \mathbb{E} \phi(\beta \Gamma_k) + \mathbb{E} \phi^2(\beta \Gamma_k) \]
(2.13)
and
\[ \Delta_{\phi,k}^S(\beta) := (k+1)E[\phi(\beta \Gamma_k)] - kE[\phi(\beta \Gamma_{k+1})]. \] (2.14)

The following analog of Proposition 2.1 holds: for all \( k, \phi \in \Phi_k \), and \( \beta > 0 \), there is a positive constant \( V_{\phi,k}(\beta) \) such that
\[ \lim_{\lambda \to \infty} \frac{\text{Var}[S_{\phi,k}(P_{\lambda} \cap [0,1])]}{\lambda} = V_{\phi,k}(\beta). \]

**Theorem 2.5** *(Gaussian limit for sum functions of spacings)* Let \( \phi \in \Phi_k \) and \( h \in B(A) \). Then as \( n \to \infty \) and for all \( \beta > 0 \)
\[
\frac{\text{Var}[\langle h, \mu_{g,\phi,k} \rangle]}{n} \to \int_A h(x)V_{\phi,k}^S(g(x))f(x)dx - \left( \int_A \Delta_{\phi,k}^S(g(x))f(x)dx \right)^2
\]
while \( n^{-1/2}p_{n,\phi,k}^g \) converges in law to a Gaussian field with covariance kernel
\[
\int_A h_1(x)h_2(x)V_{\phi,k}^S(g(x))f(x)dx - \int_A h_1(x)\Delta_{\phi,k}^S(g(x))f(x)dx\int_A h_2(x)\Delta_{\phi,k}^S(g(x))f(x)dx.
\]

**Remarks.** (i) Darling [16] undertook the first systematic study of the statistics \( S_{\phi,k} \) when \( k = 1 \), but restricted attention to uniform samples. Theorem 2.3 generalizes Holst [24], as well as earlier work of Cressie [12], who proves asymptotic normality (but not weak convergence) for sum functions of \( k \)-spacings over uniform points. Holst uses a generalization of LeCam’s method and a CLT for \( k \)-dependent random variables. In \( d = 1 \), Holst and Rao [25] prove asymptotic normality of \( S_{\phi,1}^g \) under their admittedly ‘somewhat stringent conditions’ on \( f \) and \( g \). For non-uniform samples the asymptotics of \( S_{\phi,1}^g \) have been widely studied under the assumption that \( G \) runs through a sequence of alternatives \( G_n \) approaching the uniform distribution; see Hall [22] and del Pino [37]. Theorem 2.3 does not impose this restriction. Khashimov [29] establishes asymptotic normality of \( S_{\phi,k}^1 \) under rather technical differentiability conditions on \( \phi \) and \( f \).

(ii) Without loss of generality we could of course assume that \( f \) is the uniform density on \( A \). However, this would not simplify our proofs and therefore we adhere to the present general terminology.

### 2.6 Corollaries of Theorem 2.5

For many tests involving goodness-of-fit (Dudewicz et al. [18], Blumenthal [10], Cressie [12], Holst and Rao [25], del Pino [37], Weiss [47] ) and parametric estimation (Ghosh and Jammalamadaka [21]) it is important to know the asymptotic distribution of the statistics \( S_{\phi,k}(X_1, ..., X_n) \) for
arbitrary $g$ and $f$. It is straightforward to verify that if we take $\phi(x) = 1_{[0,t]}(x)$ and use the relations
\[ \mathbb{E}\phi(\Gamma_1/\beta) = 1 - e^{-\beta t}, \quad \mathbb{E}\phi(\Gamma_2/\beta) = 1 - (1 + \beta t)e^{-\beta t}, \]
in (2.13) and (2.14), then Theorem 2.5 yields the classical CLT for the empirical distribution function for the $k$-spacings process [38, 7]. We now determine the limit theory for other choices of $\phi$.

2.6.1 Limit theory for logarithms of spacings

Setting $\phi(x) = \log x$ in Theorem 2.5 yields a CLT for logarithms of $k$-spacings as follows. Let $\psi$ be the di-gamma function with $\psi(k) := \sum_{j=1}^{k-1} j^{-1} - \gamma$, where $\gamma$ is Euler’s constant, and let $\psi'(k) := -\sum_{j=1}^{k-1} j^{-2} + \pi^2/6$.

By Cressie [12] and Holst [24],
\[ \sum_{l=1}^{k-1} \text{Cov}(\log(\Gamma_k), \log(\Gamma_{k+1} - \Gamma_l)) = k(k-1)\psi'(k) - (k-1). \]

Since $\mathbb{E}[\log(\Gamma_k)] = \psi(k)$, we also have
\[ 2k(\mathbb{E}\log(\Gamma_k) - \mathbb{E}\log(\Gamma_{k+1}))\mathbb{E}\log(\Gamma_k) = -2\psi(k). \]

Also, $\mathbb{E}[\log^2(\Gamma_k)] = \psi'(k) + (\psi(k))^2$. So, combining terms and using (2.13) gives
\[ V_{S_{\log,k}}(1) = 2k(k-1)\psi'(k) - (k-1) - 2\psi(k) + \psi'(k) + (\psi(k))^2. \]

Using (2.14) gives $\Delta_{S_{\log,k}}(1) = (k+1)\psi(k) - k\psi(k+1) = \psi(k) - 1$. Letting $U_1, U_2, \ldots$ be i.i.d. with the uniform density on $[0,1]$ and putting $\tau_k := (2k^2 - 2k + 1)\psi'(k) - 2k + 1$ we have:

**Corollary 2.3 (Gaussian limits for logarithmic k-spacings, uniform densities)** As $n \to \infty$,
\[ \frac{\text{Var}[S_{\log,k}(U_1, \ldots, U_n)]}{n} \to \tau_k, \]

and
\[ \frac{S_{\log,k}(U_1, \ldots, U_n) - \mathbb{E}S_{\log,k}(U_1, \ldots, U_n)}{n^{1/2}} \xrightarrow{D} N(0, \tau_k) \]

while $n^{-1/2}T_n_{\log,k}$ converges in law to a Gaussian field with covariance kernel
\[ (2[k(k-1)\psi'(k) - (k-1)] - 2\psi(k) + \psi'(k) + (\psi(k))^2) \int_0^1 h_1(x)h_2(x)dx - (\psi(k)-1)^2 \int_0^1 h_1(x)dx \int_0^1 h_2(x)dx. \]
By using simple relations such as
\[
\text{Cov}(\log \beta \Gamma_k, \log \beta (\Gamma_{k+1} - \Gamma_1)) = \text{Cov}(\log \Gamma_k, \log (\Gamma_{k+1} - \Gamma_1)), \quad \beta > 0,
\]
it follows that \( V_{\text{log},k}^S(\beta) = V_{\text{log},k}^S(1) + \log^2 \beta + 2 \log \beta (k(1 - \psi(k) - 1) \) and \( \Delta_{\text{log},k}^S(\beta) = \Delta_{\text{log},k}^S(1) + \log \beta. \)

We may thus generalize Corollary 2.3 to logarithmic \( k \)-spacings over non-uniform points.

**Corollary 2.4** (CLT limits for logarithmic \( k \)-spacings, general densities) Let \( X_1, X_2, \ldots \) be i.i.d. with density \( f \) on \([0,1]\). As \( n \to \infty \)
\[
\frac{\text{Var}[S_{\text{log},k}^g(X_1,\ldots,X_n)]}{n} \to \tau_k + \text{Var} \left[ \frac{\log(f(X))}{g(X)} \right]
\]
and
\[
\frac{S_{\text{log},k}^g(X_1,\ldots,X_n) - \mathbb{E} S_{\text{log},k}^g(X_1,\ldots,X_n)}{n^{1/2}} \overset{D}{\to} N \left( 0, \tau_k + \text{Var}[\log(f(X)/g(X))] \right).
\]

**Remarks.** When \( X_i \) are i.i.d. uniform on \([0,1]\), and when \( g = 1 \), then the CLT for \( S_{\text{log},k}^g(X_1,\ldots,X_n) \) was established by Darling (sect. 7 of [16]) for \( k = 1 \) and later by Holst [24] and Cressie [12] for general \( k \). When the \( X_i \) have a step density then Cressie shows asymptotic normality of \( S_{\text{log},k}^g(X_1,\ldots,X_n) \) including cases when \( k \to \infty \). Czekala (Thm. 1 of [15]) apparently re-discovered Cressie’s result. Shao and Hahn [44] treat general densities for \( k = 1 \), although their proof depends upon interchanging limits in order to pass from step densities to arbitrary densities. When \( k = 1 \), Blumenthal (Thm. 2 of [10]), proves Corollary 2.5 for densities \( f \) satisfying special conditions. Corollary 2.4 extends all of these results to general \( f \) and \( g \), resolving a conjecture of Darling ([16], p. 249) affirmatively.

### 2.6.2 Information gain of order \( r \) (\( d = 1 \))

Let \( \phi(x) = x^r, \ r > 0 \). We write \( S_r^g \) to denote \( S_{\phi}^g \), Rényi’s information gain (I-divergence) of order \( r \) in \( d = 1 \). Denote the associated point measures by
\[
\mu_{n,r}^g := \sum_{i=1}^{n-1} (n \cdot [G(X_{i+k}) - G(X_i)])^r \delta X_i.
\]

Let \( V_r := -2r \Gamma(2r+1) + \Gamma(2r+1) \) and \( D_r := \Gamma(r+1)(1-r) \). It is a simple matter to verify via (2.13) and (2.14), respectively, that for all \( \beta > 0 \)
\[
V_r^S(\beta) := \lim_{\lambda \to \infty} \frac{\text{Var}[S_r^g(P_\lambda \cap [0,1])]}{\lambda} = V_r \beta^{2r}
\]
and
\[
\Delta_r^S(\beta) := 2 \mathbb{E} \phi(\beta \Gamma_1) - \mathbb{E} \phi(\beta \Gamma_2) = D_r \beta^r.
\]
Using these values for $V^S_\beta$ and $\Delta^S_\beta$, Theorem 2.5 yields the following improvement upon Theorem 2.3. Put

$$
\sigma^2_r(f, g) := V^r_r \int_A \frac{g(x)}{f(x)} \frac{g(x)}{f(x)} f(x) dx - D^2_r \left( \int_A \frac{g(x)}{f(x)} f(x) dx \right)^2.
$$

As usual, $X_1, X_2, \ldots$ are i.i.d. with density $f$ on $A$.

**Corollary 2.5** (Gaussian limits for information gain, general densities) As $n \to \infty$

$$
\frac{\text{Var}[S^\beta_r(X_1, \ldots, X_n)]}{n} \to \sigma^2_r(f, g)
$$

and

$$
\frac{S^\beta_r(X_1, \ldots, X_n) - E S^\beta_r(X_1, \ldots, X_n)}{n^{1/2}} \overset{d}{\to} N(0, \sigma^2_r(f, g)),
$$

while $n^{-1/2} \rho_{n,r}$ converges in law to a Gaussian field with covariance kernel

$$
V^r_r \int_A h_1(x)h_2(x) \left( \frac{g(x)}{f(x)} \right)^2 f(x) dx - D^2_r \int_A h_1(x) \left( \frac{g(x)}{f(x)} \right)^r f(x) dx \int_A h_2(x) \left( \frac{g(x)}{f(x)} \right)^r f(x) dx.
$$

**Remarks.** It is easy to verify using [8] that $\sigma^2_r(f, g) > 0$ except when $r = 1$. Corollary 2.5 extends upon the CLTs of Darling [16] (uniform case) and Weiss [47]. The latter assumes that $GF^{-1}$ has a step function derivative and incorrectly asserts that his results hold in more general situations (cf. p. 417 of Pyke [38]). Moran [30] proved a CLT for the statistic $S_2$ over uniform random variables.

### 2.6.3 Limit theory for log-likelihood ratio statistics ($d = 1$)

Let $\phi(x) = x \log x$. Using the relations

$$
E[(\beta \Gamma_1 \log(\beta \Gamma_1))] = \beta \log \beta + \beta (1 - \gamma) \quad \text{and} \quad E[(\beta \Gamma_2 \log(\beta \Gamma_2))] = 2 \beta \log \beta + \beta (3 - 2 \gamma),
$$

as well as

$$
E[(\beta \Gamma_1 \log(\beta \Gamma_1))^2] = \beta^2 [2(\log \beta)^2 + (6 - 4 \gamma) \log \beta + 2 + \pi^2/3 - 6 \gamma + 2 \gamma^2],
$$

it is a simple matter to verify via (2.13) and (2.14), respectively, that for all $\beta > 0$

$$
V^S_\phi(\beta) := \lim_{\lambda \to \infty} \frac{\text{Var}[S^\beta_\phi(\mathcal{P}_\lambda \cap [0, 1])]}{\lambda} = \left( \frac{\pi^2}{3} - 2 \right) \beta^2
$$

and that

$$
\Delta^S_\phi(\beta) := 2E \phi(\beta \Gamma_1) - E \phi(\beta \Gamma_2) = -\beta.
$$
Put
\[ \sigma_\phi^2(f, g) := \left( \frac{\pi^2}{3} - 2 \right) \int_A \frac{g^2(x)}{f(x)} dx - \left( \int_A g(x) dx \right)^2 \]
and note that since \( g \) is a density we have
\[ \sigma_\phi^2(f, g) = \left( \frac{\pi^2}{3} - 2 \right) \text{Var} \left[ \frac{g(X)}{f(X)} \right] + \frac{\pi^2}{3} - 3. \]

When \( \phi(x) = x \log x \), we write \( \mu_{n,\phi}^g \) for the point measure
\[ \sum_{i=1}^{n-1} \phi(n \cdot [G(X_{i+k}) - G(X_{i})]) \delta_{X_i}. \]

The following is an improvement upon Theorem 2.4 in the setting \( d = 1 \). Let \( X_1, X_2, \ldots \) be i.i.d. with density \( f \) on \( A \).

**Corollary 2.6** (Gaussian limit for log likelihood statistic, general densities) As \( n \to \infty \)
\[ \frac{\text{Var}[S_\phi^g(X_1, \ldots, X_n)]}{n} \to \sigma_\phi^2(f, g) \]
and
\[ \frac{S_\phi^g(X_1, \ldots, X_n) - \mathbb{E}S_\phi^g(X_1, \ldots, X_n)}{n^{1/2}} \xrightarrow{D} N(0, \sigma_\phi^2(f, g)) \]
while \( n^{-1/2} \mu_{n,\phi}^g \) converges in law to a Gaussian field with covariance kernel
\[ \left( \frac{\pi^2}{3} - 2 \right) \int_A h_1(x)h_2(x) \frac{g^2(x)}{f(x)} dx - \int_A h_1(x)g(x) dx \int_A h_2(x)g(x) dx. \]

**Remarks.** Corollary 2.6 extends the results of Gebert and Kale [20], who assume uniformity of \( X_i, \ i \geq 1 \), and Czekala (Thm. 2 of [15]), who assumes that \( X_i, \ i \geq 1 \), have a step density. van Es [46] establishes asymptotic normality for \( S_\phi^g \) whenever \( k, n \to \infty \), \( k = o(n^{1/2}) \), and \( f \) is Lipschitz on its support \( A \).

### 2.7 Number of pairs of sample points within a fixed distance

Instead of considering spacing statistics defined with respect to the sample points, we now define a statistic using cells of fixed radius \( t \). Thus, for all \( t > 0 \) and all finite point sets \( \mathcal{X} \) and \( x \in \mathcal{X} \) we set \( \xi^t(x, \mathcal{X}) := \text{card}(B_t(x) \cap \mathcal{X}) \) and for all \( \lambda > 0 \) we define \( \xi^t_\lambda(x, \mathcal{X}) := \xi^t(\lambda^{1/d}x, \lambda^{1/d}\mathcal{X}) \). The statistic
\[ H^t(\mathcal{X}) := \frac{1}{2} \sum_{x \in \mathcal{X}} \xi^t(x, \mathcal{X}) \]
counts the total number of pairs of points in $\mathcal{X}$ distant $t$ from each other. Define the scaled version $H^\lambda_t(\mathcal{X}) := H^\lambda(\lambda^{1/d}\mathcal{X})$. Given i.i.d. random variables $X_1, ..., X_n$ with density $f : A \to \mathbb{R}$, we seek the asymptotic distribution of the functional $H^\lambda_t(X_1, ..., X_n)$ as well as that of the point measure

$$
\mu^n_t := \sum_{i=1}^n \xi^n_t(X_i, \{X_i\}_{i=1}^n)\delta_{X_i}.
$$

Let $\omega_d$ be the volume of the unit ball in $\mathbb{R}^d$ so that $v_t := t^d \omega_d$ stands for the volume a $d$-dimensional ball of radius $t$. For all $f \in B(A)$ and $t > 0$ put

$$
\sigma_t^2(f) := v_t \int_A f(x)^2 dx + v_t^2 \left(2 \int_A f(x)^3 dx - (\int_A f(x)^2 dx)^2\right).
$$

The proof of the following CLT is deferred to section six.

**Corollary 2.7** (Gaussian limit for the number of pairs of points within distance $t$) As $n \to \infty$

$$
\frac{\text{Var}[H^\lambda_t(X_1, ..., X_n)]}{n} \to \sigma^2_t(f) \quad (2.15)
$$

and

$$
\frac{H^\lambda_t(X_1, ..., X_n) - \mathbb{E} H^\lambda_t(X_1, ..., X_n)}{n^{1/2}} \overset{d}{\to} N(0, \sigma^2_t(f))
$$

while $n^{-1/2} \mu_{n,t}$ converges in law to a Gaussian field with covariance kernel

$$
\int_A h_1(x)h_2(x)[v_t(f(x))^2 + 2v_t^2(f(x))^3]dx - v_t^2 \int_A h_1(x)f(x)^2 dx \int_A h_2(x)f(x)^2 dx.
$$

**Remarks.** The statistic $H^\lambda_t$ has been considered by various authors. L’Ecuyer et al. [19] considers $H^\lambda_t$ from the point of view of multidimensional goodness-of-fit tests, but restricts attention to uniform samples. Penrose [33] (Ch.4) proves that the finite dimensional distributions of the process $H^\lambda_t(X_1, ..., X_n), t > 0$, converge weakly to those of a Gaussian process without the explicit formulas above.

### 3 Proof of Theorem 2.1

We first show a Poissonized version of (2.5), namely we show that for all $h \in B(A)$,

$$
\lim_{\lambda \to \infty} \lambda^{-1} \text{Var}[\langle h, \mu^n_{\lambda, \phi} \rangle] = \int_A (h(x))^2 \varphi(x) \left(\frac{g(x)}{f(x)}\right) f(x) dx. \quad (3.1)
$$
3.1 Variance convergence over Poisson samples

Recall that for all $\lambda > 0$, $P^f_\lambda$ is a Poisson point process on $A$ with intensity measure $\lambda f : A \to \mathbb{R}^+$. For all $\lambda > 0$, let $P'^f_\lambda$ be a Poisson point process equidistributed with and independent of $P^f_\lambda$, i.e., $P'^f_\lambda$ is a copy of $P^f_\lambda$. To simplify the notation, for all $\lambda > 0$ and all $x \in A$ define

$$\xi_\lambda(x) := \xi_\lambda^g(x) := \xi_\lambda^g(x, P^f_\lambda) := \phi \left( \lambda \int_{C(x, P^f_\lambda)} g(u)du \right)$$

so that (2.7) and (2.8) become

$$N^g_\lambda(P^f_\lambda) = \sum_{x \in P^f_\lambda} \xi_\lambda^g(x, P^f_\lambda)$$

and

$$\mu^g_{\lambda, \phi} = \sum_{x \in P^f_\lambda} \xi_\lambda^g(x, P^f_\lambda) \delta_x.$$

Define also

$$\xi'_\lambda(x) := \xi'_\lambda^g(x) := \xi'_\lambda^g(x, P'^f_\lambda) := \phi \left( \lambda \int_{C(x, P'^f_\lambda)} g(u)du \right)$$

and

$$\xi'_\lambda(y) := \xi'_\lambda^g(y, P_{\lambda f(x)}) := \phi \left( \lambda \int_{C(y, P_{\lambda f(x)})} g(x)du \right)$$

and

$$\xi'_\lambda(y) := \phi \left( \lambda \int_{C(y, P'^f_{\lambda f(x)})} g(x)du \right).$$

We refer to $\xi'_\lambda(y)$ and $\xi'_\lambda^g(y)$ as first order approximations to $\xi_\lambda^g(y)$ and $\xi'_\lambda^g(y)$, in that $g$ is approximated by the scalar $g(x)$. When $x$ is close to $y$, the continuity of $\phi$ and $g$ ensures that this is a good approximation. We seek to approximate $\mu^g_{\lambda, \phi}$ by the first order approximation

$$\sum_{x \in P^f_\lambda} \xi'_\lambda(x, P_{\lambda f(x)}) \delta_x.$$

Given $x \in \mathbb{R}^d$, notice that for all $y \in \mathbb{R}^d$, $\xi'_\lambda(y) \overset{\mathbb{E}}{=} \phi(\frac{g'(x)}{d x}) \Gamma_1$. By assumption $\mathbb{E} |\xi'^g_\lambda(x)|^p < \infty$ for all $x \in \mathbb{R}^d$ and all $p > 0$. We will use these bounds heavily in all that follows.

Notice that when $g$ is not constant, we have $\xi'_\lambda^g(x, P^f_\lambda) \neq \xi'_\lambda^g(x - y, P^f_\lambda - y)$, that is the functions $\xi'_\lambda^g(\cdot, \cdot)$ are not translation invariant.

For every $x \in \mathbb{R}^d$ and $\tau > 0$ there exists an a.s. finite random variable $R := R^f_x$ such that

$$\phi \left( \int_{C(x, P \cap B_R(x) \cup A)} du \right)$$

is independent of $A$ for all finite $A \subset \mathbb{R}^d \setminus B_R(x)$. Likewise, given $f$ and $g$ bounded away from zero and infinity, for all $x \in A$, there exists a random variable $R := R^f_x := R^f_x \lambda$, called a radius of stabilization for $\xi^g$ with respect to $P^f_\lambda$ at $x$ such that

$$\xi^g(\lambda^{1/d}x, \lambda^{1/d}P^f_\lambda \cap B_R(\lambda^{1/d}x) \cup A)$$

is independent of $A$ for all finite $A \subset \lambda^{1/d}A \setminus B_R(\lambda^{1/d}x)$. Also, the bounds on $f$ and $g$ imply that $R^f_x$ have uniformly exponentially decaying tails (section 6 of [34]). In other words, the $\xi$ are ‘exponentially stabilizing’.
To show (3.1), we will show for all $h \in B(A)$ that
\[
\lambda^{-1}\text{Var}([h, \mu^\lambda_{x, \phi}]) = \lambda \int_{A \times A} h(x)h(y)\mathbb{E} \left[ \xi^\lambda_P(x, P^\lambda_A \cup y)\xi^\lambda_P(y, P^\lambda_A \cup x) - \xi^\lambda_P(x, P^\lambda_A)\xi^\lambda_P(y, P^\lambda_A) \right] f(x)f(y)dxdy + \\
+ \int_A h(x)^2 E[(\xi^\lambda_P(x, P^\lambda_A))^2]f(x)dx.
\] is approximated by its ‘first order version’
\[
\lambda \int_{A \times \mathbb{R}^d} h(x)h(y)\mathbb{E} \left[ \xi^\lambda_P(x, P_{\lambda f(x)} \cup y)\xi^\lambda_P(y, P_{\lambda f(x)} \cup x) - \xi^\lambda_P(x, P^\lambda_A)\xi^\lambda_P(y, P^\lambda_A) \right] f(x)f(y)dxdy + \\
+ \int_A h(x)^2 E[(\xi^\lambda_P(x, P_{\lambda f(x)})^2)]f(x)dx
\] and then we show that this first order version reduces to $\int_A (h(x))^2 V_\phi(\frac{q(x)}{f(x)}) f(x)dx$.

For all $\lambda \in \mathbb{R}^+$ and $x \in A$ we introduce two auxiliary homogeneous independent Poisson point processes $P_{\lambda f(x)}$ and $P'_{\lambda f(x)}$ such that:

- $P_{\lambda f(x)}, P'_{\lambda f(x)}$ have constant intensity density on $A$ equal to $\lambda f(x)$,
- $P^\lambda_A$ and $P_{\lambda f(x)}$ are coupled in the sense that for any $B \in \mathcal{B}(A)$,
  \[ P \left[ P^\lambda_A(B) \neq P_{\lambda f(x)}(B) \right] \leq \lambda \int_B |f(y) - f(x)|dy, \] (3.4) and the same is true for $P'_{\lambda f(x)}$ and $P_{\lambda f(x)}$.

We will measure the closeness of (3.2) and (3.3) in terms of correlation functions for $\xi^\lambda_P(x)$:

\[ q^\lambda_P(x) := \mathbb{E} [\xi^\lambda_P(x)], \quad x \in \mathbb{R}^d, \]
\[ c^\lambda_P(x, y) := \mathbb{E} \left[ \xi^\lambda_P(x, P^\lambda_A \cup y)\xi^\lambda_P(y, P^\lambda_A \cup x) - \xi^\lambda_P(x, P^\lambda_A)\xi^\lambda_P(y, P^\lambda_A) \right], \quad x, y \in \mathbb{R}^d, \quad x \neq y, \]

as well as the corresponding correlation functions for the first order approximations:
\[ q^\lambda_P(x) := \mathbb{E} [(\xi^\lambda_P(x))^2], \quad x \in \mathbb{R}^d, \]
\[ c^\lambda_P(x, y) := \mathbb{E} \left[ \xi^\lambda_P(x, P_{\lambda f(x)} \cup y)\xi^\lambda_P(y, P_{\lambda f(x)} \cup x) - \xi^\lambda_P(x, P_{\lambda f(x)})\xi^\lambda_P(y, P_{\lambda f(x)}) \right], \quad x, y \in \mathbb{R}^d, \quad x \neq y. \]

Write simply $q(x)$ and $c(x, y)$ for $q_1(x)$ and $c_1(x, y)$, respectively, and similarly for $q^\lambda_P$ and $c^\lambda_P$.

Since $\lambda \int_{C(x, P_{\lambda f(x)})} du \overset{\mathcal{P}}{=} \int_{C(x, P_{f(x)})} du$ holds for all $\lambda > 0$, it follows that
\[ q^\lambda_P(x) = q^\lambda(x) \quad \text{and} \quad c^\lambda_P(x, x + y) = c^\lambda(x, x + \lambda^{1/d}y) \]
for all $x, y$ and $\lambda$. By the bounds on $g$ and $f$ and the definition of $\Phi$ we have
\[ \sup_{x \in \mathbb{R}^d} [q_\lambda(x), q^\lambda_P(x)] \leq C. \] (3.5)
Similarly, applying Cauchy-Schwarz,

$$\sup_{x,y \in \mathbb{R}^d} \sup_{\lambda} \left| \langle c_{\lambda}(x,y), c_{\lambda}^T(x,y) \rangle \right| \leq C.$$  \hfill (3.6)

We next describe some properties of the correlation functions.

**Lemma 3.1** The correlations $c_{\lambda}(x,y)$ and $c_{\lambda}^T(x,y)$ satisfy the bounds:

$$|c_{\lambda}(x,y)| \leq C \exp(-C \lambda^{1/d}|x-y|)$$  \hfill (3.7)

and

$$|c_{\lambda}^T(x,y)| \leq C \exp(-C \lambda^{1/d}|x-y|).$$  \hfill (3.8)

**Proof.** We prove the bound (3.7); the proof of (3.8) is identical. Let $R_x$ and $R_y$ be the radii of stabilization for $x$ and $y$. $R_x$ and $R_y$ are both less than $\lambda^{1/d}|x-y|/2$ except on a set with probability at most $2 \exp(-C \lambda^{1/d}|x-y|)$. Let $E$ denote the event that $R_x$ and $R_y$ are both less than $\lambda^{1/d}|x-y|/2$. On $E$ the sets $\mathcal{P}_x^f \cap \mathcal{B}_{R\lambda-1/4}(x)$ and $\mathcal{P}_y^f \cap \mathcal{B}_{R\lambda-1/4}(y)$ do not intersect and thus on $E$ we have $\xi_{\lambda}^0(x,\mathcal{P}_x^f \cup y) = \xi_{\lambda}^0(x,\mathcal{P}_x^f)$ and similarly $\xi_{\lambda}^0(y,\mathcal{P}_y^f \cup x) = \xi_{\lambda}^0(y,\mathcal{P}_y^f)$ and therefore

$$|c_{\lambda}(x,y)| \leq \mathbb{E} \left[ |\xi_{\lambda}^0(x,\mathcal{P}_x^f \cup y)\xi_{\lambda}^0(y,\mathcal{P}_y^f \cup x) - \xi_{\lambda}^0(x,\mathcal{P}_x^f)\xi_{\lambda}^0(y,\mathcal{P}_y^f) \cdot 1_E | \right].$$

Hölder’s inequality and the moment bound $\mathbb{E}[|\xi_{\lambda}(x)|^3] < \infty$ yield $|c_{\lambda}(x,y)| \leq C(\mathbb{P}[E^c])^{1/3}$. The desired bounds (3.7) follow. \hfill \Box

Compactness of $A$ and the continuity of $f$ and $h$ implies uniform continuity, so we fix moduli of continuity $t_f, t_h : \mathbb{R}^+ \to \mathbb{R}^+$ such that for any $x,y \in A : |x-y| \leq \delta$, $|f(x) - f(y)| \leq t_f(\delta)$ and $|h(x) - h(y)| \leq t_h(\delta)$.

The next coupling lemma shows that the correlation functions $q_{\lambda}$ and $c_{\lambda}$ are uniformly approximated by the first order approximations $c_{\lambda}^T$ and $q_{\lambda}^T$, respectively.

**Lemma 3.2** Let $\phi \in \Phi$. There exists a function $\delta : \mathbb{R}^+ \to \mathbb{R}^+$, increasing to $\infty$, such that as $\lambda \to \infty$, $\delta(\lambda)/\lambda \to 0$ and

(i) $\forall x \in A$ distant at least $(\delta/\lambda)^{1/d}$ from $\partial A$, $|q_{\lambda}(x) - q_{\lambda}^T(x)| \to 0,$

(ii) $\forall x, y \in A$ distant at least $(\delta/\lambda)^{1/d}$ from $\partial A,$

$$\delta(\lambda)|c_{\lambda}(x,y) - c_{\lambda}^T(x,y)| \to 0,$$

(iii) $\delta(\lambda)t_f((\delta(\lambda)/\lambda)^{1/d}) \to 0$ and $\delta(\lambda)t_h((\delta(\lambda)/\lambda)^{1/d}) \to 0.$

The following lemma establishes the variance limit (3.2).
Lemma 3.3 Let $\phi \in \Phi$ and $h \in B(A)$. Then
\[
\lim_{\lambda \to \infty} \lambda^{-1} \text{Var}[h, \mu^\phi_\lambda] = \int_A (h(x))^2 V_\phi \left( \frac{g(x)}{f(x)} \right) f(x) dx. \tag{3.9}
\]
The integral on the right hand side of (3.9) is finite.

**Proof.** Using the definition of $q_\lambda$ and $c_\lambda$, we may rewrite the integral (3.2) as
\[
\int_A h(x)^2 q_\lambda(x) f(x) dx + \lambda \int \int_{A \times A} h(y) h(x) c_\lambda(x, y) f(x) f(y) dy dx =
\int_A h(x) f(x) \left[ h(x) q_\lambda(x) + \lambda \int_A h(y) c_\lambda(x, y) f(y) dy \right] dx.
\]

Let $\delta := \delta(\lambda)$ be as in Lemma 3.2. We want to compare the bracketed expression in the last integral, namely
\[
h(x) q_\lambda(x) + \lambda \int_A h(y) c_\lambda(x, y) f(y) dy \tag{3.10}
\]
with its corresponding uniform version
\[
h(x) q_\lambda^u(x) + \lambda \int \int_{\mathbb{R}^d} h(x) c_\lambda^u(x, y) f(y) dy dx.
\tag{3.11}
\]
It will be enough to show for all $x \in A$ distant at least $2(\delta/\lambda)^{1/d}$ from $\partial A$, that the difference of the expressions (3.10) and (3.11) converges to zero as $\lambda \to \infty$.

Lemma 3.2(i) implies for all $x$ distant at least $(\delta/\lambda)^{1/d}$ from $\partial A$ that the difference of the first terms in (3.10) and (3.11) goes to zero as $\lambda \to \infty$. The difference of the integrals in (3.10) and (3.11) equals:
\[
\lambda \int_{\mathbb{R}^d} [c_\lambda(x, y) h(y) f(y) - c_\lambda^u(x, y) h(x) f(x)] dy. \tag{3.12}
\]

To evaluate the integral (3.12), we integrate separately over $B(\delta/\lambda)^{1/d}(x)$ and $\mathbb{R}^d \setminus B(\delta/\lambda)^{1/d}(x)$. The integral over $B(\delta/\lambda)^{1/d}(x)$ involves the difference
\[
\lambda \int_{B(\delta/\lambda)^{1/d}(x)} [c_\lambda(x, y) h(y) f(y) - c_\lambda^u(x, y) h(x) f(x)] dy,
\]
which we split as
\[
\lambda \int_{B(\delta/\lambda)^{1/d}(x)} (c_\lambda(x, y) - c_\lambda^u(x, y)) h(y) f(y) dy + \lambda \int_{B(\delta/\lambda)^{1/d}(x)} c_\lambda^u(x, y) (h(y) - h(x)) f(y) dy + \lambda \int_{B(\delta/\lambda)^{1/d}(x)} c_\lambda^u(x, y) h(x) (f(y) - f(x)) dy. \tag{3.13}
\]
The first integral is bounded by the product of $\lambda$, the volume of $B_{(\delta/\lambda)^{1/d}}(x)$, and the maximum of the integrand $(c_\lambda(x, y) - c_\lambda^*(x, y))h(y)f(y)$. However, since $y$ is distant at least $(\delta/\lambda)^{1/d}$ from $\partial A$, the product goes to zero by Lemma 3.2(ii). The second and third integrals also tend to zero as $\lambda \to \infty$ by the bound (3.6) and Lemma 3.2(iii).

Since $f$ and $h$ are bounded, the integral in (3.12) over $\mathbb{R}^d \setminus B_{(\delta/\lambda)^{1/d}}(x)$ is bounded by

$$C \int_{\mathbb{R}^d \setminus B_{(\delta/\lambda)^{1/d}}(x)} [c_\lambda(x, y) + c_\lambda^*(x, y)]d(\lambda^{1/d}y)$$

which by Lemma 3.1 is bounded by

$$C \int_{\mathbb{R}^d \setminus B_{(\delta/\lambda)^{1/d}}(x)} \exp(-C|z - x|) + \exp(-C|z - x|)dz.$$

Recalling that $\omega_d$ denotes the volume of the unit ball in $\mathbb{R}^d$, this last integral is bounded by $C_\delta^\infty f_{\delta/\lambda}^d \exp(-Ct)f^{d-1}dt$, which tends to zero as $\lambda \to \infty$ since $\delta \to \infty$. We conclude that (3.12) converges to 0 uniformly in $x \in A$. Hence,

$$\lambda^{-1}\text{Var}[\langle \mu_{\lambda, \phi}^h \rangle, h] - \int_{A} f(x)h(x) \left[ h(x)q_\lambda^*(x) + \lambda \int_{\mathbb{R}^d} h(x)c_\lambda^*(x, x + y)f(x)dy \right]dx$$

converges to 0 as $\lambda \to \infty$. Using the equivalence $q_\lambda^*(x) = q^*(x)$ and $c_\lambda^*(x, x + y) = c^*(x, x + \lambda^{1/d}y)$ we see that

$$\lambda^{-1}\text{Var}[h, \mu_{\lambda, \phi}^h] \to \int_{A} h(x) \left[ h(x)q^*(x) + \int_{\mathbb{R}^d} h(x)c^*(x, y)f(x)dy \right]f(x)dx.$$

If we take $A = [0, 1]^d$, fix $x \in A$, and set $h = 1$, then these arguments show that for all $x \in A$

$$\lim_{\lambda \to \infty} \frac{\text{Var}[X^{g(x)}(\mathcal{P}_{\lambda}^1(x)) \cap [0, 1]^d]}{\lambda} = \left[ q^*(x) + \int_{\mathbb{R}^d} c^*(x, y)f(x)dy \right]f(x).$$

Setting $V(\frac{g(x)}{f(x)}) = q^*(x) + \int_{\mathbb{R}^d} c^*(x, y)f(x)dy$, this yields the desired variance convergence (3.9). This also furnishes a proof of the limit (2.4).

Now we return to the

Proof of Lemma 3.2. To prove implication (i) we need to show for all $x \in A$ distant at least $(\delta/\lambda)^{1/d}$ from $\partial A$, that

$$\left| \mathbb{E} \phi^2 \left( \lambda \int_{C(x, \mathcal{P}_{\lambda}^x)} g(u)du \right) - \mathbb{E} \phi^2 \left( \lambda \int_{C(x, \mathcal{P}_{\lambda}(x))} g(x)du \right) \right| \to 0$$

as $\lambda \to \infty$. Let $x' := x_{\lambda} \in C(x, \mathcal{P}_{\lambda}^x)$ satisfy

$$\int_{C(x, \mathcal{P}_{\lambda}^x)} g(u)du = g(x') \int_{C(x, \mathcal{P}_{\lambda}^x)} du.$$  (3.15)
Such a point exists because of the continuity of \( g \) and the mean value theorem for integrals.

Also, \( g(x')_\lambda - g(x') \xrightarrow{P} 0 \) as \( \lambda \to \infty \) and by the boundedness of \( g \) we have \( \mathbb{E} |g(x')_\lambda - g(x')|^2 \to 0. \) Chebyshev’s inequality and Cauchy-Schwarz imply that as \( \lambda \to \infty, \)

\[
\lambda g(x') \int_{C(x, P'_\lambda(x))} du - \lambda g(x) \int_{C(x, P'_\lambda(x))} du \xrightarrow{P} 0,
\]

(3.16)

since \( \lambda \int_{C(x, P'_\lambda(x))} du \) has a finite second moment. Now it may be shown (Lemma 5.1 of [28]) for all \( x \in A \) distant at least \((\delta / \lambda)^{1/d}\) from \( \partial A, \)

\[
\lambda g(x) \int_{C(x, P'_\lambda(x))} du \xrightarrow{P} \frac{g(x)}{f(x)} \Gamma_1.
\]

(3.17)

Combining (3.16)-(3.18) gives as \( \lambda \to \infty, \)

\[
\lambda \int_{C(x, P'_\lambda(x))} g(u)du \xrightarrow{P} \frac{g(x)}{f(x)} \Gamma_1
\]

and thus by continuity of \( \phi: \)

\[
\phi^2 \left( \lambda \int_{C(x, P'_\lambda(x))} g(u)du \right) \xrightarrow{P} \phi^2 \left( \frac{g(x)}{f(x)} \Gamma_1 \right).
\]

Now for all \( x \in A \)

\[
\sup_{\lambda \ge 1} \mathbb{E} \left[ \phi^2 \left( \lambda \int_{C(x, P'_\lambda(x))} g(u)du \right) \right] < \infty,
\]

since \( \phi \in \Phi. \) It follows that for each \( x, \) \( \phi^2 \left( \lambda \int_{C(x, P'_\lambda(x))} g(u)du \right), \lambda > 0, \) are uniformly integrable, showing that (3.14) holds and so Lemma 3.2(i) holds.

Now we prove implications (ii) and (iii). Clearly, there is a function \( \delta := \delta(\lambda) \to \infty \) as \( \lambda \to \infty \) such that implication (iii) is true and, by exponential decay of \( c_\lambda, \) that

\[
\delta(\lambda) \sup_{|x-y| \ge \delta \lambda^{-1/d}} c_\lambda(x, y) \to 0 \quad \text{and} \quad \delta(\lambda) \sup_{|x-y| \ge \delta \lambda^{-1/d}} c_\lambda^\xi(x, y) \to 0.
\]

From the definition of \( c_\lambda(x, y) \) and \( c_\lambda^\xi(x, y), \) it is enough to show that

\[
\sup_{|x-y| < \delta \lambda^{-1/d}} \delta(\lambda) |\mathbb{E} [\xi^\xi(x, P'_{\lambda} \cup y) \xi^\xi(y, P'_{\lambda} \cup x) - \xi^\xi(x, P_{\lambda f(x)} \cup y) \xi^\xi(y, P_{\lambda f(x)} \cup x)]| \to 0 \quad (3.18)
\]

and

\[
\sup_{|x-y| < \delta \lambda^{-1/d}} \delta(\lambda) |\mathbb{E} [\xi_\lambda^\xi(x, P'_{\lambda}) \xi_\lambda^\xi(y, P'_{\lambda} - \xi_\lambda^\xi(x, P_{\lambda f(x)}) \xi_\lambda^\xi(y, P_{\lambda f(x)})]| \to 0 \quad (3.19)
\]

as \( \lambda \to \infty. \) We will first show (3.18); (3.19) has a similar proof.

Let \( x \in A \) be distant at least \((\delta / \lambda)^{1/d}\) from \( \partial A. \) Let \( A(x, \lambda, \delta) \) be the event that \( P_{\lambda}^I = P_{\lambda f(x)} \) on the ball \( B_x((\delta / \lambda)^{1/d}) \) and that the radii of stabilization \( R_{x, \lambda} \) for \( \lambda^{1/d}x \) is less than \( \delta. \) The coupling estimate (3.4) and the exponential decay of the tails of \( R_{x, \lambda} \) imply that

\[
P[A(x, \lambda, \delta)^c] \le C[\omega_d \delta t_f((\delta / \lambda)^{1/d}) + \exp(-C\delta)].
\]

(3.20)
Now bound the absolute value in (3.18) as

$$\left| \mathbb{E} [\xi_2^x(x, \mathcal{P}_\lambda^f \cup y) \xi_2^y(y, \mathcal{P}_\lambda^A \cup x) - \xi_2^x(x, \mathcal{P}_\lambda^{af(x)} \cup y) \xi_2^y(y, \mathcal{P}_\lambda^{af(x)} \cup x)] \right|$$

(3.21)

$$\leq \left| \mathbb{E} [(\xi_2^x(x, \mathcal{P}_\lambda^f \cup y) - \xi_2^x(x, \mathcal{P}_\lambda^{af(x)} \cup y)) \cdot \xi_2^y(y, \mathcal{P}_\lambda^A \cup x)] + \mathbb{E} [(\xi_2^y(y, \mathcal{P}_\lambda^A \cup x) - \xi_2^x(x, \mathcal{P}_\lambda^{af(x)} \cup x)) \cdot \xi_2^x(x, \mathcal{P}_\lambda^{af(x)})] \right|.$$  

(3.22)

Since $\mathbb{E} [(\xi_2^y(y, \mathcal{P}_\lambda^A \cup x))^2] < \infty$ and $\mathbb{E} [(\xi_2^x(x, \mathcal{P}_\lambda^{af(x)})]^2] < \infty$ for all $y$, it suffices by Cauchy-Schwarz to show for all $x \in A$ that as $\lambda \to \infty$

$$\delta(\lambda) \mathbb{E} \left[ \left| \xi_2^x(x, \mathcal{P}_\lambda^f \cup y) - \xi_2^x(x, \mathcal{P}_\lambda^{af(x)} \cup y) \right|^2 \right] \to 0$$

(3.23)

and that

$$\sup_{|x-y| < \delta^{-1/4}} \delta(\lambda) \mathbb{E} \left[ \left| \xi_2^x(x, \mathcal{P}_\lambda^f \cup x) - \xi_2^x(y, \mathcal{P}_\lambda^{af(x)} \cup x) \right|^2 \right] \to 0.$$  

Now we show (3.22). Now the left-hand side of (3.22) equals

$$\mathbb{E} \left[ \phi \left( \lambda \int_{C(x, \mathcal{P}_\lambda^f \cup y)} g(u) du \right) - \phi \left( \lambda \int_{C(x, \mathcal{P}_\lambda^{af(x)} \cup y)} g(x) du \right) \right]^2$$

which is bounded by the sum of

$$\mathbb{E} \left[ \phi \left( \lambda \int_{C(x, \mathcal{P}_\lambda^f \cup y)} g(u) du \right) - \phi \left( \lambda \int_{C(x, \mathcal{P}_\lambda^{af(x)} \cup y)} g(x) du \right) \right]^2 1_{A(x, \lambda)}$$

and

$$\mathbb{E} \left[ \phi \left( \lambda \int_{C(x, \mathcal{P}_\lambda^f \cup y)} g(u) du \right) - \phi \left( \lambda \int_{C(x, \mathcal{P}_\lambda^{af(x)} \cup y)} g(x) du \right) \right]^2 1_{A^c(x, \lambda)}.$$  

The second expectation goes to zero by Cauchy-Schwarz, $\phi \in \Phi$, and the estimate (3.20).

Since $C(x, \mathcal{P}_\lambda^f) \subset B_x((\delta/\lambda)^{1/4})$ on the set $A(x, \lambda, \delta)$, the definition of $A(x, \lambda, \delta)$ shows that the first term is bounded by

$$\mathbb{E} \left[ \phi \left( \lambda \int_{C(x, \mathcal{P}_\lambda^{af(x)} \cup y)} g(u) du \right) - \phi \left( \lambda \int_{C(x, \mathcal{P}_\lambda^{af(x)} \cup y)} g(x) du \right) \right]^2.$$  

The mean value theorem for integrals implies the existence of some $x_\lambda \in C(x, \mathcal{P}_\lambda^{af(x)} \cup y)$ such that the above equals

$$\mathbb{E} \left[ \left| \phi \left( g(x_\lambda) \cdot |C(x, \mathcal{P}_\lambda^{af(x)} \cup y)| \right) - \phi \left( g(x) \cdot |C(x, \mathcal{P}_\lambda^{af(x)} \cup y)| \right) \right|^2 \right].$$

Since $x_\lambda \overset{P}{\rightarrow} x$, the continuity of $\phi$ and $g$ imply that

$$\left| \phi \left( g(x_\lambda) \cdot |C(x, \mathcal{P}_\lambda^{af(x)} \cup y)| \right) - \phi(g(x) \cdot |C(x, \mathcal{P}_\lambda^{af(x)} \cup y)| \right|^2 \overset{P}{\rightarrow} 0.$$  

(3.24)

Since the left hand side of the above is uniformly integrable, it follows that

$$\lim_{\lambda \to 0} \mathbb{E} \left[ \left| \phi \left( g(x_\lambda) \cdot |C(x, \mathcal{P}_\lambda^{af(x)} \cup y)| \right) - \phi(g(x) \cdot |C(x, \mathcal{P}_\lambda^{af(x)} \cup y)| \right|^2 \right] = 0$$

(3.25)
and therefore there is some $\delta_0(\lambda) \to \infty$ such that
\[
\delta_0(\lambda) \mathbb{E} \left[ \left( \phi(g(x_\lambda')) - \phi(g(x)) \right)^2 \right] \to 0.
\]
This shows (3.22), changing our choice of $\delta(\lambda)$, if necessary.

Now we prove (3.23), i.e., we show
\[
\lim_{\lambda \to \infty} \sup_{|x - y| < \lambda^{-1/d}} \delta(\lambda) \mathbb{E} \left[ \left( \phi \left( \int_{C(y, P_{x,\lambda})} g(u) du \right) - \phi \left( \int_{C(y, P_{x,\lambda})} g(x) du \right) \right)^2 \right] = 0.
\]
By considering the set $A(x, \lambda, \delta)$ and by following the proof of (3.22) it is enough to show that
\[
\lim_{\lambda \to \infty} \sup_{|x - y| < \lambda^{-1/d}} \delta(\lambda) \mathbb{E} \left[ \left( \phi \left( \int_{C(y, P_{x,\lambda})} g(y) du \right) - \phi \left( \int_{C(y, P_{x,\lambda})} g(x) du \right) \right)^2 \right] = 0
\]
i.e., that
\[
\lim_{\lambda \to \infty} \sup_{|x - y| < \lambda^{-1/d}} \delta(\lambda) \mathbb{E} \left[ \phi \left( \frac{g(y)}{f(y)} \Gamma_1 \right) - \phi \left( \frac{g(x)}{f(x)} \Gamma_1 \right) \right]^2 = 0. \tag{3.26}
\]
The continuity of $f$, $g$ and $\phi$ yields for all $|x - y| < \lambda^{-1/d}$
\[
\left[ \phi \left( \frac{g(y)}{f(y)} \Gamma_1 \right) - \phi \left( \frac{g(x)}{f(x)} \Gamma_1 \right) \right]^2 \to 0.
\]
Since $\phi \in \Phi$, the dominated convergence theorem yields
\[
\mathbb{E} \left[ \phi \left( \frac{g(y)}{f(y)} \Gamma_1 \right) - \phi \left( \frac{g(x)}{f(x)} \Gamma_1 \right) \right]^2 \to 0.
\]
Changing our choice of $\delta(\lambda)$ if necessary, (3.26) follows. This completes the proof of Lemma 3.2 $\square$.

### 3.2 de-Poissonization: variance convergence over samples of fixed size

We show here how to pass from (3.2) to the de-Poissonized variance limit (2.5). Let $\mathcal{X}_m$ be the point process consisting of $m$ i.i.d. random variables with density $f$ on $A$. De-Poissonization involves coupling the $\mathcal{X}_m$, $m$ large, with a Poisson point process [34, 33, 8]. It will suffice to prove the following coupling lemma and then follow section six of [8] verbatim.

We first fix the terminology. For all $h \in B(A)$, let
\[
H_n^h(\mathcal{X}_m) := \sum_{i=1}^m \phi(n \cdot D_{i,n}^g) h(X_i)
\]
and put $R_{m,n}^h := H_n^h(\mathcal{X}_{m+1}) - H_n^h(\mathcal{X}_m)$. Define for all $\beta > 0$
\[
\Delta_{\phi}(\beta) := \lim_{R \to \infty} \mathbb{E} \left[ N_\phi^\beta(\mathcal{P} \cap B_R(0) \cup 0) - N_\phi^\beta(\mathcal{P} \cap \mathcal{B}_R(0)) \right].
\]
The existence of $\Delta_{\phi}(\beta)$ is guaranteed by Lemma 6.1 and Definition 2.1 of [34]. Note that $\mathbb{E} \left[ \Delta_{\phi}(\beta) \right] = \Delta_{\phi}(\beta)$. 

26
Lemma 3.4 Fix $\phi \in \Phi$ and let $\varepsilon > 0$. Let $h, g$ and $f$ be fixed. There exists $\delta > 0$ and $n_0 \geq 1$ such that for all $n \geq n_0$ and all $m, m' \in [1 - \delta)n, (1 + \delta)n]$ with $m < m'$, there exists random variables $X, X'$ with density $f$ and a coupled family of variables $D := D(X)$, $D' := D(X')$, $R := R(X, X')$, $R' := R'(X, X')$ with the following properties:

(i) $D$ and $D'$ each have the same distribution as $h(X)\Delta_{\phi}(\frac{f(X)}{f(X')})$,

(ii) $D$ and $D'$ are independent,

(iii) $(R, R')$ have the same joint distribution as $(R_{m,n}^h, R_{m',n}^h)$,

(iv) $P[|D - R| > \varepsilon] \cup |D' - R'| > \varepsilon| < \varepsilon$.

Proof. The existence of $D, D', R,$ and $R'$ as well as properties (i)-(iii) follow the proof of Lemma 6.1 of [8] (alternatively, see Theorem 2.16 of [33]). Suppose we are given $n$. Let $X, X', Y_1, Y_2, ...$ be i.i.d. random variables with density $f$ on $A$. On a suitable probability space, let $\mathcal{P} := \mathcal{P}^f_n$ and $\mathcal{P}' := \mathcal{P}'^f_n$ be independent Poisson processes on $A$ with intensity measure $nf(x)dx$.

Let $\mathcal{P}''$ be the point process consisting of those points of $\mathcal{P}$ which lie closer to $X$ than to $X'$ (in the Euclidean norm), together with those points of $\mathcal{P}'$ which lie closer to $X'$ than to $X$. Clearly $\mathcal{P}''$ is a Poisson process also having intensity measure $nf(x)dx$ on $A$, and moreover it is independent of $X$ and of $X'$.

Let $N$ denote the number of points of $\mathcal{P}''$ (a Poisson variable with mean $n|A|$). Choose an ordering on the points of $\mathcal{P}''$, uniformly at random from all $N!$ possible such orderings. Use this ordering to list the points of $\mathcal{P}''$ as $W_1, W_2, ..., W_N$. Also, set $W_{N+1} = Y_1, W_{N+2} = Y_2, W_{N+3} = Y_3$ and so on.

Let

$$R := R(X, X') := H^h_n(\{W_1, ..., W_m, X\}) - H^h_n(\{W_1, ..., W_m\})$$

and

$$R' := R(X, X') := H^h_n(\{W_1, ..., W_{m'-1}, X, X'\}) - H^h_n(\{W_1, ..., W_{m'-1}\}).$$

$x, X', W_1, W_2, W_3, ...$ are i.i.d. variables on $A$ with density $f$, and therefore the pairs $(R, R')$ and $(R_{m,n}^h, R_{m',n}^h)$ have the same joint distribution as claimed.

To prove property (iv), we need to first fix the notation. Fix $h \in B(A)$. We will show $P[|D - R| > \varepsilon] < \varepsilon$; the proof of $P[|D' - R'| > \varepsilon] < \varepsilon$ is identical.

For all $x \in \mathbb{R}^d, \beta > 0$, and $\tau > 0$, let $B(x, \tau)$ denote a ball such that

$$N_{\phi}^\beta(\mathcal{P}_\tau \cap B(x, \tau) \cup x) - N_{\phi}^\beta(\mathcal{P}_\tau \cap B(x, \tau)) = \Delta_{\phi}(\frac{\beta}{\tau}).$$

27
For all \( x \in A \) define

\[
D^g_n(x) = [N_{n,\phi}^g(P_{nf}(x) \cap n^{-1/d}B(x, f(x)) \cup x) - N_{n,\phi}^g(P_{nf}(x) \cap n^{-1/d}B(x, f(x)))] \cdot h(x)
\]

and

\[
D(x) = [N_{\phi}^g(P_{f(x)} \cap B(x, f(x)) \cup x) - N_{\phi}^g(P_{f(x)} \cap B(x, f(x)))] \cdot h(x) \overset{D}{=} h(x) \Delta'(g(x)/f(x)).
\]

By following Lemma 6.1 of [8] we can show the following approximations:

(i) the set of points of \( W_1, W_2, ..., W_m \) on \( n^{-1/d}B(x, f(x)) \) differs little from the set of points \( P := P_n^f \) on \( n^{-1/d}B(x, f(x)) \), and

(ii) the set of points \( P := P_n^f \) on \( n^{-1/d}B(x, f(x)) \) differs little from the set of points of \( P_{nf}(x) \) on \( n^{-1/d}B(x, f(x)) \).

Thus, combining (i) and (ii), one might expect that \(|R(x, x') - D^g_n(x)|\) is small. Indeed, by following Lemma 6.1 of [8] we can show for all \( \varepsilon > 0 \) that there exists \( \delta > 0, n_0 \geq 1 \) such that for all \( n \geq n_0 \) and all \( m, m' \in [(1 - \delta)n, (1 + \delta)n] \) with \( m < m' \) and all pairs \((x,x') \in A \times A\) that

\[
P[|R(x, x') - D^g_n(x)| > \varepsilon/2] < \varepsilon/2.
\]

To complete the proof that \( P[|D - R| > \varepsilon] < \varepsilon \), it remains to show for all \( x \in A \) that

\[
P[|D^g_n(x) - D(x)| > \varepsilon/2] < \varepsilon/2. \quad (3.27)
\]

The difference \( D(x) - D^g_n(x) \) can be estimated by considering differences of the type

\[
\sum_{X_i \in B(x, f(x)) \cap P_{f(x)}} \phi \left( \int_{C(X_i, B(x, f(x)) \cap P_{f(x)})} g(x)du - \int_{C(X_i, n^{-1/d}B(x, f(x)) \cap P_{nf}(x)}) g(u)du \right)
\]

\[
\sum_{X_i \in n^{-1/d}B(x, f(x)) \cap P_{nf}(x)} \phi \left( \int_{C(X_i, n^{-1/d}B(x, f(x)) \cap P_{nf}(x))} g(x)du - \int_{C(X_i, P_{nf}(x))} g(u)du \right).
\]

Using the continuity of \( \phi \) and \( g \) as well as the boundedness of \( f \) and \( g \), standard arguments (see (3.24) - (3.25)) imply that the summands are uniformly small in probability for large \( n \). Since the number of summands is a.s. finite, the above difference is small in probability, showing (3.27). Thus condition (iv) holds and Lemma 3.4 is proved.

By following section 6 of [8], and by combining Lemma 3.4 and the limit (3.2) we may establish the variance limit (2.5) in Theorem 2.1.
3.3 Completion of proof of Theorem 2.1

To prove the convergence in law (2.6) in Theorem 2.1, it suffices to show that the higher order cumulants for the random variables measures \( \langle h, \mu_{g, \varphi} \rangle, n \geq 1 \), converge to zero. If the moments \( \mathbb{E}[\phi^p(\alpha \Gamma_1)] \) were finite for all \( \alpha > 0 \) and all \( p > 0 \), then the result would follow immediately from the cumulant methods from section five of [8].

Without this strong moment condition, we will instead show that (2.6) follows immediately from Theorem 1.2 of [36]. Define for all \( x \in \lambda^{1/d}A \) and all finite point sets \( \mathcal{X} \subset \lambda^{1/d}A \) the weight

\[
U(x; \mathcal{X}) = \phi \left( \int_{C(x; \mathcal{X})} g(u \lambda^{-1/d}) du \right).
\]

Then \( U \) is exponentially stabilizing for the density \( f \) in the sense that for all \( \lambda > 0 \) and all \( x \in \lambda^{1/d}A \), there exists a random variable \( R := R(x, \lambda) \) such that

\[
\sup_{\lambda > 0, x \in \lambda^{1/d}A} P[R(x, \lambda) > t] \leq C \exp(-t/C).
\]

Set \( U_\lambda(x; \mathcal{X}) := U(\lambda^{1/d}x; \lambda^{1/d}A) \) and note that by hypothesis, \( \sup_{\lambda > 0, x \in \mathbb{R}^d} \mathbb{E}[|U_\lambda(x; P^t \cup x)|^p] < \infty \) for some \( p > 3 \), where here \( X \) has density \( f \). It follows that

\[
\mu_{\lambda, \varphi}^g = \sum_{x \in P^t \lambda} U_\lambda(x, P^t \lambda) \delta_x
\]

and thus by Theorem 1.2 of [36], for all \( h \in B(A) \)

\[
\sup_t \left| P \left[ \frac{\langle h, \mu_{\lambda, \varphi}^g \rangle}{\text{Var}[\langle h, \mu_{\lambda, \varphi}^g \rangle]} \leq t \right] - P[N(0, 1) \leq t] \right| \leq C(\log \lambda)^{3d} \lambda^{-1/2}.
\]

Thus, combining with Lemma 3.3, we obtain that \( \langle h, \mu_{\lambda, \varphi}^g \rangle \) converges to a normal random variable with mean zero and variance

\[
\int_A (h(x))^2 V_\varphi \frac{g(x)}{f(x)} f(x) dx,
\]

proving Theorem 1.2.

4 Proof of Proposition 2.3

For all \( x \in \mathbb{R}^d \), let \( B_x := B_x(P) \) denote the volume of the largest ball around the point \( x \) not containing any points from \( P \).
The last display in the proof of Lemma 3.3 (with \( f = 1 \) and \( g = \beta \)) and the translation invariance of \( N^\beta_\phi \), which yields for all \( x \in \mathbb{R}^d \), \( q^x(0) = q(0) = \mathbb{E}[\phi^2(\beta \Gamma_1)] \) and \( c^x(x, y) = c(0, y) \), implies that
\[
V_\phi(\beta) = \lim_{\lambda \to \infty} \frac{\text{Var}[N^\beta_\phi(\mathbb{P}_\lambda \cap [0, 1]^d)]}{\lambda} = \mathbb{E}[\phi^2(\beta \Gamma_1)] + \int_{\mathbb{R}^d} c(0, y) dy,
\]
where \( c(0, y) = \mathbb{E}[\phi(\beta B_0)\phi(\beta B_y)] - \mathbb{E}[\phi(\beta B_0)]\mathbb{E}[\phi(\beta B_y)] \). Since \( B_0 \overset{D}{=} B_y \overset{D}{=} \Gamma_1 \) for all \( y \in \mathbb{R}^d \), it follows that
\[
\mathbb{E}[\phi(\beta B_0)] = \int_0^\infty \phi(\beta s) e^{-s} ds = \mathbb{E}[\phi(\beta B_y)].
\]
For all \( s, t \in \mathbb{R}^+ \), let \( p(s, t) = P[B_0 > s, B_y > t] \). Then for all \( s, t \in [0, |y|^d \omega_d] \) we have
\[
p(s, t) = e^{-(s+t) + I(s, t, |y|)}.
\]
Otherwise \( p(s, t) = 0 \). Hence, for \( y \in \mathbb{R}^d \)
\[
c(0, y) = \int_0^\infty \int_0^\infty \phi(\beta s)\phi(\beta t) \frac{\partial^2 p}{\partial s \partial t} dsdt - \left( \int_0^\infty \phi(\beta s) e^{-s} ds \right)^2
\]
\[
= \beta^2 \int_0^\infty \int_0^\infty \phi'(\beta s)\phi'(\beta t)[p(s, t) - e^{-(s+t)}] dsdt,
\]
using integration by parts on both integrals and \( \phi(0) = 0 \).

By definition of \( p(s, t) \), this yields
\[
c(0, y) = \beta^2 \int_0^{|y|^d \omega_d} \int_0^{|y|^d \omega_d} \phi'(\beta s)\phi'(\beta t)[e^{-(s+t) + I(s, t, |y|)} - e^{-(s+t)}] dsdt -
\]
\[
- \beta^2 \int \int_{\text{max}(s, t) \geq |y|^d \omega_d} \phi'(\beta s)\phi'(\beta t)e^{-(s+t)} dsdt.
\]
Therefore
\[
\int_{\mathbb{R}^d} c(0, y) dy = \beta^2 \int_{\mathbb{R}^d} \int_{0}^{|y|^d \omega_d} \int_{0}^{|y|^d \omega_d} \phi'(\beta s)\phi'(\beta t)[e^{-(s+t) + I(s, t, |y|)} - e^{-(s+t)}] dsdt dy -
\]
\[
- \beta^2 \int_{\mathbb{R}^d} \int_{\text{max}(s, t) \geq |y|^d \omega_d} \phi'(\beta s)\phi'(\beta t)e^{-(s+t)} dsdt dy
\]
and making a change of variable on the outside integral yields
\[
\int_{\mathbb{R}^d} c(0, y) dy = \beta^2 \int_0^\infty \int_0^{|y|^d \omega_d} \phi'(\beta s)\phi'(\beta t)[e^{-(s+t) + I(s, t, |y|)} - e^{-(s+t)}] dsdt dy -
\]
\[
- \beta^2 \int_0^\infty \int_{\text{max}(s, t) \geq |y|^d \omega_d} \phi'(\beta s)\phi'(\beta t)e^{-(s+t)} dsdt dy -
\]
\[
\text{and letting } u = |y|^d \omega_d, \text{ the above becomes}
\]
\[
\int_0^\infty c(0, y) dy = \beta^2 \int_0^\infty \int_0^u \phi'(\beta s)\phi'(\beta t)e^{-(s+t)} [e^{I(s, t, u/\omega_d)^{1/d}} - 1] dsdt du -
\]

30
\[-\beta^2 \int_0^\infty \int_{\max(s,t) \geq u} \phi'(\beta s)\phi'(\beta t)e^{-(s+t)}dsdtdu.\]

Finally, change the order of integration to obtain
\[
\int_0^\infty c(0, y)dy = \beta^2 \int_0^\infty \int_0^\infty \phi'(\beta s)\phi'(\beta t)e^{-(s+t)}\int_{\max(s,t)}^\infty \frac{[e^{I(s,t,(u/\omega d)}^{1/d} - 1]}{dudsdt} - \beta^2 \int_0^\infty \int_0^\infty \phi'(\beta s)\phi'(\beta t)e^{-(s+t)}\int_0^{\max(s,t)} \max(s,t) \int e^{I(s,t, (u/\omega d)}^{1/d} - 1]dudsdt,
\]

which is exactly the desired limit. This proves Proposition 2.3.

## 5 Proof of Theorem 2.5

Throughout let \( P \) denote a rate one Poisson point process on \( \mathbb{R} \) and let \( 0 < S_1 < S_2 < ... \) denote the order statistics for the restriction of \( P \) to \( \mathbb{R}^+ \). Let \( S_k = \{S_1, ..., S_k\} \), and put \( S_0 = 0 \). Note that for all \( k \geq 1 \), \( S_k \) has a \( \Gamma_k \) distribution.

**Lemma 5.1** Let \( f : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R} \) be measurable. Then
\[
\int_0^\infty \mathbb{E} P f(S_m, y) 1_{y < S_m} dy = \sum_{l=1}^m \mathbb{E} P f(S_{m+1,l}, S_l)
\]
where \( S_{m+1,l} := \{S_1, ..., S_{l-1}, S_{l+1}, ..., S_{m+1}\} \).

**Proof.** Indeed,
\[
\mathbb{E} P f(S_m, y) = \int_{0 < s_1 < ... < s_m} e^{-s_m} f(s_1, ..., s_m, y) ds_1 ... ds_m.
\]

Hence,
\[
\int_0^\infty \mathbb{E} P f(S_m, y) 1_{y < S_m} ds_1 ... ds_m dy = \int_{0 < s_1 < ... < s_m} e^{-s_m} f(s_1, ..., s_m, y) 1_{y < s_m} ds_1 ... ds_m dy = \sum_{l=1}^m \int_{0 < s_1 < ... < s_l-1 < y < s_l < ... < s_m} e^{-s_m} f(s_1, ..., s_m, y) 1_{y < s_m} ds_1 ... ds_m dy,
\]

which is exactly the desired identity.

From the proof of Theorem 2.1 we know that for all \( \beta > 0 \),
\[
\lim_{\lambda \to \infty} \frac{\text{Var}[S_{\phi,k}(P_{\lambda \cap [0,1]})]}{\lambda} = \mathbb{E} [\phi^2(\beta k)] + \int_{-\infty}^\infty c(0, y) dy
\]
where
\[
c(0, y) = \mathbb{E} [\phi(\beta C_0)\phi(\beta C_y)] - \mathbb{E} [\phi(\beta C_0)] \cdot \mathbb{E} [\phi(\beta C_y)],
\]
where $C_0$ (respectively, $C_y$) denotes the length of the $k$-spacings starting at the origin (respectively, starting at $y$) with respect to the augmented point set $P \cup 0 \cup y$. Write $e_k := \mathbb{E}[\phi(\beta S_k)]$. Now

$$c(0, y) = \mathbb{E}[(\phi(\beta C_0)\phi(\beta C_y) - \phi(\beta S_k)e_k)(1(y < S_k) + 1(y > S_k))].$$

The expectation over the latter term vanishes since

$$\mathbb{E}[(\phi(\beta C_0)\phi(\beta C_y) - \phi(\beta S_k)e_k)(1(y > S_k))\mathbb{E}(\phi(\beta C_y) - e_k|\mathcal{F}_k)] = 0,$$

as the restriction of $P$ to $[y, \infty)$ is independent of $\mathcal{F}_k$.

Therefore for $y > 0$

$$\int_0^\infty c(0, y)dy = \int_0^\infty \mathbb{E}[\phi(\beta C_0)\phi(\beta C_y) - \phi(\beta S_k)e_k](1(y < S_k))dy.$$

By Lemma 5.1, the above equals

$$\sum_{l=1}^k \mathbb{E}[\phi(\beta S_k)\phi(\beta S_{k+l} - \beta S_l) - \phi(\beta S_{k+1})e_k]$$

which clearly equals

$$\sum_{l=1}^k \mathbb{E}[\phi(\beta S_k)\phi(\beta S_{k+l} - \beta S_l) - \phi(\beta S_{k+1})e_k + (\phi(\beta S_k) - \phi(\beta S_{k+1})e_k]$$

$$= \sum_{l=1}^k \text{Cov}(\phi(\beta S_k), \phi(\beta S_{k+l} - \beta S_l) + k(e_k - e_{k+1})e_k$$

$$= \sum_{l=1}^k \text{Cov}(\phi(\beta S_k), \phi(\beta S_{k+l} - \beta S_l) + k(e_k - e_{k+1})e_k.$$

By symmetry, $\int_{-\infty}^0 c(0, y)dy = \int_0^\infty c(0, y)dy$ and thus

$$\lim_{\lambda \to \infty} \frac{\text{Var}[S_{\phi,k}^d(\mathcal{P} \cap [0, 1])]}{\lambda} = V_{\phi,k}(\beta)$$

where $V_{\phi,k}(\beta)$ is defined in (2.13). To prove Theorem 2.5, it is enough to combine Theorem 2.1, the above limit, and the following lemma, whose proof we leave to the reader.

**Lemma 5.2** For all $k = 1, 2, ..., \phi \in \Phi_k$, and $\beta > 0$ we have

$$\mathbb{E}
\left[
\lim_{R \to \infty} S_{\phi,k}^d(\mathcal{P} \cap B_R(0) \cup 0) - S_{\phi,k}^d(\mathcal{P} \cap B_R(0))
\right]
= (k + 1)\mathbb{E}[\phi(\beta \Gamma_k) - k\mathbb{E}(\phi(\beta \Gamma_{k+1})].$$

This completes the proof of Theorem 2.5.

\[\square\]
6 Proof of Corollary 2.7

Recall that $\xi^t(x, \mathcal{X}) := \text{card}(B_t(x) \cap \mathcal{X})$. Since the spatial correlations for $\xi^t$ are of finite range, we may easily determine the asymptotic distributions by appealing to the general results of [8].

We find the two point correlation function $c(0, y)$ for $\xi^t$ defined over Poisson points $P_\beta$ with constant intensity $\beta$ as follows. Now

$$c(0, y) = \mathbb{E}[|B_t(0) \cap P_\beta \cup 0 \cup \{y]\| - |B_t(0) \cap P_\beta \cup 0 \cup \{y]\|] - \mathbb{E}[|B_t(0) \cap P_\beta|] \cdot \mathbb{E}[|B_t(0) \cap P_\beta|],$$

and writing $B_t(0) \cap P_\beta$ as the union of disjoint sets $B_t(0) \cap B_t^\beta(y) \cap P_\beta$ and $B_t(0) \cap B_t(y) \cap P_\beta$ and similarly for $B_t(0) \cap P_\beta$, it follows that

$$c(0, y) = \text{Var}[|B_t(0) \cap B_t(y) \cap P_\beta|] = \beta \cdot |B_t(0) \cap B_t(y)|.$$  

Letting $v_t := t^d \omega_d$ denote the volume of a ball of radius $t$, we thus obtain

$$\int_{\mathbb{R}^d} c(0, y) dy = \beta \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} 1_{B_t(0)}(u) 1_{B_t(y)}(u) du dy$$

$$= \beta \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} 1_{B_t(0)}(u) 1_{B_t(y)}(u) dy du = \beta \int_{\mathbb{R}^d} 1_{B_t(0)}(u) \int_{\mathbb{R}^d} 1_{B_t(y)}(u) dy du = \beta v_t^2.$$

Since $\mathbb{E}[|B_t(0) \cap P_\beta|^2] = \beta v_t + (\beta v_t)^2$, it follows directly from the proof of Lemma 3.3 of [8] that

$$\beta \nabla_t(\beta) := \lim_{\lambda \to \infty} \frac{\text{Var}[H_t^\lambda(P_\beta \cap [0, 1]^d)]}{\lambda} = \beta \mathbb{E}[|B_t(0) \cap P_\beta|^2] + \beta^2 \int_{\mathbb{R}^d} c(0, y) dy$$

$$= \beta^2 (v_t + 2 \beta v_t^2),$$

so that $\nabla_t(\beta) = \beta(v_t + 2 \beta v_t^2)$. Letting $Po(\tau)$ denote a Poisson random variable with parameter $\tau$, the expected add-one cost of inserting $0$ into $P_\beta$ is

$$\Delta_t(\beta) := \mathbb{E} \left[ \lim_{R \to \infty} [H^t(P_\beta \cap B_R(0) \cup 0) - H^t(P_\beta \cap B_R(0))] \right] = \mathbb{E}[Po(\beta v_t)] = \beta v_t,$$

since the difference in the inside braces is the number of points in $P_\beta$ within $t$ of the origin. It follows that

$$\lim_{n \to \infty} \frac{\text{Var}[H_t^\lambda(X_1, ..., X_n)]}{n} = \int_A \nabla_t(f(x)) f(x) dx - (\int_A \Delta_t(f(x)) f(x) dx)^2$$

$$= v_t \int_A f(x)^2 dx + v_t^2 \left(2 \int_A f(x)^3 dx - (\int_A f(x)^2 dx)^2\right),$$

which is the desired limit (2.29). The remainder of Corollary 2.7 follows from Theorem 2.5. 

**Acknowledgements:** Yu. Baryshnikov thanks the Lehigh Mathematics department for hosting his research visit. J. E. Yukich gratefully thanks R. Jimenez for multiple inspiring discussions related to spacings.
References


34


