Asymptotic moments of near neighbour distance distributions

By Dafydd Evans[1], Antonia J. Jones[1] and Wolfgang M. Schmidt[2]

[1]*Department of Computer Science, Cardiff University, PO Box 916, Cardiff CF24 3XF, UK.* [2]*Department of Mathematics, University of Colorado, Boulder, CO 80309-0395, USA.*

Let C be a compact convex body in $\mathbb{R}^m$ and consider a set of points selected at random from C according to some well behaved sampling distribution. We obtain an asymptotic expression for the positive moments of the kth near neighbour distance distribution as the number of points increases to infinity.

Keywords: Near neighbour distance distributions, moments, point processes

1. Introduction

Let $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ be a set of M points selected uniformly at random from the unit hypercube in $\mathbb{R}^m$ and let $d_{M,k}$ denote the distance between any point of the set and its kth nearest neighbour in the set. In Percus & Martin (1998) it is shown that under periodic boundary conditions, the expected value of $d_{M,k}$ satisfies the asymptotic expression

$$\mathcal{E}(d_{M,k}) = V_m^{-1/m} \frac{\Gamma(k+1/m)}{\Gamma(k)} \frac{1}{M^{1/m}} + O\left(\frac{1}{M^{1+1/m}}\right) \quad \text{as} \quad M \to \infty \tag{1.1}$$

where V_m denotes the volume of the unit ball in $\mathbb{R}^m$. In this paper we prove a similar result for the αth non–negative moment $\mathcal{E}(d^{\alpha}_{M,k})$ of the kth nearest neighbour distance distribution with the following important generalisations:

- The points $\boldsymbol{x}_i$ may be selected from any compact convex body C in $\mathbb{R}^m$.

- The points $\boldsymbol{x}_i$ may be selected according to any well behaved sampling density ϕ on C.

We show that for all $0 < \rho < 1/m$,

$$\mathcal{E}(d^{\alpha}_{M,k}) = \frac{c(m,\alpha,k,\phi)}{M^{\alpha/m}} + O\left(\frac{1}{M^{(\alpha+1)/m-\rho}}\right) \quad \text{as} \quad M \to \infty \tag{1.2}$$

where

$$c(m,\alpha,k,\phi) = V_m^{-\alpha/m} \frac{\Gamma(k+\alpha/m)}{\Gamma(k)} \int_C \phi(\boldsymbol{x})^{1-\alpha/m} \, d\boldsymbol{x} \tag{1.3}$$

is a finite constant not depending on M.

In what follows, μ represents Lebesgue measure in $\mathbb{R}^m$ and ∂A denotes the boundary of a set $A \subset \mathbb{R}^m$. The ball of radius r centred at $\boldsymbol{x} \in \mathbb{R}^m$ is denoted by $B_x(r)$ and V_m denotes the Lebesgue measure of the unit ball in $\mathbb{R}^m$. Finally, for any $A \subset \mathbb{R}^m$ and $\delta > 0$, the *boundary region of width* δ is defined to be the set $A(\delta)$ consisting of those points in A which are within (Euclidean) distance δ of the boundary,

$$A(\delta) = \{\boldsymbol{x} \in A : \inf_{\boldsymbol{y} \in \partial A} |\boldsymbol{x} - \boldsymbol{y}| < \delta\} \tag{1.4}$$

2. The set C.

Let C be any compact subset of $\mathbb{R}^m$ having diameter c_1 $(0 < c_1 < \infty)$ and without loss of generality suppose that $\mu(C) = 1$. We first impose the following geometric conditions on C.

C.1 There exists some constant $c_2 > 0$ such that for all $\boldsymbol{x} \in C$ and $0 < r < c_1$,

$$\mu(B_x(r) \cap C) > c_2 r^m \tag{2.1}$$

i.e. at least a uniformly constant proportion of the ball $B_x(r)$ must intersect C.

C.2 There exist constants $\lambda = \lambda(C) > 0$ and $c_3 > 0$ such that for all $0 < \delta < \lambda$,

$$\mu(C(\delta)) \leq c_3 \delta \tag{2.2}$$

i.e. the measure of the boundary region must be uniformly bounded above by some constant multiple of δ.

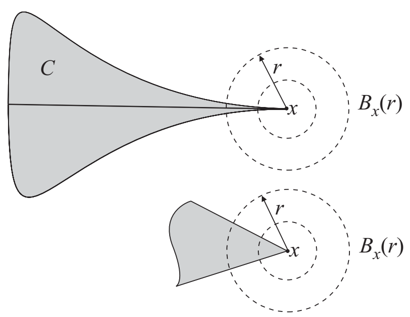

Figure 1. Condition **C.1** eliminates certain types of boundary points (top)

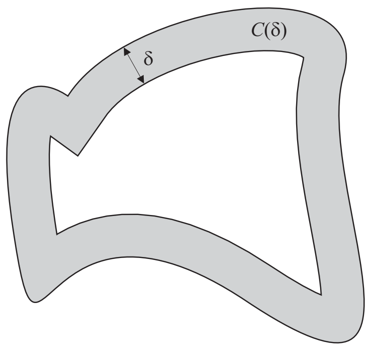

Figure 2. Condition **C.2** requires that the measure of $C(\delta)$ is bounded above by some constant multiple of δ.

Proposition 2.1. *Conditions* **C.1** *and* **C.2** *are satisfied by compact convex bodies in* $\mathbb{R}^m$.

Proof. Let C be a compact convex body in $\mathbb{R}^m$. To prove condition **C.1** we note that by definition, a convex body has non–empty interior so there exist points $\boldsymbol{a}, \boldsymbol{b} \in C$ along with some $r_a, r_b > 0$ such that the balls $B_a(r_a)$ and $B_b(r_b)$ are disjoint and completely contained in C.

First suppose that $\boldsymbol{x} \in C \setminus B_a(r_a)$ and consider the cone C_x having vertex $\boldsymbol{x}$ and base equal to the intersection of $B_a(r_a)$ with the hyperplane through $\boldsymbol{a}$ perpendicular to the line joining $\boldsymbol{x}$ and $\boldsymbol{a}$. By convexity, C_x is completely contained in C and since $r_a > 0$ we have that C_x is of positive volume for each $\boldsymbol{x} \in C \setminus B_a(r_a)$.

Let $c_x(r) > 0$ denote the proportion of the ball $B_x(r)$ occupied by the cone C_x. As r increases from 0 to c_1, the proportion $c_x(r)$ remains constant while $r \leq |\boldsymbol{x} - \boldsymbol{a}|$ and then decreases monotonically for $r > |\boldsymbol{x} - \boldsymbol{a}|$. Let $c_x = c_x(c_1)$ denote the minimum value of $c_x(r)$. Since the volume of C_x is positive and the volume of the ball $B_x(c_1)$ is finite, it is clear that $c_x > 0$ and furthermore that $\mu(B_x(r) \cap C) \geq c_x \mu(B_x(r)) > 0$ for all $0 < r \leq c_1$.

Define c_a to be the minimum value of c_x over all points $\boldsymbol{x} \in C \setminus B_a(r_a)$. This corresponds to those points $\boldsymbol{x}$ which lie on the boundary of the ball $B_a(r_a)$ (for which the cones C_x are of minimum volume) and since $B_a(r_a)$ is of positive radius we have that $c_a > 0$. Hence, $\mu(B_x(r) \cap C) \geq c_a \mu(B_x(r)) > 0$ for all $0 < r \leq c_1$ and $\boldsymbol{x} \in C \setminus B_a(r_a)$.

Now suppose that $\boldsymbol{x} \in B_a(r_a)$ and define C_x and c_x as above, this time relative to the ball $B_b(r_b)$. Define c_b to be the minimum value of c_x over each $\boldsymbol{x} \in B_a(r_a)$. Since $B_a(r_a)$ and $B_b(r_b)$ are disjoint and since $B_b(r_b)$ is of positive radius we have that $c_b > 0$ and hence $\mu(B_x(r) \cap C) \geq c_b \mu(B_x(r)) > 0$ for all $0 < r \leq c_1$ and $\boldsymbol{x} \in B_a(r_a)$.

Finally, letting $c = \min\{c_a, c_b\} > 0$ we have that for all $0 < r \leq c_1$ and for every $\boldsymbol{x} \in C$,

$$\mu(B_x(r) \cap C) \geq c\mu(B_x(r)) \geq cV_m r^m > 0 \tag{2.3}$$

where V_m is the volume of the unit ball in $\mathbb{R}^m$ and hence condition **C.1** is satisfied with $c_2 = cV_m > 0$.

Next we show that C satisfies condition **C.2**. By definition, C has non–empty interior and we may suppose (without loss of generality) that the origin $\mathbf{0} \in \operatorname{int} C$. Let $B_0(\lambda)$ denote the ball of maximal radius $\lambda > 0$ centred at the origin and contained in C.

Let $0 < \delta < \lambda$ and suppose that $\boldsymbol{x} \in C(\delta)$. Let $\boldsymbol{y}$ be a boundary point of C such that the distance from $\boldsymbol{x}$ to $\boldsymbol{y}$ is strictly less than δ (this exists by definition of $C(\delta)$) and define $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z}$ where $|\boldsymbol{z}| < \delta$.

Let us write $-\boldsymbol{z} = (\delta/\lambda)\boldsymbol{a}$ where $\boldsymbol{a} \in \mathbb{R}^m$. Then $\boldsymbol{a} = (\lambda/\delta)(-\boldsymbol{z})$ and $|\boldsymbol{a}| = (\lambda/\delta)|\boldsymbol{z}|$. By definition, $|\boldsymbol{z}| < \delta$ so we have $|\boldsymbol{a}| < \lambda$ and hence $\boldsymbol{a} \in \operatorname{int} B_0(\lambda)$. Since $B_0(\lambda) \subseteq C$ this implies that $\boldsymbol{a} \in \operatorname{int} C$ and therefore $-\boldsymbol{z} \in \operatorname{int}(\delta/\lambda)C$.

Now suppose that $\boldsymbol{x} \in (1 - \delta/\lambda)C$. Then $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{z}$ may be expressed as $\boldsymbol{y} = (1 - \delta/\lambda)\boldsymbol{b} + (\delta/\lambda)\boldsymbol{a}$ for some $\boldsymbol{b} \in C$ and $\boldsymbol{a} \in \operatorname{int} C$. Since $0 < \delta < \lambda$ we have $0 < \delta/\lambda < 1$ and $0 < 1 - \delta/\lambda < 1$ and since C is convex and $\boldsymbol{a} \in \operatorname{int} C$ this implies that $\boldsymbol{y} \in \operatorname{int} C$, contradicting the fact that $\boldsymbol{y}$ is a boundary point of C.

Hence $\boldsymbol{x} \notin (1-\delta/\lambda)C$ so the sets $C(\delta)$ and $(1-\delta/\lambda)C$ are disjoint. $C(\delta)$ is therefore contained in $C \setminus (1-\delta/\lambda)C$ so

$$\mu\left(C(\delta)\right) \leq \mu(C) - \mu\left((1-\delta/\lambda)C\right) \tag{2.4}$$

Since $\mu((1-\delta/\lambda)C) = (1-\delta/\lambda)^m \mu(C)$ and $\mu(C) = 1$ we have

$$\mu(C(\delta)) \leq 1 - (1-\delta/\lambda)^m \tag{2.5}$$

For all $0 < \delta < \lambda$ we may therefore conclude that $\mu(C(\delta)) \leq c_3\delta$ for some constant $c_3 > 0$, as required. □

3. The sampling distribution Φ.

Suppose that points $\boldsymbol{x}$ are selected at random from C according to a sampling distribution Φ. We restrict our attention to those distributions Φ for which the corresponding density function ϕ satisfies the following conditions.

P.1 ϕ is continuous on C.

P.2 ϕ has bounded partial derivatives at each point of C.

P.3 $\phi(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in C$.

Since C is compact and since ϕ is continuous and strictly positive on C, it is easily shown that there exist constants a_1, a_2 such that

$$0 < a_1 < \phi(\boldsymbol{x}) < a_2 < \infty \quad \text{for all} \quad \boldsymbol{x} \in C \tag{3.1}$$

Let $\omega_x(r)$ denote the probability measure induced by the sampling distribution Φ on the spherical neighbourhoods of C,

$$\omega_x(r) = \mathbf{P}(B_x(r)) = \int_{B_x(r)\cap C} \phi(\boldsymbol{t})\, d\boldsymbol{t} \tag{3.2}$$

Then $\omega_x(r)$ is the probability that a point selected from C is selected from the ball $B_x(r)$. We remark that for all $\boldsymbol{x} \in C$ and all $0 \leq r \leq c_1$, by (3.1) and condition **C.1**,

$$\omega_x(r) \geq a_1\mu(B_x(r)\cap C) \geq a_1 c_2 r^m \tag{3.3}$$

and we say that $\omega_x(r)$ satisfies a *positive density* condition on C.

4. The nearest neighbour distance distribution

Suppose that M points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ are selected at random from C according to the sampling distribution Φ. For any fixed point $\boldsymbol{x} \in C$ the *kth nearest neighbour* of $\boldsymbol{x}$ in the set $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ is defined to be that point $\boldsymbol{x}_j$ with the property that exactly $k-1$ points of the set are closer to $\boldsymbol{x}$ than $\boldsymbol{x}_j$ is to $\boldsymbol{x}$. We note that since $\mu(C) = 1$ and $\phi(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in C$, the kth nearest neighbour of $\boldsymbol{x}$ is uniquely defined with probability 1.

Let $d_{M,k}(\boldsymbol{x})$ denote the distance from $\boldsymbol{x}$ to its kth nearest neighbour in the set $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$. Then $d_{M,k}(\boldsymbol{x})$ is a random variable taking values in the range $[0, c_1]$ and defined on the space of all samples of size M selected from C. Its distribution function is defined by

$$q_x(r) = \mathbf{P}(d_{M,k}(\boldsymbol{x}) \leq r) = q_x(M, k, r) \tag{4.1}$$

Following Percus & Martin (1998) we derive the corresponding density function as follows.

Lemma 4.1. *For fixed $\boldsymbol{x} \in C$, the probability density function of the random variable $d_{M,k}(\boldsymbol{x})$ is given by*

$$dq_x(r) = k\binom{M}{k}\omega_x(r)^{k-1}(1-\omega_x(r))^{M-k}\, d\omega_x(r) \tag{4.2}$$

Proof. Let $\epsilon > 0$ and consider

$$q_x(r+\epsilon) - q_x(r) = \mathbf{P}(r \leq d_{M,k}(\boldsymbol{x}) \leq r+\epsilon) \tag{4.3}$$

Since ϕ is continuous on C, for ϵ sufficiently small we may suppose that the kth nearest neighbour of $\boldsymbol{x}$ is the only point lying in the spherical shell of radius r and width $\epsilon > 0$ centred at $\boldsymbol{x}$. In this case we must have

- $k-1$ points in the ball $B_x(r)$, each selected with probability $\omega_x(r)$.
- Exactly one of the remaining $M-k+1$ points in the shell $B_x(r+\epsilon) \setminus B_x(r)$, selected with probability $\omega_x(r+\epsilon) - \omega_x(r)$.

- The remaining $M-k$ points in the region $C \setminus B_x(r+\epsilon)$, each selected with probability $1-\omega_x(r+\epsilon)$

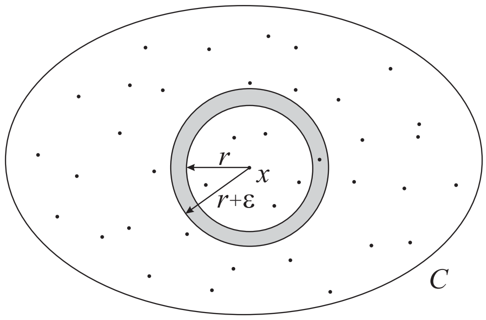

Figure 3. Exactly one point falls in the shaded region $B_x(r+\epsilon) \setminus B_x(r)$.

Using elementary combinatorial arguments we have

$$q(\boldsymbol{x}, r+\epsilon) - q(\boldsymbol{x}, r) = k\binom{M}{k}\omega_x(r)^{k-1}(1-\omega_x(r+\epsilon))^{M-k}(\omega_x(r+\epsilon)-\omega_x(r)) \quad (4.4)$$

and letting $\epsilon \to 0$ we obtain (4.2) as required. □

5. Moments of the near neighbour distance distribution

Consider the αth moment about zero of the kth nearest neighbour distance $d_{M,k}(\boldsymbol{x})$, defined by

$$\mathcal{E}(d_{M,k}^{\alpha}(\boldsymbol{x})) = \int_0^{c_1} r^{\alpha}\, dq_x(r) \quad (5.1)$$

By Lemma 4.1 we can write this as

$$\mathcal{E}(d_{M,k}^{\alpha}(\boldsymbol{x})) = k\binom{M}{k}\int_{r=0}^{c_1} r^{\alpha}\omega_x(r)^{k-1}(1-\omega_x(r))^{M-k} d\omega_x(r) \quad (5.2)$$

Since $\omega_x(r)$ satisfies a positive density condition on C, and since C is convex, we see that $\omega_x(r)$ is strictly monotonic increasing for $0 \le r \le r_0$ for some $r_0 > 0$ and $\omega_x(r) = 1$ for $r_0 \le r \le c_1$. Writing $r = h(\omega)$ for the inverse function where it exists, we change the variable of integration in (5.2) to obtain

$$\mathcal{E}(d_{M,k}^{\alpha}(\boldsymbol{x})) = k\binom{M}{k}\int_0^1 h(\omega_x)^{\alpha}\omega_x^{k-1}(1-\omega_x)^{M-k}\, d\omega_x \quad (5.3)$$

where we have used the fact that $\omega_x(0) = 0$ and $\omega_x(r_0) = 1$.

Thus the expectation $\mathcal{E}(d^{\alpha}_{M,k}(\boldsymbol{x}))$ is defined as an integral over the probability measure $\omega_x(r)$ of the spherical neighbourhoods of $\boldsymbol{x}$. We aim to find an asymptotic expression for (5.3) in terms of the number of points M, as $M \to \infty$. In order to achieve this we need the following technical results, which we state without proof.

Lemma 5.1. *For any fixed* $\sigma > 0$,

$$\frac{\Gamma(N)}{\Gamma(N+\sigma)} = \frac{1}{N^{\sigma}}\left(1 + O\left(\frac{1}{N}\right)\right) \quad \textit{as} \quad N \to \infty \tag{5.4}$$

Lemma 5.2 (The exponential convergence lemma). *Let* $c > 0$ *and* $0 < \sigma < 1$ *be constants. Then for every* $\beta > 0$

$$\left(1 - \frac{c}{N^{\sigma}}\right)^{N} = O\left(\frac{1}{N^{\beta}}\right) \quad \textit{as} \quad N \to \infty \tag{5.5}$$

Using these results, we first show that any ball whose radius does not shrink to zero sufficiently rapidly (relative to M) can be ignored in the limit as $M \to \infty$. More precisely, for any $0 < \rho < 1/m$ we define

$$\delta = \frac{1}{M^{1/m-\rho}} \tag{5.6}$$

and show that any ball of radius $r > \delta$ or equivalently of probability measure $\omega_x > \omega_x(\delta)$ can be neglected in the asymptotic limit as $M \to \infty$.

Lemma 5.3. *For every* $\beta > 0$,

$$k\binom{M}{k}\int_{\omega_x(\delta)}^{1} h(\omega)^{\alpha}\omega^{k-1}(1-\omega)^{M-k}\,d\omega = O\left(\frac{1}{M^{\beta}}\right) \quad \textit{as} \quad M \to \infty \tag{5.7}$$

Proof. Let

$$I(\delta) = k\binom{M}{k}\int_{\omega_x(\delta)}^{1} h(\omega)^{\alpha}\omega^{k-1}(1-\omega)^{M-k}\,d\omega \tag{5.8}$$

By (3.3), $\omega_x(\delta) \geq a_1 c_2 \delta^m$ so for each ω in the range $\omega_x(\delta) \leq \omega \leq 1$ we have that $1 - \omega \leq 1 - a_1 c_2 \delta^m$. Clearly, $h(\omega) \leq c_1$ and $|\omega| \leq 1$ so

$$I(\delta) \leq c_1^{\alpha} k\binom{M}{k}(1 - a_1 c_2 \delta^m)^{M-k} \tag{5.9}$$

By Lemma 5.1,

$$k\binom{M}{k} = \frac{\Gamma(M+1)}{\Gamma(M+1-k)\Gamma(k)} = O((M+1)^k) = O(M^k) \quad \text{as} \quad M \to \infty \tag{5.10}$$

Furthermore, $0 < \rho < 1/m$ implies that $0 < 1 - m\rho < 1$ so $\delta^m = o(1)$ and hence $(1 - a_1 c_2 \delta^m)^{-k} = O(1)$ as $M \to \infty$. Since $c_1 < \infty$ and α is fixed we thus have that

$$I(\delta) \leq O(M^k)(1 - a_1 c_2 \delta^m)^M \quad \text{as} \quad M \to \infty \tag{5.11}$$

Substituting for δ we get

$$I(\delta) \leq O(M^k)\left(1 - \frac{a_1 c_2}{M^{1-m\rho}}\right)^M \quad \text{as} \quad M \to \infty \tag{5.12}$$

and the result follows by Lemma 5.2. □

In evaluating (5.3) we may therefore restrict our attention to those neighbourhood balls of $\boldsymbol{x}$ having probability measure in the range $[0, \omega_x(\delta)]$

(*a*) *Main Theorem*

Theorem 5.4. *Let C be a compact convex body in $\mathbb{R}^m$ with $\mu(C) = 1$ and let $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ be a set of points selected independently at random from C according to the sampling distribution Φ whose density function ϕ satisfies* **P.1** – **P.3**. *Let $d_{M,k}$ denote the distance between any point $\boldsymbol{x}_i$ and its kth nearest neighbour in the set $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$. Then for all $0 < \rho < 1/m$ and integer $\alpha \geq 0$,*

$$\mathcal{E}(d^{\alpha}_{M,k}) = \frac{c(m, \alpha, k, \phi)}{M^{\alpha/m}} + O\left(\frac{1}{M^{(\alpha+1)/m-\rho}}\right) \quad \textit{as} \quad M \to \infty \tag{5.13}$$

where

$$c(m, \alpha, k, \phi) = V_m^{-\alpha/m} \frac{\Gamma(k + \alpha/m)}{\Gamma(k)} \int_C \phi(\boldsymbol{x})^{1-\alpha/m}\, d\boldsymbol{x} \tag{5.14}$$

is a constant not depending on M and V_m is the volume of the unit ball in $\mathbb{R}^m$.

Proof. Fix some point $\boldsymbol{x}$ of the sample $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ and define $d_{M,k}(\boldsymbol{x})$ to be the distance from $\boldsymbol{x}$ to its kth nearest neighbour. By (5.3) applied to a set of $M - 1$ points we have

$$\mathcal{E}(d^{\alpha}_{M,k}(\boldsymbol{x})) = C_{M,k} \int_0^1 h(\omega_x)^{\alpha} \omega_x^{k-1} (1 - \omega_x)^{M-k-1}\, d\omega_x \tag{5.15}$$

where $h(\omega_x)$ is the radius of the ball centred at $\boldsymbol{x}$ of probability measure ω and

$$C_{M,k} = \frac{\Gamma(M)}{\Gamma(M-k)\Gamma(k)} \tag{5.16}$$

The theorem is proved by computing the expected value of (5.15) over all points $\boldsymbol{x}$ selected from C according to Φ, given by

$$\mathcal{E}(d^{\alpha}_{M,k}) = \int_C \mathcal{E}(d^{\alpha}_{M,k}(\boldsymbol{x}))\phi(\boldsymbol{x})\,d\boldsymbol{x} \tag{5.17}$$

Since C is convex it is connected and since ϕ is continuous on C we can apply the first mean value theorem of the integral calculus to ϕ. Hence, by (3.2) there exists a point $\boldsymbol{\xi}_1 \in B_x(r) \cap C$ such that

$$\omega_x(r) = \phi(\boldsymbol{\xi}_1)\mu(B_x(r) \cap C) \tag{5.18}$$

Furthermore, since ϕ is differentiable at every point of C we can also apply the first mean value theorem of the differential calculus to ϕ. Hence there exists a point $\boldsymbol{\xi}_2$ on the line segment joining $\boldsymbol{x}$ and $\boldsymbol{\xi}_1$ such that

$$\phi(\boldsymbol{\xi}_1) = \phi(\boldsymbol{x}) + (\boldsymbol{x} - \boldsymbol{\xi}_1)\phi'(\boldsymbol{\xi}_2) \tag{5.19}$$

and since C is convex, $\boldsymbol{\xi}_2$ is contained in $B_x(r) \cap C$. For $0 < \rho < 1/m$ we define δ as in (5.6) and by Lemma 5.3 we may suppose without loss of generality that $r \in [0, \delta]$. Since $\boldsymbol{\xi}_1 \in B_x(r) \cap C$ we may therefore assume that $|\boldsymbol{x} - \boldsymbol{\xi}_1| \leq \delta$ and write

$$|\boldsymbol{x} - \boldsymbol{\xi}_1| = O(\delta) \quad \text{as} \quad M \to \infty \tag{5.20}$$

By hypothesis all partial derivatives of ϕ are bounded at each point of C and since $\boldsymbol{\xi}_2 \in C$, by (5.19) and (5.20) we have

$$\phi(\boldsymbol{\xi}_1) = \phi(\boldsymbol{x}) + O(\delta) \quad \text{as} \quad M \to \infty \tag{5.21}$$

and hence by (5.18),

$$\omega_x(r) = (\phi(\boldsymbol{x}) + O(\delta))\mu(B_x(r) \cap C) \quad \text{as} \quad M \to \infty \tag{5.22}$$

Let $B = C(\delta)$ denote the boundary region of width δ as defined in (1.4) and let $A = C \setminus B$. We define the conditional expectations $\mathcal{E}_A(\boldsymbol{x}) = \mathcal{E}(d^\alpha_{M,k}(\boldsymbol{x}) \,|\, \boldsymbol{x} \in A)$ and $\mathcal{E}_B(\boldsymbol{x}) = \mathcal{E}(d^\alpha_{M,k}(\boldsymbol{x}) \,|\, \boldsymbol{x} \in B)$ so that

$$\mathcal{E}(d^\alpha_{M,k}) = \int_A \mathcal{E}_A(\boldsymbol{x})\phi(\boldsymbol{x})\, d\boldsymbol{x} + \int_B \mathcal{E}_B(\boldsymbol{x})\phi(\boldsymbol{x})\, d\boldsymbol{x} \tag{5.23}$$

Case (1): $\boldsymbol{x} \in A$. Since $\boldsymbol{x}$ is at least a distance δ from the boundary of C, by Lemma 5.3 we can assume (without loss of generality) that the ball $B_x(r)$ is completely contained in C. Thus $\mu(B_x(r) \cap C) = V_m r^m$ and hence by (5.22)

$$\omega_x = (\phi(\boldsymbol{x}) + O(\delta))V_m r^m \quad \text{as} \quad M \to \infty \tag{5.24}$$

Since $\phi(\boldsymbol{x}) \geq a_1 > 0$ for all $\boldsymbol{x} \in C$ we have

$$r^m = \frac{\omega_x}{V_m\phi(\boldsymbol{x})}\left(1 + \frac{O(\delta)}{\phi(\boldsymbol{x})}\right)^{-1} = \frac{\omega_x}{V_m\phi(\boldsymbol{x})}\,(1 + O(\delta)) \tag{5.25}$$

as $M \to \infty$ so the inverse function $r = h(\omega_x)$ is given by

$$h(\omega_x) = (V_m\phi(\boldsymbol{x}))^{-1/m}\omega_x^{1/m}(1 + O(\delta)) \quad \text{as} \quad M \to \infty \tag{5.26}$$

and since $\alpha \geq 0$ is fixed we have that $(1 + O(\delta))^\alpha = (1 + O(\delta))$ and hence

$$h(\omega_x)^\alpha = (V_m\phi(\boldsymbol{x}))^{-\alpha/m}\omega_x^{\alpha/m}(1 + O(\delta)) \quad \text{as} \quad M \to \infty \tag{5.27}$$

Substituting this into (5.15) we obtain

$$\mathcal{E}_A(\boldsymbol{x}) = (V_m\phi(\boldsymbol{x}))^{-\alpha/m}(1 + O(\delta))I_{M,k} \quad \text{as} \quad M \to \infty \tag{5.28}$$

where

$$I_M(k) = C_{M,k}\int_0^1 \omega_x^{k+\alpha/m-1}(1-\omega_x)^{M-k-1}\, d\omega_x \tag{5.29}$$

The integral in (5.29) is simply the Beta function $B(a, b)$ with parameters $a = k + \alpha/m$ and $b = M - k$, given by

$$B(k + \alpha/m, M - k) = \frac{\Gamma(k + \alpha/m)\Gamma(M - k)}{\Gamma(M + \alpha/m)} \tag{5.30}$$

and substituting for $C_{M,k}$ from (5.16) we get

$$I_M(k) = \frac{\Gamma(k + \alpha/m)}{\Gamma(k)}\frac{\Gamma(M)}{\Gamma(M + \alpha/m)} \tag{5.31}$$

By Lemma 5.1 we obtain

$$I_{M,k} = \frac{\Gamma(k+\alpha/m)}{\Gamma(k)} \frac{1}{M^{\alpha/m}} \left(1 + O\left(\frac{1}{M}\right)\right) \quad \text{as} \quad M \to \infty \tag{5.32}$$

and by definition, $1/M = o(\delta)$ as $M \to \infty$ so that

$$\mathcal{E}_A(\boldsymbol{x}) = (V_m \phi(\boldsymbol{x}))^{-\alpha/m} \frac{\Gamma(k+\alpha/m)}{\Gamma(k)} \frac{1}{M^{\alpha/m}} (1 + O(\delta)) \quad \text{as} \quad M \to \infty \tag{5.33}$$

Hence,

$$\int_A \mathcal{E}_A(\boldsymbol{x}) \phi(\boldsymbol{x})\, d\boldsymbol{x} = \frac{c'(m, \alpha, k, \phi)}{M^{\alpha/m}} (1 + O(\delta)) \quad \text{as} \quad M \to \infty \tag{5.34}$$

where

$$c'(m, \alpha, k, \phi) = V_m^{-\alpha/m} \frac{\Gamma(k+\alpha/m)}{\Gamma(k)} \int_A \phi(\boldsymbol{x})^{1-\alpha/m}\, d\boldsymbol{x} \tag{5.35}$$

In fact the error incurred in replacing the integral in (5.35) by the equivalent integral over C is at most of order $O(\delta)$ as $M \to \infty$. To see this we note that since $0 < a_1 \leq \phi(\boldsymbol{x}) \leq a_2 < \infty$ for each $\boldsymbol{x} \in C$ then $|\phi(\boldsymbol{x})^{1-\alpha/m}| \leq a_3 < \infty$ where $a_3 = \max\{1/a_1^{1-\alpha/m}, a_2^{1-\alpha/m}\}$. Furthermore, by (5.6) we see that $\delta \to 0$ as $M \to \infty$, so by condition **C.2** there exists some constant $c_3 > 0$ such that $\mu(B) \leq c_3 \delta$ for M sufficiently large. Hence

$$\int_B \phi(\boldsymbol{x})^{1-\alpha/m}\, d\boldsymbol{x} \leq a_3 c_3 \delta = O(\delta) \quad \text{as} \quad M \to \infty \tag{5.36}$$

and since $C = A \cup B$ is a disjoint union we obtain

$$\int_A \phi(\boldsymbol{x})^{1-\alpha/m}\, d\boldsymbol{x} = \int_C \phi(\boldsymbol{x})^{1-\alpha/m}\, d\boldsymbol{x} + O(\delta) \quad \text{as} \quad M \to \infty \tag{5.37}$$

Thus, by (5.34) and (5.35) we have that

$$\int_A \mathcal{E}_A(\boldsymbol{x}) \phi(\boldsymbol{x})\, d\boldsymbol{x} = \frac{c(m, \alpha, k, \phi)}{M^{\alpha/m}} + O\left(\frac{\delta}{M^{\alpha/m}}\right) \quad \text{as} \quad M \to \infty \tag{5.38}$$

where

$$c(m, \alpha, k, \phi) = V_m^{-\alpha/m} \frac{\Gamma(k+\alpha/m)}{\Gamma(k)} \int_C \phi(\boldsymbol{x})^{1-\alpha/m}\, d\boldsymbol{x} \tag{5.39}$$

Case (2): $\boldsymbol{x} \in B$. In this case the ball $B_x(r)$ is not necessarily contained in C. Condition **C.1** states that for all $r > 0$ there exists some constant $c_2 > 0$ such that $\mu(B_x(r) \cap C) \geq c_2 r^m$ and hence

$$\omega_x \geq (\phi(\boldsymbol{x}) + O(\delta)) c_2 r^m \tag{5.40}$$

Proceeding as in (5.25) of case (1) we obtain an upper bound for the inverse function $r = h(\omega_x)$ given by

$$h(\omega_x) \leq (c_2 \phi(\boldsymbol{x}))^{-1/m} \omega_x^{1/m} (1 + O(\delta)) \quad \text{as} \quad M \to \infty \tag{5.41}$$

from which we conclude that

$$\mathcal{E}_B(\boldsymbol{x}) = O\left(\frac{1}{M^{\alpha/m}}\right) \quad \text{as} \quad M \to \infty \tag{5.42}$$

and thus

$$\int_B \mathcal{E}_B(\boldsymbol{x}) \phi(\boldsymbol{x})\, d\boldsymbol{x} = O\left(\frac{1}{M^{\alpha/m}}\right) \int_B \phi(\boldsymbol{x})\, d\boldsymbol{x} \quad \text{as} \quad M \to \infty \tag{5.43}$$

Since $|\phi(\boldsymbol{x})| \leq a_2 < \infty$ for each $\boldsymbol{x} \in B$, and by condition **C.2** as before, we have that

$$\int_B \phi(\boldsymbol{x})\, d\boldsymbol{x} \leq a_2 c_3 \delta = O(\delta) \quad \text{as} \quad M \to \infty \tag{5.44}$$

and hence

$$\int_B \mathcal{E}_B(\boldsymbol{x}) \phi(\boldsymbol{x})\, d\boldsymbol{x} = O\left(\frac{\delta}{M^{\alpha/m}}\right) \quad \text{as} \quad M \to \infty \tag{5.45}$$

From (5.23), (5.38) and (5.45) we thus obtain

$$\mathcal{E}(d_{M,k}^{\alpha}) = \frac{c(m, \alpha, k, \phi)}{M^{\alpha/m}} + O\left(\frac{\delta}{M^{\alpha/m}}\right) \quad \text{as} \quad M \to \infty \tag{5.46}$$

and substituting for δ from (5.6) the result follows. □

6. Discussion

(*a*) *A uniform sampling distribution*

Suppose now that the points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ are selected according to a *uniform* sampling distribution. Since $\mu(C) = 1$ in this case we must have that $\phi(\boldsymbol{x}) = 1$ for all $\boldsymbol{x} \in C$ and the constant (5.14) reduces to

$$c(m, \alpha, k, \phi) = V_m^{-\alpha/m} \frac{\Gamma(k + \alpha/m)}{\Gamma(k)} \tag{6.1}$$

which agrees with the constant of (1.1) obtained by Percus & Martin (1998) on a torus in the case $\alpha = 1$. Note also that the convexity condition on C may be relaxed in the case of a uniform sampling distribution, i.e. we need only that conditions **C.1** and **C.2** are satisfied. This is because $\omega_x(r) = \mu(B_x(r) \cap C)$ for each $\boldsymbol{x} \in C$ for the uniform distribution and the application of the mean value theorems (equations (5.18) and (5.19)) in the proof of Theorem 5.4 are unnecessary.

(b) A Law of large numbers for kth nearest neighbour distances

For the points $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_M\}$ let

$$\Delta_M(k) = \frac{1}{M} \sum_{i=1}^{M} d_{M,k}(\boldsymbol{x}_i) \tag{6.2}$$

be the empirical mean distance between kth nearest neighbours in the set. In another paper (Evans D. & Jones A. J., 2002) we show that bounded functions of a point and its kth nearest neighbour satisfy a weak law of large numbers with explicit bounds. Applied to (6.2) this means that for all $\kappa > 0$,

$$\Delta_M(k) = \mathcal{E}(\Delta_M(k)) + O\left(\frac{\mathcal{E}(d_{M,k}^2)^{1/4}}{M^{1/2-\kappa}}\right) \quad \text{in probability as} \quad M \to \infty \tag{6.3}$$

so by Theorem 5.4 it follows that

$$\Delta_M(k) = \mathcal{E}(d_{M,k}) + O\left(\frac{1}{M^{1/2+1/(2m)-\kappa}}\right) \quad \text{in probability as} \quad M \to \infty \tag{6.4}$$

Hence, for all $m \geq 2$ the empirical mean $\Delta_M(k)$ converges in probability to the distribution mean $\mathcal{E}(d_{M,k})$ as $M \to \infty$.

D. Evans gratefully acknowledges the support of EPSRC, U.K. Studentship number 98700188.

The work of W. M. Schmidt was partially supported by NSF-DMS 0074531.

References

Evans D. & Jones A. J. 2002 A proof of the Gamma test. Accepted by *Proc. Roy. Soc. Series A*

Percus A. G. & Martin O. C. 1998 Scaling universalities of kth nearest neighbor distances on closed manifolds. *Adv. Appl. Math.* **21**, 424–436.

Figure captions

Figure 1. Condition **C.1** eliminates certain types of boundary points (top).

Figure 2. Condition **C.2** requires that the measure of $C(\delta)$ is bounded above by some constant multiple of δ.

Figure 3. Exactly one point falls in the shaded region $B_x(r+\epsilon) \setminus B_x(r)$.

Short title for page headings

Near neighbour distances