# The average distance of the $n$-th neighbour in a uniform distribution of random points

P. Bhattacharyya, B. K. Chakrabarti and A. Chakraborti

Saha Institute of Nuclear Physics,

Sector - I, Block - AF, Bidhannagar, Kolkata 700 064, India

December 10, 2001

**Abstract**

We first review the derivation of the exact expression for the average distance $\langle r_n \rangle$ of the $n$-th neighbour of a reference point among a set of $N$ random points distributed uniformly in a unit volume of a $D$-dimensional geometric space. Next we propose a 'mean-field' theory of $\langle r_n \rangle$ and compare it with the exact result. The result of the 'mean-field' theory is found to agree with the exact expression only in the limit $D \to \infty$ and $n \to \infty$. Thus the 'mean-field' approximation is useless in this context.

# 1 Introduction to the average $n$-th neighbour distance

Consider $N$ (a large number) points distributed randomly and uniformly in a unit volume of a $D$-dimensional geometric space. A point is said to be the $n$-th neighbour of another (the reference point) if there are exactly $n-1$ other points that are closer to the latter than the former. The average distance to the first neighbour is exactly known [1]; though originally calculated in three dimensions the method can be used for any finite dimension $D$ : The probability distribution $P(r_1)\mathrm{d}r_1$ of the first neighbour distance is defined by the probability of finding the first neighbour of a given reference point at a distance between $r_1$ and $r_1 + \mathrm{d}r_1$ :

$$P(r_1)\,\mathrm{d}r_1 = [1 - V(r_1)]^{N-1}\ (N-1)\,\mathrm{d}V(r_1), \tag{1}$$

where $V(r_1) = \pi^{D/2} \cdot (r_1)^D / \Gamma(D/2+1)$ is the volume of a $D$-dimensional hypersphere of radius $r_1$ centered at the reference point. The average first neighbour distance is defined as :

$$\langle r_1 \rangle = \int_0^R r_1\ P(r_1)\,\mathrm{d}r_1, \tag{2}$$

where $R$ is the radius of a $D$-dimensional hypersphere of unit volume :

$$R = \frac{1}{\pi^{1/2}}\ \left[\Gamma\left(\frac{D}{2}+1\right)\right]^{1/D}. \tag{3}$$

With the probability distribution of equation 1 we get

$$\begin{aligned}\langle r_1 \rangle &= \int_0^1 r_1\,[1 - V(r_1)]^{N-1}\,(N-1)\,\mathrm{d}V(r_1) \\ &= \frac{1}{\pi^{1/2}}\ \left[\Gamma\left(\frac{D}{2}+1\right)\right]^{1/D} \Gamma\left(1+\frac{1}{D}\right)\ \left(\frac{1}{N}\right)^{1/D}. \end{aligned} \tag{4}$$

Now we address the general problem : What is the form of the average $n$-th neighbour distance, for any finite $n$? Though this is a problem of purely geometric nature, the quantity $\left\langle r_N^{(D)}(n) \right\rangle$ is relevant in physical and computational contexts; for example, in astrophysics we need to know the average distance between neighbouring stars distributed independently in a homogeneous universe [2], and in the traveling salesman problem we need the average distance of the neighbours of each site for estimating the optimal path-length [3].

We proceed by extending the line of argument used in the case of the first neighbour [1] to the $n$-th neighbour. The probability distribution of the $n$-th neighbour distance $r_n$ is defined as the probablity $P(r_n)\mathrm{d}r_n$ of finding the $n$-th neighbour of a given reference point at a distance between $r_n$ and $r_n+\mathrm{d}r_n$. This is a *conditional probability* because we look for the $n$-th neighbour of a point when its first $(n-1)$ neighbours have already been located :

$$P\left(r_n\right)\,\mathrm{d}r_n = \left[1 - \frac{V(r_n) - V(r_{n-1})}{1 - V(r_{n-1})}\right]^{N-n}\,\frac{(N-n)\,\mathrm{d}V(r_n)}{1 - V(r_{n-1})}. \tag{5}$$

The quantity $V(r_n)$ is the volume of a $D$-dimensional hypersphere of radius $r_n$ centered at the reference point. For a given reference point and its first $n-1$ neighbours the average $n$-th neighbour distance is obtained as :

$$\langle r_n\rangle_{\text{(particular)}} = \int_{r_{n-1}}^{R} r_n\,P\left(r_n\right)\,\mathrm{d}r_n \tag{6}$$

where, as before, $R$ is the radius of a $D$-dimensional hypersphere of unit volume. The quantity $\langle r_n\rangle_{\text{(particular)}}$ is a function of a particular $r_{n-1}$, $r_{n-2}$, ..., $r_1$ which are the distances of the first $n-1$ neighbours of the given reference point. To calculate the ensemble average of $r_n$ the quantity $\langle r_n\rangle_{\text{(particular)}}$ must be averaged successively over the probability distributions of each of the first $n-1$ neighbours :

$$\begin{aligned}\langle r_n\rangle \;&=\; \int_0^R \mathrm{d}r_1\,P(r_1)\,\int_{r_1}^R \mathrm{d}r_2\,P(r_2)\,\cdots\int_{r_{n-3}}^R \mathrm{d}r_{n-2}\,P(r_{n-2}) \\ &\quad\times\int_{r_{n-2}}^R \mathrm{d}r_{n-1}\,P(r_{n-1})\,\int_{r_{n-1}}^R \mathrm{d}r_n\,r_n\,P(r_n)\end{aligned} \tag{7}$$

where the probability distribution of the $i$-th neighbour is given by equation 5 with $i$ replacing $n$. After a change in the order of the integrals in equation 7 :

$$
\begin{aligned}
\langle r_n \rangle &= (N-1)(N-2)\cdots(N-n)\frac{\left[\Gamma\left(\frac{D}{2}+1\right)\right]^{1/D}}{\pi^{1/2}} \\
&\times \int_0^1 \mathrm{d}V(r_n)\ [V(r_n)]^{1/D}\ [1-V(r_n)]^{N-n}\int_0^{V(r_n)} \mathrm{d}r_1 \\
&\times \int_{V(r_1)}^{V(r_n)} \mathrm{d}r_2 \cdots \int_{V(r_{n-3})}^{V(r_n)} \mathrm{d}r_{n-2} \int_{V(r_{n-2})}^{V(r_n)} \mathrm{d}r_{n-1} \qquad (8)
\end{aligned}
$$

which gives the final form of the average $n$-th neighbour distance :

$$
\begin{aligned}
\langle r_n \rangle &= \int_0^1 \begin{pmatrix} N-1 \\ n-1 \end{pmatrix} [V(r_n)]^{n+(1/D)-1}\ [1-V(r_n)]^{N-n}\,(N-n)\,\mathrm{d}V(r_n) \\
&= \frac{1}{\pi^{1/2}}\left[\Gamma\left(\frac{D}{2}+1\right)\right]^{1/D}\ \frac{\Gamma\left(n+\frac{1}{D}\right)}{\Gamma(n)}\ \left(\frac{1}{N}\right)^{1/D}. \qquad (9)
\end{aligned}
$$

This result was reported in [4].

Next we consider fluctuations $\delta r_n$ occuring in $r_n$. This can be calculated exactly for any neighbour $n$ ; the mean square deviation in $r_n$ from its average value is given by :

$$
\begin{aligned}
(\delta r_n)^2 &= \left\langle r_n^2 \right\rangle - \langle r_n \rangle^2 \\
&= \frac{1}{\pi}\left[\Gamma\left(\frac{D}{2}+1\right)\right]^{2/D}\left[\frac{\Gamma\left(n+\frac{2}{D}\right)}{\Gamma(n)} - \frac{\Gamma^2\left(n+\frac{1}{D}\right)}{\Gamma^2(n)}\right]\left(\frac{1}{N}\right)^{2/D} \qquad (10)
\end{aligned}
$$

which vanishes as $D \to \infty$. This suggests that the form of $\langle r_n \rangle$ for large $D$ can be arrived at by neglecting fluctuations, an approach which corresponds to mean-field theories in statistical mechanics.

## 2 A 'mean-field' theory

By the following 'mean-field' argument we derive an expression for the average $n$-th neighbour distance in large dimensions $D$. Since the average first neighbour distance can be found easily, we derive $\langle r_n \rangle$ in terms of $\langle r_1 \rangle$. As before we consider $N$ (a large number of) random points distributed uniformly within a unit volume of a $D$-dimensional geometric space. We choose any one of them as the reference point and locate its $n$-th neighbour. Neglecting fluctuations, which we can do for large $D$, the distance between them is $r_n(N) \approx \langle r_n(N) \rangle$. Keeping these two points fixed we change the number of points in the unit volume to $N\alpha$ by adding or removing points at random; the factor $\alpha$ is arbitrary to the extent that $N\alpha$ and $n\alpha$ are natural numbers. Since the distribution of points is uniform, the hypersphere that had originally enclosed just $n$ points will now contain $n\alpha$ points. Therefore, what was originally the $n$-th neighbour of the reference point now becomes the $n\alpha$-th neighbour. Since the two points under consideration are fixed, so is the distance between them. Consequently,

$$\langle r_n(N) \rangle \approx \langle r_{n\alpha}(N\alpha) \rangle \, . \tag{11}$$

Now we take $\alpha = 1/n$, so that

$$\langle r_n(N) \rangle \approx \langle r_1(N/n) \rangle \, , \tag{12}$$

which shows that the average $n$-th neighbour distance for a set of $N$ random points distributed uniformly is approximately given by the average distance for a depleted set of $N/n$ random points in the same volume. Using the expression for $\langle r_1(N) \rangle$ from equation 4 we get

$$\langle r_n \rangle \approx \frac{1}{\pi^{1/2}} \left[ \Gamma\left(\frac{D}{2}+1\right)\right]^{1/D} \Gamma\left(1+\frac{1}{D}\right) \left(\frac{n}{N}\right)^{1/D} . \tag{13}$$

Since the above argument neglects fluctuations the result of equation 13 ought to be exact in the limit $D \to \infty$.

The exact expression of $\langle r_n \rangle$ for a finite dimension $D$ is expected to reduce to the form of equation 13 as $D \to \infty$ where fluctuations do not affect. For large $D$ equation 9 takes the following form :

$$\langle r_n \rangle \approx \frac{1}{\pi^{1/2}} \left[ \Gamma\left(\frac{D}{2}+1\right)\right]^{1/D} \Gamma\left(1+\frac{1}{D}\right) \left(1+\frac{1}{D}\sum_{k=1}^{n-1}\frac{1}{k}\right) \left(\frac{1}{N}\right)^{1/D} . \tag{14}$$

For the above expression to reduce to the form of equation 13 the sum $\sum_{k=1}^{n-1}\frac{1}{k}$ must be equal to $\log_e n$ which happens only in the limit $n \to \infty$. Thus for any finite $n$ the exact result of equation 9 fails to produce the fluctuation-free form of equation 13 in large dimensions $D$. This shows that the 'mean-field' approximation is useless in the present context. However the 'mean-field' approach for $\langle r_n \rangle$ may be used as a crude approximation in other distributions (non-uniform) of random points where an exact calculation is not possible beyond the first neighbour.

# Acknowledgement

We thank D. Dhar, S. S. Manna and A. Percus for their comments.

# References

[1] S. Chandrasekhar, Rev. Mod. Phys. **15** (1943) 1.

[2] S. Chandrasekhar, An introduction to the theory of stellar structure (University of Chicago Press, Chicago, 1957).

[3] J. Beardwood, J. H. Halton and J. M. Hammersley, Proc. Camb. Phil. Soc. **55** (1959) 299.

[4] A. Percus and O. Martin, Phys. Rev. Lett. **76** (1996) 1188.