Nearest-Neighbor Graphs

Steven Finch

April 25, 2008

Consider a set P of n points that are independently and uniformly distributed in the d-dimensional unit cube. Let $p \in P$. There exists almost-surely $q \in P$ such that $q \neq p$ and $|p-q| < |p-r|$ for all $r \in P$, $r \neq p$, $r \neq q$. The point q is called the **nearest neighbor** of p and we write $p \prec q$. Note that $p \prec q$ does not imply $q \prec p$. Draw an edge connecting p and q if and only if $p \prec q$; the resulting graph of n vertices and $\leq n$ edges is called the **nearest-neighbor graph** G on P.

What is the probability, $\alpha(d)$, given $p \in P$, that $p \prec q$ implies $q \prec p$? Such a pair is **isolated** from the rest of G, in the sense that the only edge touching p or q is the edge that connects p and q. We have [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]

$$\alpha(1) = \frac{2}{3}, \qquad \alpha(2) = \frac{6\pi}{8\pi + 3\sqrt{3}} = 0.6215048968..., \qquad \alpha(3) = \frac{16}{27}$$

and, more generally [9],

$$\alpha(d) = \begin{cases} \left[\dfrac{3}{2} + \dfrac{1}{2}\displaystyle\sum_{k=1}^{\ell} \frac{1 \cdot 3 \cdots (2k-1)}{2 \cdot 4 \cdots (2k)} \left(\frac{3}{4}\right)^k\right]^{-1} & \text{if } d = 2\ell + 1, \\ \left[\dfrac{4}{3} + \dfrac{\sqrt{3}}{2\pi}\left(1 + \displaystyle\sum_{k=1}^{\ell-1} \frac{2 \cdot 4 \cdots (2k)}{3 \cdot 5 \cdots (2k+1)} \left(\frac{3}{4}\right)^k\right)\right]^{-1} & \text{if } d = 2\ell. \end{cases}$$

Here is a variation of the preceding. Draw an edge connecting p and q if and only if $q \prec p$; the resulting graph of n vertices and $\leq n$ edges is called the **nearest-neighbor anti-graph** H on P. What is the probability, $\beta(d)$, that $p \in P$ is isolated from the rest of H? That is, what proportion of points in P are not nearest neighbors of any other points? We have [16, 17, 18, 19, 20, 21]

$$\beta(1) = \frac{1}{4}, \qquad \beta(2) \approx 0.28, \qquad \beta(3) \approx 0.30$$

but the latter two estimates are only simulation-based. To further understand $\beta(2)$ will occupy us for the remainder of this essay.

Define constants $C(0,d)=1$ and

$$C(k,d)=\int\limits_{\Omega(k,d)}\exp\left[-\operatorname{Vol}\left(\bigcup_{j=1}^{k}S(x_j)\right)\right]dx_1dx_2\ldots dx_k$$

for $k\geq 1$, where $S(x_j)$ is the ball in $\mathbb{R}^d$ of radius $|x_j|$, centered at x_j, and

$$\Omega(k,d)=\left\{(x_1,x_2,\ldots,x_k)\in\mathbb{R}^{dk}:|x_i|\leq|x_i-x_j|\text{ for all }1\leq i\neq j\leq k\right\}.$$

It is known that [19, 22, 23, 24, 25]

$$\beta(2)=\sum_{k=0}^{6}\frac{(-1)^k}{k!}C(k,2),\qquad\beta(3)=\sum_{k=0}^{12}\frac{(-1)^k}{k!}C(k,3)$$

and clearly $C(1,d)=1$, $C(2,1)=1/2$. The upper limits of summation are the *kissing numbers* in $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. A proof that 24 is the kissing number in $\mathbb{R}^4$ was given only recently [26, 27]. Also, $C(6,2)=0$ since $\Omega(6,2)$ is of measure zero.

Henze [24, 25] showed that

$$C(2,d)=\frac{2^{d+1}\pi^{d-1}}{\Gamma(d-1)}\int\limits_0^{\infty}\int\limits_0^{\xi}\int\limits_{\theta_0}^{\pi}\xi^{d-1}\eta^{d-1}\sin(\theta)^{d-2}F_d(\xi,\eta)\,d\theta\,d\eta\,d\xi$$

where

$$\theta_0=\arccos\left(\frac{\eta}{2\xi}\right),$$

$$F_d(\xi,\eta)=\exp\left[-f_d(\xi,\gamma)-f_d(\eta,\delta)\right],$$

$$\gamma=\frac{\xi(\xi-\eta\cos(\theta))}{\sqrt{\xi^2+\eta^2-2\xi\,\eta\cos(\theta)}},\qquad\delta=\frac{\eta(\eta-\xi\cos(\theta))}{\sqrt{\xi^2+\eta^2-2\xi\,\eta\cos(\theta)}},$$

$$f_d(x,y)=\frac{\pi^{d/2}x^d}{2\Gamma(d/2+1)}\left[1+I\left(\frac{y^2}{x^2},\frac{1}{2},\frac{d+1}{2}\right)\right]$$

and I is the regularized beta function

$$I(z,a,b)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int\limits_0^{z}w^{a-1}(1-w)^{b-1}dw.$$

(In [24], the definitions of γ and δ were mistakenly reversed; also, the expression within square brackets for $f_d(x,y)$ was unclear.) We obtain

$$C(2,2)=0.63317...=2(0.316585...),\qquad C(2,3)=0.70888....$$

Tao & Wu [19] independently showed that

$$\begin{aligned} C(2,2) &= \pi \int_{\pi/2}^{\pi} \int_{0}^{\infty} \frac{\tau}{\left(g(\tau,\theta)+\tau^2 h(\tau,\theta)\right)^2} d\tau\, d\theta + \\ & \quad \pi \int_{\pi/3}^{\pi/2} \int_{2\cos(\theta)}^{1/(2\cos(\theta))} \frac{\tau}{\left(g(\tau,\theta)+\tau^2 h(\tau,\theta)\right)^2} d\tau\, d\theta \end{aligned}$$

where

$$g(\tau,\theta) = \pi - \varphi + \tfrac{1}{2}\sin(2\varphi), \qquad h(\tau,\theta) = \pi - \psi + \tfrac{1}{2}\sin(2\psi),$$

$$\varphi = \arcsin\left(\frac{\tau\sin(\theta)}{\sqrt{1+\tau^2-2\tau\cos(\theta)}}\right), \qquad \psi = \arcsin\left(\frac{\sin(\theta)}{\sqrt{1+\tau^2-2\tau\cos(\theta)}}\right).$$

(Several underlying details in [19] are clarified in [28].) Even more elaborate integral formulas apply for $C(3,2)$, $C(4,2)$, $C(5,2)$. Given the discrepancy between our estimate of $C(2,2)$ and their estimate (see the Table), it seems doubtful that their approximation $\beta(2) = 0.284051...$ is entirely correct.

Table 1 *Old and New Calculations of Constants*

k	Tao & Wu estimate of $C(k,2)/k!$	Current estimate of $C(k,2)/k!$
2	0.3163335...	0.316585...
3	0.0329390...	0.033056...
4	0.0006575...	still open
5	0.0000010...	still open

A discrete version of the latter problem appears in [29, 30, 31, 32]. Let all the vertices of the lattice $\mathbb{Z}^d$ be initially occupied by particles which can annihilate one-by-one their $2d$ nearest neighbors. More precisely, for each unit-length edge $\{u,v\}$ of the lattice, there is a Uniform $[0,1]$ random variable $T_{\{u,v\}}$ representing the time of an attack along the edge. If vertices u, v are both occupied immediately prior to time $T_{\{u,v\}}$, then at time $T_{\{u,v\}}$ either vertex u or vertex v (each with probability $1/2$) becomes vacant (that is, one particle annihilates the other). If u, v are not both occupied at time $T_{\{u,v\}}$, then there is no change. Once a vertex becomes vacant, it remains vacant permanently. The variables $T_{\{u,v\}}$, considered over all unit-length edges $\{u,v\}$, are independent. By time 1, no two surviving particles can be adjacent. When $d=1$, the probability that a given vertex remains occupied is $1/e = 0.3678794411....$ When $d=2$, this probability is known to be in the interval $(0.227, 0.306)$ and is approximately 0.25 via simulation. Greater accuracy is desired.

References

[1] P. J. Clark and F. C. Evans, On some aspects of spatial pattern in biological populations, *Science* 121 (1955) 397–398.

[2] P. J. Clark, Grouping in spatial distributions, *Science* 123 (1956) 373–374.

[3] M. F. Dacey, Proportion of reflexive n^{th} order neighbors in spatial distribution, *Geographical Analysis* 1 (1969) 385–388.

[4] G. F. Schwarz and A. Tversky, On the reciprocity of proximity relations, *J. Math. Psych.* 22 (1980) 157–175; MR0609119 (82f:92050).

[5] T. F. Cox, Reflexive nearest neighbours, *Biometrics* 37 (1981) 367–369; MR0673043 (83k:62122).

[6] D. P. Shine and J. Herbert, Birds on a wire, *J. Recreational Math.* 11 (1978-79) 227–228; 15 (1982-83) 232.

[7] S. Morris, Competition winners, *Omni* v. 2 (1980) n. 9, p. 108.

[8] C. Kluepfel, Birds on a wire, cows in the field, and stars in the heavens, *J. Recreational Math.* 13 (1980-81) 241–245.

[9] D. K. Pickard, Isolated nearest neighbors, *J. Appl. Probab.* 19 (1982) 444–449; MR0649985 (83g:60063).

[10] M. F. Schilling, Mutual and shared neighbor probabilities: finite- and infinite-dimensional results, *Adv. Appl. Probab.* 18 (1986) 388–405; MR0840100 (87k:60041).

[11] N. Henze, On the probability that a random point is the j^{th} nearest neighbour to its own k^{th} nearest neighbour, *J. Appl. Probab.* 23 (1986) 221–226; MR0826925 (87k:60133).

[12] D. Eppstein, M. S. Paterson and F. F. Yao, On nearest-neighbor graphs, *Discrete Comput. Geom.* 17 (1997) 263–282; http://www.ics.uci.edu/~eppstein/pubs/EppPatYao-DCG-97.pdf; MR1432064 (98d:05121).

[13] D. P. Shine and M. P. Cohen, Spread the news, *J. Recreational Math.* 36 (2007) 277–278.

[14] C. A. S. Tercariol, F. de Moura Kiipper and A. Souto Martinez, An analytical calculation of neighbourhood order probabilities for high dimensional Poissonian processes and mean field models, *J. Phys. A* 40 (2007) 1981–1989; cond-mat/0609210; MR2316309 (2008a:82035).

[15] P. J. Campbell and B. Atwood, The farmer problem, *UMAP Journal* 33 (2012) 313–331.

[16] F. D. K. Roberts, Nearest neighbours in a Poisson ensemble, *Biometrika* 56 (1969) 401–406.

[17] R. Abilock and M. Goldberg, N riflemen, *Amer. Math. Monthly* 75 (1968) 1009; 89 (1982) 274–275.

[18] S. Morris, Rifle puzzle, *Omni* v. 8 (1986) n. 4, p. 113; v. 9 (1987) n. 7, p. 141.

[19] R. Tao and F. Y. Wu, The vicious neighbour problem, *J. Phys. A* 20 (1987) L299–L306; MR0888078 (88d:82021).

[20] E. G. Enns, P. F. Ehlers and T. Misi, A cluster problem as defined by nearest neighbours, *Canad. J. Statist.* 27 (1999) 843–851; MR1767151 (2001b:60017).

[21] S. Portnoy, A squirtgun battle, *J. Recreational Math.* 37 (2008) 39–45.

[22] C. M. Newman, Y. Rinott and A. Tversky, Nearest neighbors and Voronoi regions in certain point processes, *Adv. Appl. Probab.* 15 (1983) 726–751; MR0721703 (85m:60023).

[23] C. M. Newman and Y. Rinott, Nearest neighbors and Voronoi volumes in high-dimensional point processes with various distance functions, *Adv. Appl. Probab.* 17 (1985) 794–809; MR0809431 (87d:60048).

[24] N. Henze, Über die Anzahl von Zufallspunkten mit typ-gleichem nächsten Nachbarn und einen multivariaten Zwei-Stichproben-Test, *Metrika* 31 (1984) 259–273; MR0773815 (86i:62075).

[25] N. Henze, On the fraction of random points with specified nearest-neighbour interrelations and degree of attraction, *Adv. Appl. Probab.* 19 (1987) 873–895; MR0914597 (89c:60063).

[26] O. R. Musin, The kissing number in four dimensions, *Annals of Math.* 168 (2008) 1–32; math.MG/0309430; MR2415397.

[27] F. Pfender and G. M. Ziegler, Kissing numbers, sphere packings, and some unexpected proofs, *Notices Amer. Math. Soc.* 51 (2004) 873–883; MR2145821 (2006a:52015).

[28] S. R. Finch, Union of n disks: remote centers, common origin, arXiv:1511.04968.

[29] M. O'Hely and A. Sudbury, The annihilating process, *J. Appl. Probab.* 38 (2001) 223–231; MR1816125 (2001m:60226).

[30] A. Sudbury, Inclusion-exclusion methods for treating annihilating and deposition processes, *J. Appl. Probab.* 39 (2002) 466–478; MR1928883 (2003k:60266).

[31] A. Sudbury, The annihilating process on random trees and the square lattice, *J. Appl. Probab.* 41 (2004) 816–831; MR2074826 (2005g:82090).

[32] M. D. Penrose and A. Sudbury, Exact and approximate results for deposition and annihilation processes on graphs, *Annals Appl. Probab.* 15 (2005) 853–889; MR2114992 (2005k:60307).