Nearest Neighbor Distances in 2D

Brian Lee

October 21, 2012

1 Introduction/Problem Statement

A coworker recently asked me the following question: given a spherical vesicle / cell with dye molecules embedded in the surface, what is the average nearest-neighbor distance (NND)? This number is important because in FRET assays, you have two types of dye molecules - the donor chromophore and the acceptor chromophore. The efficiency of energy transfer goes as $1/r^6$, and so for FRET purposes, only the nearest neighbor matters - even if you were to add up the contributions from every other possible pairing, it would be tiny compared to the contribution from the NN pair.

To simplify the spherical geometry, I would like to use an infinite, flat plane. FRET is typically effective over distances of $<$ 10nm, and given that most vesicles/cells have diameters of 100nm - 10μm, we can consider the NND interactions to happen in a region that is approximately 'locally flat'. (Interesting tidbit - the deviation from flatness, as measured by the angle excess of a triangle, is proportional to the area of the triangle divided by the radius of curvature squared. This is Girard's theorem, a special case of the Gauss-Bonnet theorem.)

Additionally, we'd like to have more than just the 'average' NND. It would be best if we could calculate the probability distribution of nearest neighbor distances. From there, we can calculate average NN distance, mode NN distance, and all sorts of other goodies.

So our problem can be restated as follows: in an infinite, flat plane with points of density $\sigma = N/4\pi r^2$, what is the probability distribution function $\rho(r)$ of NN-distances?

2 Dimensional analysis

NN distance has units of length (L). σ has units of $1/L^2$. Therefore, NN distance must be proportional to $1/\sqrt{\sigma}$. Or more precisely,

$$NND = \frac{C}{\sqrt{\sigma}}$$

where C is some as-of-yet undetermined constant that depends on the geometry of our problem.

3 Some simple cases

Let's solve some simple cases, to gain some intuition for our problem.

3.1 Square grid

Assume our points are spaced in a square grid. Then, the NN distance is equal to the side length of a square, and we know the density of our grid must equal the overall density. Each square has four points associated with it, and each point is shared between four squares.

$$\frac{4 \cdot \frac{1}{4}}{NND^2} = \sigma$$

$$NND = \frac{1}{\sqrt{\sigma}}$$

3.2 Triangular grid

Assume our points are spaced in a triangular grid. Then, the NN distance is equal to the side length of a triangle, and we have the following relationship:

$$\frac{3 \cdot \frac{1}{3}}{\frac{NND^2\sqrt{3}}{4}} = \sigma$$

$$NND = \frac{2}{3^{1/4}}\frac{1}{\sqrt{\sigma}} \approx 1.5197\frac{1}{\sqrt{\sigma}}$$

3.3 Hexagonal grid

Assume our points are on the vertices of a hexagonal grid. The NN distance is equal to the side length of a hexagon.

$$\frac{6 \cdot \frac{1}{3}}{6 \cdot \frac{NND^2\sqrt{3}}{4}} = \sigma$$

$$NND = \frac{2}{27^{1/4}} \frac{1}{\sqrt{\sigma}} \approx .8774 \frac{1}{\sqrt{\sigma}}$$

3.4 Simple cases - conclusion

It looks like with more efficient grids (triangular), our NND actually becomes larger! This makes sense - for a specified density σ, the most efficient grids give each point the most breathing room. The least efficient grid (hexagonal) gives us about half as much breathing room as the most efficient grid. We should expect randomized points to be even less efficient than a hexagonal grid.

4 The Random Case

The random case seems quite difficult at first. While math makes it easy to express infinite concepts like "take the average distance over all pairs of points", it's not easy to express finite concepts like "take the nearest neighbor, and neglect the others".

The key observation is this: the probability that the NN is located at some distance r must equal the probability that no NN exists closer than r, times the probability that there exists some point at a distance r.

The probability that the NN is located between distances r and $r + dr$ is, by definition, equal to

$$\rho(r)dr$$

The probability that some point exists between distance r and $r + dr$ is equal to the area times density of points.

$$2\sigma\pi r dr$$

The probability that no point exists closer than distance r is one minus the probability that there does exist a point closer than r.

$$(1 - \int_0^r \rho(r)dr)$$

Putting it all together and dividing by $2\sigma\pi r dr$, we have this equation for ρ:

$$\frac{\rho(r)}{2\sigma\pi r} = 1 - \int_0^r \rho(r)dr$$

Take a derivative of the whole thing to yield the following differential equation:

$$\frac{d}{dr}\left(\frac{\rho(r)}{2\sigma\pi r}\right) = -\rho(r) = -2\sigma\pi r\frac{\rho(r)}{2\sigma\pi r}$$

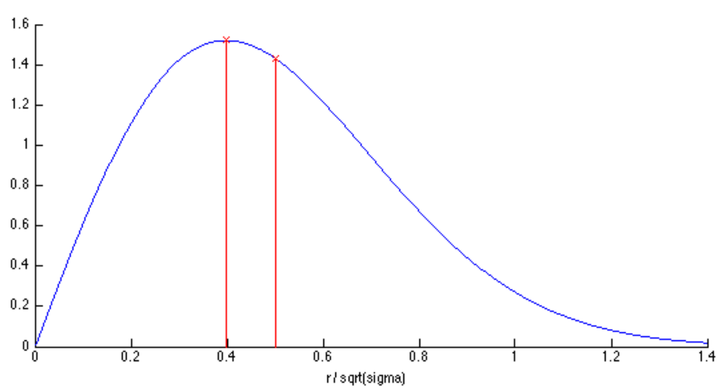
I rewrote the right hand side in a way that suggests that the solution to this differential equation is: (check!)

$$\rho(r) = 2\sigma\pi r e^{-\sigma\pi r^2}$$

Conveniently, this solution is already normalized - our probabilities had better add up to 1!

$$\int_0^\infty \rho(r)dr = -e^{-\sigma\pi r^2}\Big|_0^\infty = 1$$

Let's rest for a moment and see what we've got on our hands. This function is the product of an increasing function r times a much more quickly decreasing function e^{-r^2}. Therefore, at some intermediate point, the probability peaks! We can figure out this maximum by taking a derivative and setting it equal to zero.



$$\frac{d\rho}{dr} = 2\sigma\pi e^{-\sigma\pi r^2} - (2\sigma\pi r)^2 e^{-\sigma\pi r^2} = 0$$

$$2\sigma\pi r^2 = 1$$

$$r = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{\sigma}} \approx .3989\frac{1}{\sqrt{\sigma}}$$

That looks familiar. There's the $\frac{1}{\sqrt{\sigma}}$ relationship, as promised, and a coefficient that's twice as small as from our hexagonal grid.

We can now answer our original question of the average NND by integrating from zero to infinity. There's a pretty nifty trick along the way.

$$
\begin{aligned}
< NND > &= \int_0^\infty r\rho(r)dr = \int_0^\infty 2\sigma\pi r^2 e^{-\sigma\pi r^2} && (1)\\
&= \int_0^\infty -\sigma\frac{\partial}{\partial\sigma}\left(e^{-\sigma\pi r^2}\right)dr && (2)\\
&= -\sigma\frac{\partial}{\partial\sigma}\int_0^\infty e^{-\sigma\pi r^2}dr && (3)\\
&= -\sigma\frac{\partial}{\partial\sigma}\left(\frac{1}{\sqrt{\sigma}}\right) && (4)\\
&= \frac{1}{2\sqrt{\sigma}} && (5)
\end{aligned}
$$

The reordering of partial derivatives and integrals in step 3 is allowed since they occur over two different, independent variables and are orthogonal in some sense. Step 4 uses the well-known fact that the integral of a Gaussian function is $\sqrt{\pi}$. We get the surprising result (confirmed by numerical simulations) that the average NND for random points is exactly one half the lattice spacing of a square grid!

Finally, as $r \to 0$, the e^{-r^2} factor approaches 1, and so $\rho(r) \to 2\sigma\pi r$. This makes sense - at very close distances, it is very unlikely that a closer point exists, and so the NND probability density function is directly proportional to the circumference of the circle at that distance.

All of this analysis can also be done in three dimensions, where the final solution turns out to be:

$$\rho(r) = 4\sigma\pi r^2 e^{-\frac{4}{3}\sigma\pi r^3}$$