17-19 April 2019

Random Walk Algorithms: Lectures 8 & 9

David A. Meyer

In which Laplace's derivation of the continuum approximation to a binomial probability distribution is applied to the goal of guessing a nontrivial solution of the heat equation, the heat kernel, to make sense of the $t \to 0$ limit of which several elementary notions from the theory of distributions are introduced.

Approximating the probability distribution function

In the previous lecture we solved the advection equation. The next goal is to solve the heat equation. Our strategy is to find the solution to $\mathrm{RW}^2(1/2)$, *i.e.*, its probability distribution function at timestep t, approximate this with an at least twice differentiable function on $\mathbb{R}$ and check that this function satisfies the heat equation. Recall that

$$\Pr(X_t = s) = u_{t,s} = (B^t \vec{u}_0)_s,$$

where

$$B = \frac{1}{4}\begin{array}{c} \\ \vdots \\ -2 \\ 0 \\ 2 \\ \vdots \end{array}\!\!\begin{array}{c} \begin{array}{ccccc} \ldots & -2 & 0 & 2 & \ldots \end{array} \\ \begin{pmatrix} & & & & \\ & \ddots & 1 & & \\ & & 2 & & \\ & & 1 & \ddots & \\ & & & & \end{pmatrix} \end{array}.$$

We can, of course, compute B^t by matrix multiplication, or by referring back to the binomial distribution we derived for $\mathrm{RW}^1(p)$ in Lecture 4. As a preview of Lecture 10, however, we notice that if we denote the unit vectors of the basis in which B is written above as $\{\ldots, \hat{e}_{-4}, \hat{e}_{-2}, \hat{e}_0, \hat{e}_2, \hat{e}_4, \ldots\}$, then

$$\begin{aligned}
\vec{u}_0 &= \hat{e}_0 \\
\vec{u}_1 = B\vec{u}_0 = B\hat{e}_0 &= \frac{1}{4}(\hat{e}_{-2} + 2\hat{e}_0 + \hat{e}_2) \\
\vec{u}_2 = B\vec{u}_1 &= \frac{1}{4}(B\hat{e}_{-2} + 2B\hat{e}_0 + B\hat{e}_2) \\
&= \frac{1}{4^2}\big((\hat{e}_{-4} + 2\hat{e}_{-2} + \hat{e}_0) + 2(\hat{e}_{-2} + 2\hat{e}_0 + \hat{e}_2) + (\hat{e}_0 + 2\hat{e}_2 + \hat{e}_4)\big) \\
&= \frac{1}{4^2}(\hat{e}_{-4} + 4\hat{e}_{-2} + 6\hat{e}_0 + 4\hat{e}_2 + \hat{e}_4),
\end{aligned}$$

etc., the crucial point being that B acts linearly, and so can be evaluated on each basis vector and the results, weighted by the probability of that basis vector, summed.

After t steps, the probabilities are binomial:

$$u_{t,s} = \frac{1}{2^{2t}} \binom{2t}{t+s/2} = \frac{1}{2^n} \binom{n}{k} = \frac{1}{2^n} \frac{n!}{k!(n-k)!},$$

where we set $n = 2t$ and $k = t + s/2$ for temporary convenience. We will approximate this near the center of the probability distribution, that is, within a constant number of standard deviations, so that $k = n/2 + O(\sqrt{n})$, *i.e.*, $s = O(\sqrt{t})$; further from the center we can approximate the distribution simply by 0. Following Laplace [1], we start by using Stirling's approximation [2,3],

$$n! = n^n e^{-n} \sqrt{2\pi n} \Big(1 + O(\frac{1}{n})\Big),$$

to write

$$u_{t,s} = \Big(\frac{n}{2k}\Big)^k \Big(\frac{n}{2(n-k)}\Big)^{n-k} \sqrt{\frac{n}{2\pi k(n-k)}} \Big(1 + O(\frac{1}{n})\Big).$$

Next, note that

$$\begin{aligned} \log\Big(\frac{n}{2k}\Big) &= \log\Big(\frac{n}{n+s}\Big) = -\log\Big(1 + \frac{s}{n}\Big) \\ \log\Big(\frac{n}{2(n-k)}\Big) &= \log\Big(\frac{n}{n-s}\Big) = -\log\Big(1 - \frac{s}{n}\Big), \end{aligned}$$

and recall that $\log(1+\epsilon) = \epsilon - \frac{1}{2}\epsilon^2 + O(\epsilon^3)$, so

$$\begin{aligned} \log\Big(\frac{n}{2k}\Big)^k \Big(\frac{n}{2(n-k)}\Big)^{n-k} &= k \log\Big(\frac{n}{2k}\Big) + (n-k) \log\Big(\frac{n}{2(n-k)}\Big) \\ &= -\frac{n+s}{2}\Big(\frac{s}{n} - \frac{s^2}{2n^2} + O(\frac{s^3}{n^3})\Big) - \frac{n-s}{2}\Big(-\frac{s}{n} - \frac{s^2}{2n^2} + O(\frac{s^3}{n^3})\Big) \\ &= -\frac{s^2}{2n} + O(\frac{s^3}{n^2}). \end{aligned}$$

Exponentiating both sides of this equation gives

$$\Big(\frac{n}{2k}\Big)^k \Big(\frac{n}{2(n-k)}\Big)^{n-k} = e^{-s^2/(2n)} \Big(1 + O(\frac{s^3}{n^2})\Big).$$

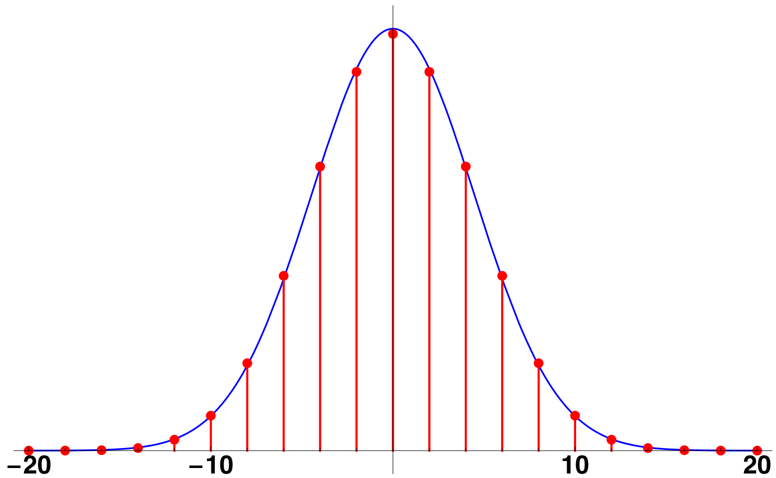
Also,

$$\sqrt{\frac{n}{2\pi k(n-k)}} = \sqrt{\frac{n}{\pi(n^2 - s^2)/2}} = \frac{1}{\sqrt{\pi n/2(1 - s^2/n^2)}} = \frac{1}{\sqrt{\pi n/2}} \Big(1 + O(\frac{s^2}{n^2})\Big).$$

Multiplying the last two results gives, finally,

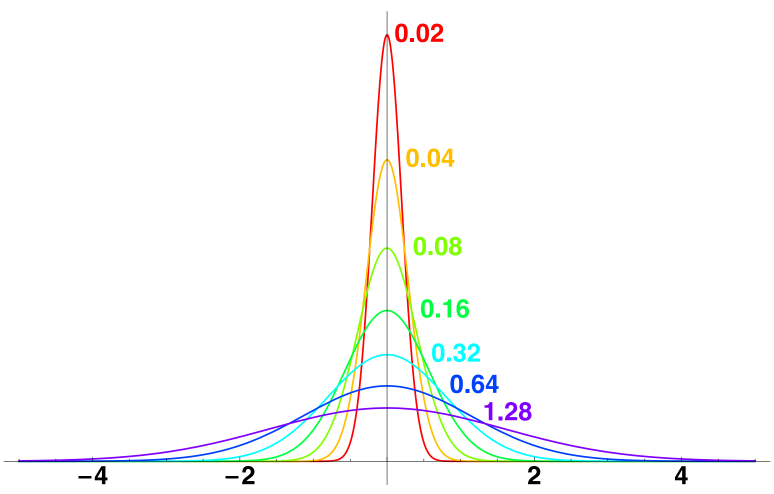
$$u_{t,s} = \frac{1}{\sqrt{\pi t}} e^{-s^2/(4t)} \Big(1 + O(\frac{s^3}{t^2})\Big).$$

The adjacent plot shows how well this approximation works, even for the fairly small value $t = 10$. Since $s \in 2\mathbb{Z}$, and the sum of the probabilities plotted in red is 1, the integral of this normal approximation must be close to 2; in fact it is exactly 2. Normalizing it to have integral 1 leads us to guess that



$$u(t,s) = \frac{1}{\sqrt{4\pi\alpha t}} e^{-s^2/(4\alpha t)}$$

satisfies the heat equation. In fact, it does, as can be readily checked by taking the partial derivatives. This one parameter family of Gaussian functions is called the *heat kernel*. Since each is a normal distribution with variance $2t$, $\Pr(-\sqrt{2t} < s < \sqrt{2t}) \approx 0.68$, $\Pr(-2\sqrt{2t} < s < 2\sqrt{2t}) \approx 0.95$, *etc.* Notice that this implies that our initial goal of approximating the binomial distribution within a constant number of standard deviations of its mean in fact represented almost all of the probability. The success of the approximation also implies that the probabilities of a binomial random variable being with a constant number of standard deviations of its mean are very close to those of the corresponding normal random variable.



This graph shows the heat kernel at a sequence of arguments t from 1.28 down to 0.02. The limit as $t \to 0$ is usually called the *Dirac delta function* [4], but it cannot be a function: the integral is 1 for all positive t, so if the limit exists it should still be 1 at the limit, but the function values go to 0 everywhere except at $x = 0$, where the limit does not exist. It is well-defined within the theory of distributions created by Schwartz [5]:

Definition. Let $C^\infty(\mathbb{R})$ be the vector space of smooth functions on $\mathbb{R}$, and let $C_0^\infty(\mathbb{R})$ be the vector space of smooth functions with *compact support* on $\mathbb{R}$. Here $\phi \in C^\infty(\mathbb{R})$ has compact support if there exists $N > 0$ such that $\phi(x) = 0$ for all x such that $|x| > N$.

Definition. A *distribution* is a linear map $C_0^\infty(\mathbb{R}) \to \mathbb{R}$, *i.e.*, an element of the dual space of $C_0^\infty(\mathbb{R})$.

EXAMPLE. Let $f : \mathbb{R} \to \mathbb{R}$ be an integrable function. Then $L_f : C_0^\infty(\mathbb{R}) \to \mathbb{R}$ is the distribution defined by

$$\langle \phi, L_f \rangle = L_f(\phi) = \int \phi(x) f(x) \mathrm{d}x.$$

EXAMPLE. $\delta_y : C_0^\infty(\mathbb{R}) \to \mathbb{R}$ is defined by

$$\langle \phi, \delta_y \rangle = \phi(y),$$

which is written, in analogy with L_f above, as

$$\phi(y) = \int \phi(x) \delta_y(x) \mathrm{d}x = \int \phi(x) \delta(y - x) \mathrm{d}x,$$

justifying the terminology “delta function”.

References

[1] P.-S. Marquis de Laplace, *Théorie Analytique des Probabilités*, Première partie (Paris: Mme. Ve. Courcier 1812) Chapitre III.

[2] A. de Moivre, *Miscellaneis Analyticus Supplementum* (London: J. Tonson & J. Watts 1730).

[3] J. Stirling, *Methodus Differentialis: Sive Tractatus de Summatione et Interpolatione Serierum Infinitarum* (London: Gul. Bowyer & G. Strahan 1730).

[4] P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford: Clarendon Press 1930).

[5] L. Schwartz, *Théorie des Distributions*, Tome 1 (Paris: Hermann & Cie 1950).