Approximations using Dawson's integral

(by Chris H. Rycroft, `chr@mit.edu`.)

Consider a random walk following a fat-tailed distribution with PDF

$$p(x) = \frac{2}{\sqrt{\pi}(1+x^4)}.$$

This PDF has zero mean and unit variance, but the "absolute moments" of the form $\langle |x|^m \rangle$ diverge for $m \geq 3$. Near the origin, the Fourier transform is not analytic, and has the expansion

$$\hat{p}(k) = 1 - \frac{k^2}{2} + \frac{|k|^3}{3\sqrt{2}} - \frac{k^4}{24} + O(|k|^5).$$

The cumulant generating function has the form

$$\psi(k) = \frac{-k^2}{2} + \frac{|k|^3}{3\sqrt{2}} + O(k^4)$$

which we can use to find approximations after N steps. If we consider just the first term, we obtain the Central Limit Theorem (CLT):

$$P_N(x) \approx \frac{1}{\sqrt{2\pi N}} e^{-x^2/2N}$$

However, if we include the next term, we find that

$$P_N(x) \approx \frac{1}{\sqrt{2\pi N}} e^{-x^2/2N} + \frac{D'''(x/\sqrt{2N})}{6\pi N\sqrt{2}}$$

where $D(x)$ is Dawson's integral, defined as

$$D(x) = e^{-x^2} \int_0^x e^{t^2}\, dt.$$

Dawson's integral satisfies $D'''(x) = 4(x^2-1) + (12x - 8x^3)D(x)$, and it can be shown that $D(x) \sim 6x^{-4}$ as $x \to \infty$. The width of the central region is determined by the value of x where the Dawson correction has a comparable size to the CLT. We find the width is $\sqrt{N \log N}$ which is not much wider than the minimum required by the CLT, and much less than $N^{3/4}$ when all the cumulants are finite.

Simulations

Figure 1 shows a semi-log plot of the PDF. To compare the above predictions to simulation, we must first find a way to simulate steps from this distribution. Unlike the Cauchy distribution, a simple transformation can not be used in this case. Let X be a random variable with the one-sided PDF

$$p_X(x) = \frac{4}{\sqrt{\pi}(1+x^4)} \qquad \text{for } x > 0.$$

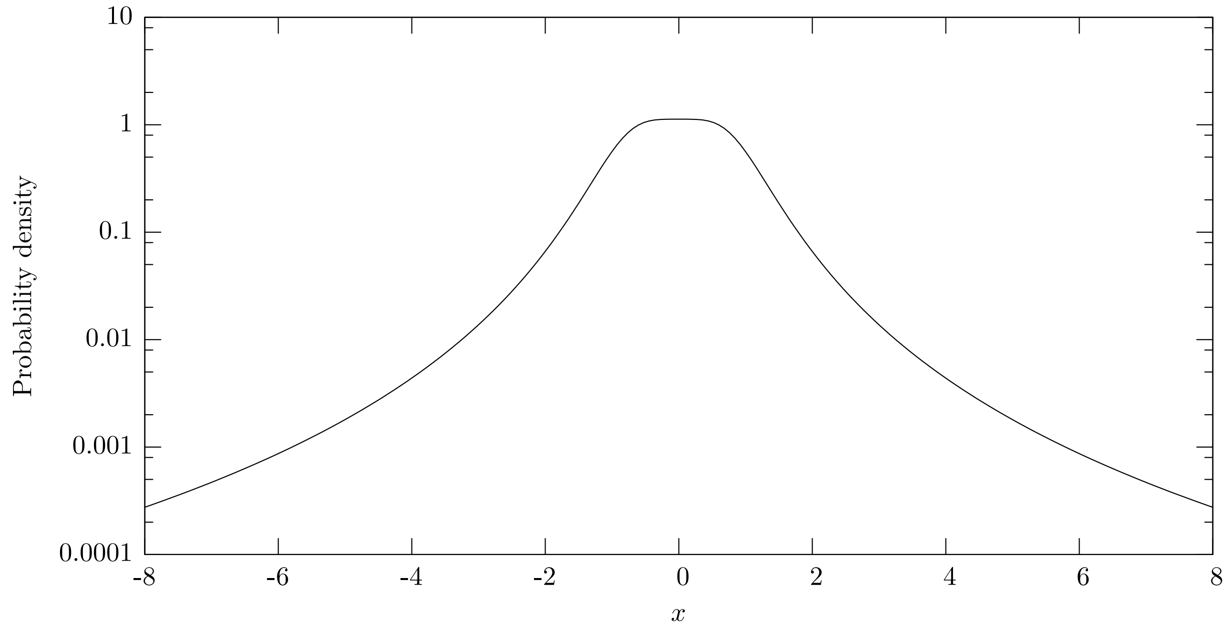

Figure 1: Semi-log plot of the PDF under consideration.

If we consider the sequence of substitutions $Z = X^2$, $\Theta = \tan^{-1} Z$, $\Phi = \sqrt{\Theta}$ then we find that Φ has PDF

$$p_\Phi(\phi) = \frac{4\phi}{\sqrt{\pi \tan \phi^2}} \qquad \text{for } 0 < \phi < \sqrt{\pi/2}.$$

Although this is a complicated function it is smooth over the given domain, and bounded between 0 and $4/\sqrt{\pi}$. To sample from this distribution, we choose random pairs (A, B) where A is uniform on $[0, \sqrt{\pi/2}]$, and B is uniform on $[0, 4/\sqrt{\pi}]$. If we keep sampling pairs until we find $B < p_\Phi(A)$, and then let $\Phi = A$, then Φ will follow the above PDF. Transforming back allows us to find X. Multiplying X by -1 with probability 1/2 lets us sample from the PDF we wish to study. This can be coded efficiently in C using the function

```
double fatrand () {
        double b;
        while (usq()*tan(b=usq()*pi/2)>b);
        b=sqrt(tan(b));return rand()%2==1?b:-b;
}
```

where `usq()` returns the square of a number uniformly distributed on $[0, 1]$ and `rand()` returns a random integer. Figure 2 shows four sample random walks from this distribution, and figures 3 to 6 show the accuracy of the Dawson correction for $N = 3$, $N = 10$, and $N = 100$.

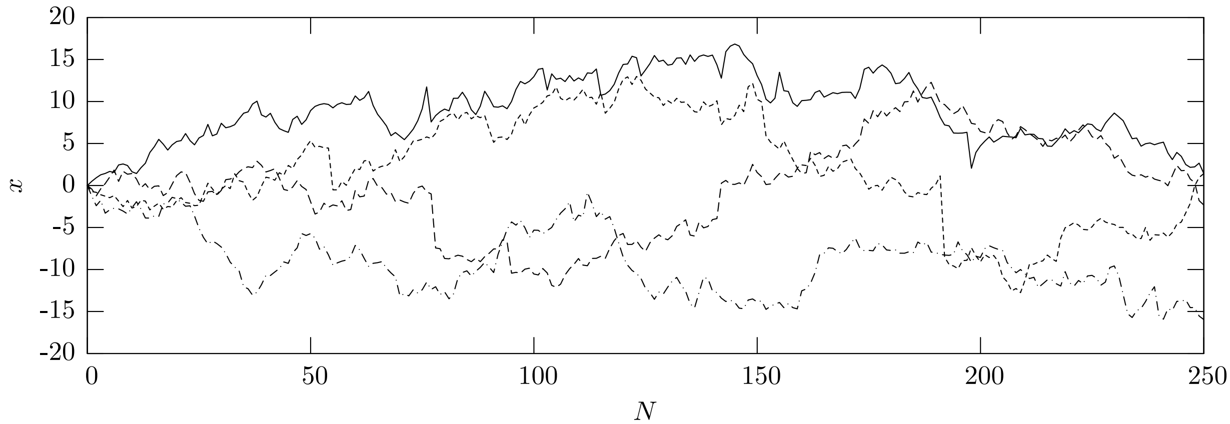

Figure 2: Four sample random walks using the PDF under consideration.

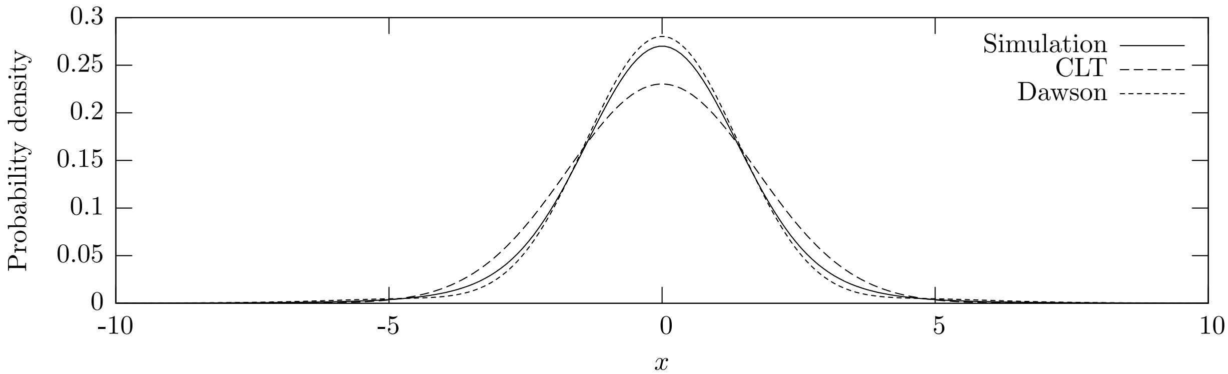

Figure 3: Comparisons between the Central Limit Theorem with and without the Dawson correction, compared to simulations of 1.6×10^{10} walks with $N = 3$ steps. Even in the central region, the Dawson curve is a better fit to the data.

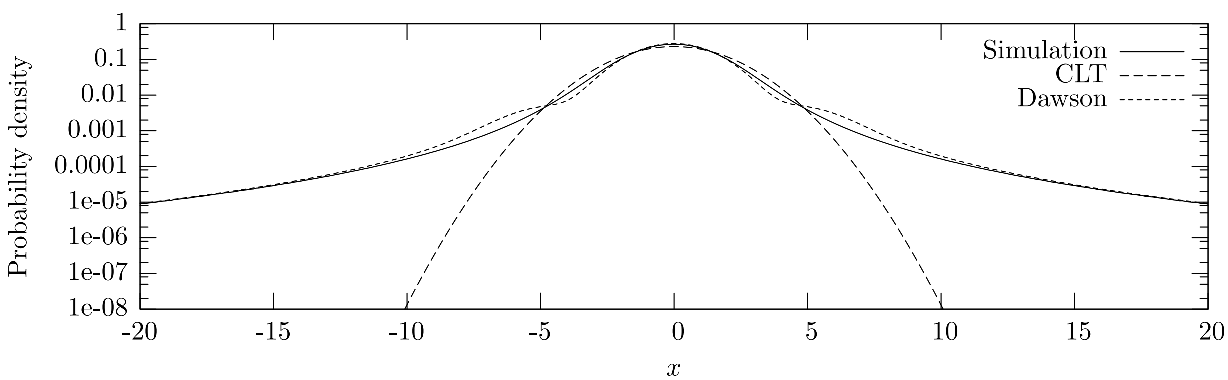

Figure 4: A semi-log plot of the $N = 3$ data shows that the Dawson correction matches the tail of the distribution to a high level of accuracy, although some small deviations are seen near $x = 5$.

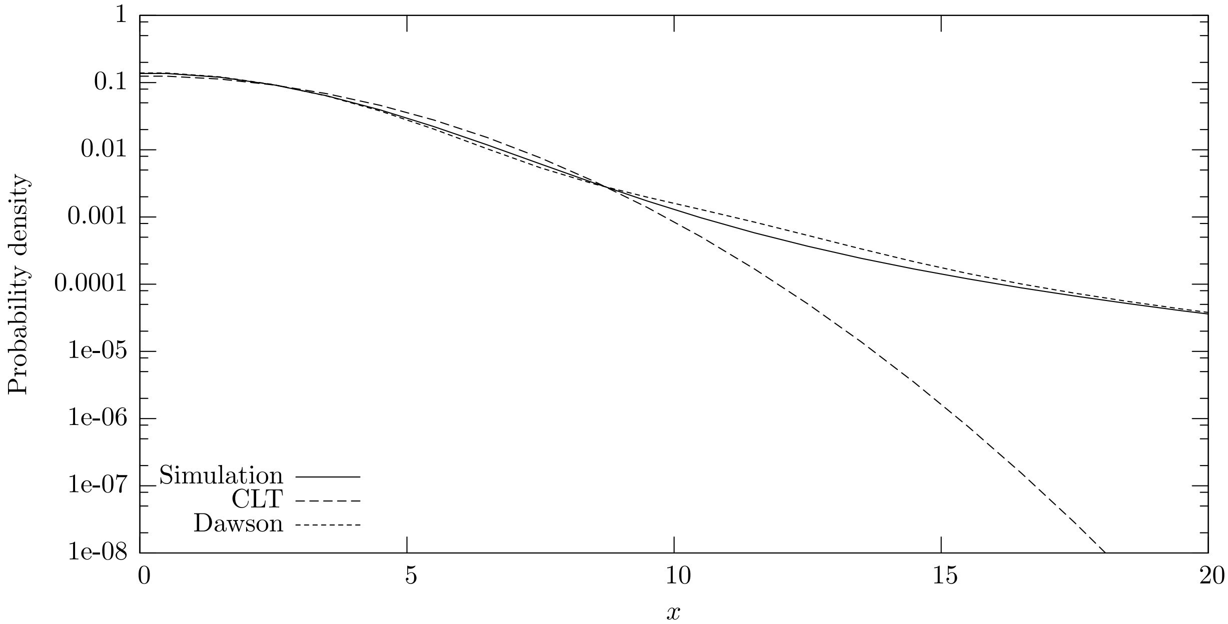

Figure 5: Comparisons between the Central Limit Theorem with and without the Dawson correction, compared to simulations of 1.6×10^{10} walks with $N = 10$ steps. Small deviations between the Dawson curve and the simulated PDF are apparent.

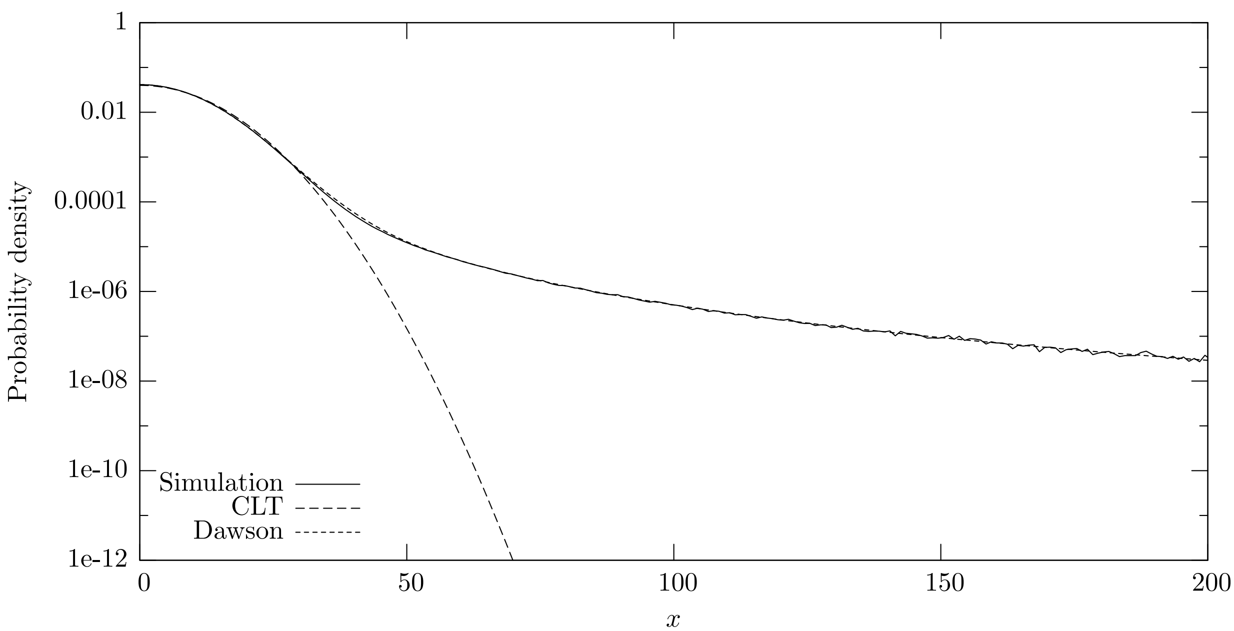

Figure 6: Comparisons between the Central Limit Theorem with and without the Dawson correction, compared to simulations of 1.6×10^{9} walks with $N = 100$ steps. The Dawson curve matches the simulated data to a high level of accuracy.