Solutions to Exam 1

Martin Z. Bazant

March 11, 2005

1 Multivariate normal random walk

To calculate the probability density function of $\vec{X}_N$, we begin by finding the characteristic function of a single step. We know that

$$p_n(\vec{x}) = \frac{\exp\left(-\frac{1}{2}\vec{x} \cdot C_n^{-1}\vec{x}\right)}{(2\pi)^{d/2}\,|C_n|^{1/2}}$$

and therefore

$$\hat{p}_n(\vec{k}) = \frac{1}{(2\pi)^{d/2}\,|C_n|^{1/2}} \int e^{-i\vec{k}\cdot\vec{x}} e^{-\vec{x}\cdot C_n^{-1}\vec{x}/2} d^d\vec{x}$$

where the integral is taken over all space. To evaluate this, we first note that the matrix C_n^{-1} is by definition symmetric and positive definite. As described in the lectures, we can therefore find a symmetric square root B, such that $C_n^{-1} = BB$. We can therefore write

$$\hat{p}_n(\vec{k}) = \frac{|B|}{(2\pi)^{d/2}} \int e^{-i\vec{k}\cdot\vec{x}} e^{-(B\vec{x})\cdot B\vec{x}/2} d^d\vec{x},$$

which suggests making a substitution of the form $\vec{y} = B\vec{x}$. Thus $d^d\vec{y} = |B|\, d^d\vec{x}$ and

$$\begin{aligned}
\hat{p}_n(\vec{k}) &= \frac{1}{(2\pi)^{d/2}} \int e^{-i\vec{k}\cdot(B^{-1}\vec{y})} e^{-\vec{y}\cdot\vec{y}/2} d^d\vec{y} \\
&= \frac{1}{(2\pi)^{d/2}} \int e^{-(\vec{y}+iB^{-1}\vec{k})\cdot(\vec{y}+iB^{-1}\vec{k})/2 - \vec{k}\cdot(BB)^{-1}\vec{k}/2} d^d\vec{y} \\
&= e^{-\vec{k}\cdot C_n \vec{k}/2}.
\end{aligned}$$

Thus the characteristic function of the PDF after N steps have been taken is given by

$$\begin{aligned}
\hat{P}_N(\vec{k}) &= \prod_{n=1}^{N} \hat{p}_n(\vec{x}) \\
&= \prod_{n=1}^{N} e^{-\vec{k}\cdot C_n \vec{k}/2} \\
&= \exp\left(-\frac{1}{2}\vec{k} \cdot \left(\sum_{n=1}^{N} C_n\right) \vec{k}\right).
\end{aligned}$$

To invert this expression, we note that this is just the characteristic function of a multivariate gaussian with correlation matrix

$$C = \sum_{n=1}^{N} C_n$$

and therefore

$$P_N(\vec{x}) = \frac{\exp\left(-\frac{1}{2}\vec{x} \cdot \left(\sum_{n=1}^{N} C_n\right)^{-1} \vec{x}\right)}{(2\pi)^{d/2} \left|\sum_{n=1}^{N} C_n\right|^{1/2}}.$$

2 Student random walk

2.1 The characteristic function

The characteristic function is given by

$$\begin{aligned}
\hat{p}(k) &= \int_{-\infty}^{\infty} \frac{Ae^{-ikx}}{(1+x^2)^2} dx \\
&= \int_{-\infty}^{\infty} \frac{Ae^{-ikx}}{(x+i)^2(x-i)^2} dx,
\end{aligned}$$

and to evaluate this we make use of residue calculus, noting that the integrand has poles of order two at $x = \pm i$. For the case when $k < 0$, we can close the contour in the upper half plane, obtaining

$$\begin{aligned}
\hat{p}(k) &= 2\pi i \operatorname{Res}\left(\frac{Ae^{-ikx}}{(x+i)^2(x-i)^2}, x = i\right) \\
&= 2\pi i A \frac{d}{dx} \left.\frac{e^{-ikx}}{(x+i)^2}\right|_{x=i} \\
&= -2\pi i A \left.\frac{e^{-ikx}(ikx - k + 2)}{(x+i)^3}\right|_{x=i} \\
&= -2\pi i A \frac{2e^{-k}(1-k)}{(2i)^3} \\
&= \frac{A\pi}{2} e^{-k}(1-k).
\end{aligned}$$

Since $\hat{p}(0) = 1$ in order for the original probability density function to be normalized, we see that the tail amplitude is given by $A = 2/\pi$. If $k > 0$, then we see that making the substitution $y = -x$ in the original expression gives

$$\hat{p}(k) = \int_{-\infty}^{\infty} \frac{e^{-i(-k)y}}{(1+y^2)^2} dy$$

and therefore we know immediately that

$$\hat{p}(k) = e^{k}(1+k)$$

from which we know that the general solution is

$$\hat{p}(k) = e^{-|k|}(1+|k|).$$

The first few terms of the Taylor expansion are

$$\begin{aligned}\hat{p}(k) &= (1 - |k| + \frac{k^2}{2} - \ldots)(1 + |k|) \\ &= (1 - \frac{k^2}{2} + \ldots)\end{aligned}$$

and thus we see that $m_1 = 0$, $m_2 = 1$, from which it follows that $\sigma^2 = c_2 = m_2 = 1$.

2.2 Leading asymptotic behavior in the tail

We see that as $x \to \infty$,

$$p(x) \sim \frac{2}{\pi x^4}$$

and by additivity of power-law tails, we know that to leading order,

$$P_N(x) \sim \frac{2N}{\pi x^4}.$$

2.3 Two terms of the asymptotic expansion in the central region

To calculate the asymptotic expansion of the rescaled variable $Z_N = X_N/\sigma N$, we first find the logarithm of the characteristic function, which is given by

$$\begin{aligned}\psi(k) &= \log(e^{-|k|}(1 + |k|)) \\ &= -|k| + \log(1 + |k|) \\ &= -\frac{k^2}{2} + \frac{|k|^3}{3} - \frac{k^4}{4} + \frac{|k|^5}{5} - \frac{k^6}{6} + \ldots.\end{aligned}$$

We know that the probability density function of X_N is given by

$$P_N(x) = \int_{-\infty}^{\infty} e^{ikx} e^{N\psi(k)} \frac{dk}{2\pi}$$

and therefore the PDF of the rescaled variable, given by $\phi_N(z) = \sqrt{N} P_N(z\sqrt{N})$, is

$$\begin{aligned}\phi_N(z) &= \int_{-\infty}^{\infty} e^{iwz} e^{N\psi(w/\sqrt{N})} \frac{dw}{2\pi} \\ &\sim \int_{-\infty}^{\infty} e^{iwz} e^{-\frac{w^2}{2} + \frac{|w|^3}{3\sqrt{N}} - \frac{w^4}{4N} + \frac{|w|^5}{5N\sqrt{N}}} \frac{dw}{2\pi} \\ &\sim \int_{-\infty}^{\infty} e^{iwz - \frac{w^2}{2}} \left(1 + \frac{|w|^3}{3\sqrt{N}} + \frac{w^6}{2!N3^2} + \frac{|w|^9}{3!N^{3/2}3^3}\right)\left(1 - \frac{w^4}{4N}\right)\left(1 + \frac{|w|^5}{5N^{3/2}}\right) \frac{dw}{2\pi} \\ &\sim \int_{-\infty}^{\infty} e^{iwz - \frac{w^2}{2}} \left(1 + \frac{|w|^3}{3\sqrt{N}} + \frac{1}{N}\left(-\frac{w^4}{4} + \frac{w^6}{18}\right) + \frac{1}{N^{3/2}}\left(\frac{|w|^5}{5} - \frac{|w|^7}{12} + \frac{|w|^9}{162}\right)\right) \frac{dw}{2\pi}.\end{aligned}$$

We evaluate this expression term by term. The leading order behavior is given by

$$\begin{aligned}\phi(z) &= \int_{-\infty}^{\infty} e^{iwz - \frac{w^2}{2}} \frac{dw}{2\pi} \\ &= \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}.\end{aligned}$$

The first order correction is

$$\begin{aligned} g_1(z) &= \int_{-\infty}^{\infty} e^{iwz-\frac{w^2}{2}} \frac{|w|^3}{3} \frac{dw}{2\pi} \\ &= \int_0^{\infty} e^{-\frac{w^2}{2}} w^3 \cos(wz) \frac{dw}{3\pi} \end{aligned}$$

From the lectures, we know that integrals of this form can be expressed in terms of Dawson's integral

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

via the identity

$$\int_0^{\infty} w^m \cos(wz) e^{-\frac{w^2}{2}} dw = \frac{1}{(-2)^{(m-1)/2}} D^{(m)}\left(\frac{z}{\sqrt{2}}\right),$$

and thus we have

$$g_1(z) = -\frac{1}{6\pi} D^{(3)}\left(\frac{z}{\sqrt{2}}\right).$$

2.4 Four terms of the asymptotic expansion

The second order correction is given by

$$g_2(z) = \int_{-\infty}^{\infty} e^{iwz-\frac{w^2}{2}} \left(-\frac{w^4}{4} + \frac{w^6}{18}\right) \frac{dw}{2\pi}$$

and since this does not have any $|w|$ terms, we can evaluate using Hermite polynomials, via the identity

$$\int_{-\infty}^{\infty} e^{iwz} w^m e^{-\frac{w^2}{2}} dw = \left(-i\frac{\partial}{\partial z}\right)^m \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} = i^m H_m(z)\phi(z),$$

obtaining

$$g_2(z) = -\phi(z)\left(\frac{H_4(z)}{4} + \frac{H_6(z)}{18}\right).$$

For third order, we use Dawson's integral again, obtaining

$$\begin{aligned} g_3(z) &= \int_{-\infty}^{\infty} e^{iwz-\frac{w^2}{2}} \left(\frac{|w|^5}{5} - \frac{|w|^7}{12} + \frac{|w|^9}{162}\right) \frac{dw}{2\pi} \\ &= \int_0^{\infty} e^{-\frac{w^2}{2}} \cos(wz) \left(\frac{|w|^5}{5} - \frac{|w|^7}{12} + \frac{|w|^9}{162}\right) \frac{dw}{\pi} \\ &= \frac{1}{\pi}\left(\frac{D^{(5)}\left(\frac{z}{\sqrt{2}}\right)}{20} + \frac{D^{(7)}\left(\frac{z}{\sqrt{2}}\right)}{96} + \frac{D^{(9)}\left(\frac{z}{\sqrt{2}}\right)}{2592}\right) \end{aligned}$$

Hence, the first four terms of the asymptotic expansion are given by

$$\begin{aligned}
\phi_N(z) &\sim \phi(z) + \frac{g_1(z)}{\sqrt{N}} + \frac{g_2(z)}{N} + \frac{g_3(z)}{N^{3/2}} \\
&\sim \phi(z) - \frac{1}{6\pi\sqrt{N}} D^{(3)}\left(\frac{z}{\sqrt{2}}\right) - \frac{\phi(z)}{N}\left(\frac{H_4(z)}{4} + \frac{H_6(z)}{18}\right) \\
&\quad + \frac{1}{\pi N^{3/2}}\left(\frac{D^{(5)}\left(\frac{z}{\sqrt{2}}\right)}{20} + \frac{D^{(7)}\left(\frac{z}{\sqrt{2}}\right)}{96} + \frac{D^{(9)}\left(\frac{z}{\sqrt{2}}\right)}{2592}\right).
\end{aligned}$$

3 The largest step

3.1 Finding the CDF

Since each of the N events x_n is sampled independently, the CDF of $x_{(N)}$ is given by

$$\begin{aligned}
F_N(x) &= \mathbb{P}(x_{(N)} < x) \\
&= \mathbb{P}\left(\max_{1\le n\le N} x_n < x\right) \\
&= \mathbb{P}((x_1 < x) \cap (x_2 < x) \cap \ldots \cap (x_N < x)) \\
&= \mathbb{P}(x_1 < x)\mathbb{P}(x_2 < x)\ldots\mathbb{P}(x_N < x) \\
&= P(x)^N.
\end{aligned}$$

We could also get the same answer using the approach of Problem 1 on Problem Set 2, where wrote a general expression for the PDF of the nth order statistic, $x_{(n)}$:

$$f_{N,n}(x) = N\left(\begin{array}{c} N-1 \\ n-1 \end{array}\right) P(x)^{n-1}(1-P(x))^{N-n}p(x).$$

Evaluating this expression for the largest outcome, $n = N$, we have

$$F_N'(x) = f_{N,N}(x) = NP(x)^{N-1}p(x)$$

which gives $F_N(x) = P(x)^N$ upon integration.

3.2 The most probable value of $x_{(N)}$

The most probable value of $x_{(N)}$ is determined by finding the maximum of its probability density function. From above, we know that the PDF of $x_{(N)}$ is given by

$$f_N(x) = \frac{d}{dx}F_N(x) = \frac{d}{dx}P(x)^N = NP(x)^{N-1}p(x)$$

Thus

$$f_N'(x) = N(N-1)P(x)^{N-2}p(x)^2 + NP(x)^{N-1}p'(x)$$

and hence the most probable value of $x_{(N)}$ will be the solution to the equation

$$0 = (N-1)p(x)^2 + P(x)p'(x). \tag{1}$$

Since it is clear from its definition that $f_N(x) \to 0$ as $|x| \to \infty$, we expect to find at least one solution to this equation which corresponds to a maximum.

3.3 A power-law tail

If the range of x is unbounded, then clearly $x_{max}(N) \to \infty$ as $N \to \infty$. Therefore, to determine the leading order behavior of $x_{max}(N)$ as $N \to \infty$, we need only consider the tail of $p(x)$.

If $p(x) \sim A/x^{1+\alpha}$, then by integrating and differentiating,

$$\begin{aligned} P(x) &\sim 1 - \frac{A}{\alpha x^\alpha} \\ p'(x) &\sim -\frac{A(1+\alpha)}{x^{2+\alpha}} \end{aligned}$$

as $x \to \infty$. Substituting into equation 1 gives

$$\begin{aligned} 0 &\approx \frac{(N-1)A^2}{x^{2+\alpha}} + \left(\frac{A}{\alpha x^\alpha} - 1\right)\frac{A(1+\alpha)}{x^{2+\alpha}} \\ 0 &\approx (N-1)A + \frac{A(1+\alpha)}{\alpha} - (1+\alpha)x^\alpha \end{aligned}$$

and hence

$$\begin{aligned} x^\alpha &\sim \frac{A(N-\alpha^{-1})}{1+\alpha} \\ x &\sim \left(\frac{A(N-\alpha^{-1})}{1+\alpha}\right)^{1/\alpha} \\ &\sim \left(\frac{AN}{1+\alpha}\right)^{1/\alpha} \end{aligned}$$

Thus for large N we see that $x_{\max}(N) = O(N^{1/\alpha})$. For $\alpha > 2$ we know that the width of the central region is $O(\sqrt{N})$ and hence we see that for large N the largest step will be smaller.

3.4 Anomalous scaling

If $0 < \alpha < 2$ we see that square root scaling is inappropriate, since the size of the largest step will be larger than the width of the central region. Assuming that the largest step dominates the position, we therefore expect that for these cases, the anomalous scaling exponent of the width of the distribution will be $\nu(\alpha) = 1/\alpha > 1/2$. On Problem Set 1, we showed that the appropriate scaled variable for the Cauchy random walk was $Z_N = X_N/N$. For the Cauchy distribution $\alpha = 1$, and thus the largest step $x_{\max}(N) = O(N)$ exactly matches the scale of the width.

Random walks whose steps have infinite variance are called "Lévy flights". They exhibit "superdiffusion", since the width of probability distribution spreads must faster than $\sqrt{N}$. This calculation also shows, however, the such processes completely lack "self-averaging", since a single walker generally leaps to near its final position in a single large step, without much exploring the accessible region set by the PDF (as would an ensemble of many such walkers).