10.3 **Random walks and their generalizations**

(a) The probability of any particular walk with n steps to the right (and thus $N-n$ steps to the left) is $p^n q^{N-n}$ (this is because the successive steps are independent, so the probabilities multiply). Each such walk corresponds to a particular sequence of n rights and $N-n$ lefts. The number of such sequences is given by the binomial coefficient $\binom{N}{n}$, so the total probability of all the walks with n steps to the right is

$$P_N(n) = \binom{N}{n} p^n q^{N-n} \ .$$

(b) Let p and q be independent variables, not necessarily satisfying $p+q=1$. Then the sum

$$\sum_{n=0}^{N} \binom{N}{n} p^n q^{N-n} = (p+q)^N$$

is a function of the two variables, p and q. When $p(\partial/\partial p)$ acts on this sum, it puts an n in front of each term. So letting $p(\partial/\partial p)$ act k times, we get

$$\left(p\frac{\partial}{\partial p}\right)^k \left(\sum_{n=0}^{N} \binom{N}{n} p^n q^{N-n}\right) = \sum_{n=0}^{N} n^k \binom{N}{n} p^n q^{N-n} \ .$$

Now there is a subtle point. We have let p and q be independent variables, so the last sum in this equation contains the normalized binomial distribution only if we evaluate the derivative at $p+q=1$. Thus we have

$$\left(p\frac{\partial}{\partial p}\right)^k \left(\sum_{n=0}^{N} \binom{N}{n} p^n q^{N-n}\right)\Bigg|_{p+q=1} = \sum_{n=0}^{N} n^k P_N(n) = \overline{n^k} \ .$$

By the binomial theorem, the sum is $(p+q)^N$, so we end up with

$$\overline{n^k} = \left(p\frac{\partial}{\partial p}\right)^k (p+q)^N \Bigg|_{p+q=1} \ .$$

Now let's do the first two moments:

$$p\frac{\partial}{\partial p}(p+q)^N = Np(p+q)^{N-1} \quad \Longrightarrow \quad \bar{n} = p\frac{\partial}{\partial p}(p+q)^N\Bigg|_{p+q=1} = Np \ ,$$

$$\left(p\frac{\partial}{\partial p}\right)^2 (p+q)^N = p\frac{\partial}{\partial p}Np(p+q)^{N-1} = Np(p+q)^{N-1} + N(N-1)p^2(p+q)^{N-2}$$

$$\Longrightarrow \quad \overline{n^2} = \left(p\frac{\partial}{\partial p}\right)^2 (p+q)^N\Bigg|_{p+q=1} = Np + N(N-1)p^2 = N^2p^2 + Np(1-p) \ .$$

Thus the variance is

$$(\Delta n)^2 = \overline{n^2} - \bar{n}^2 = Npq \ .$$

(c) It's a good idea to write the probability in terms of either dt or N, so we can clearly see how the limit works. Let's choose N:

$$P_T(n) = \frac{N!}{n!(N-n)!}(Rdt)^n(1-Rdt)^{N-n} = \frac{N!}{n!(N-n)!}(RT/N)^n(1-RT/N)^{N-n} \ .$$

Now we do a few more manipulations meant to take advantage of the fact that the distribution is significantly nonzero only when $n \ll N$:

$$\begin{aligned}
P_T(n) &= \frac{N(N-1)\cdots(N-n+1)}{n!}\left(\frac{RT}{N}\right)^n\left(1-\frac{RT}{N}\right)^{N-n} \\
&= \frac{N^n}{n!}\left(\frac{RT}{N}\right)^n\left(1-\frac{RT}{N}\right)^N\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)\left(1-\frac{RT}{N}\right)^{-n} \\
&= \frac{(RT)^n}{n!}\left(1-\frac{RT}{N}\right)^N \times \left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)\left(1-\frac{RT}{N}\right)^{-n} .
\end{aligned} \tag{1}$$

All the terms after the $\times$ limit to 1 as $N \to \infty$ (with n held fixed).

You should notice that the limit we are using here for the choose coefficient, i.e.,

$$\frac{N!}{n!(N-n)!} = \frac{N^n}{n!}\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{n-1}{N}\right) \underset{N\to\infty}{\longrightarrow} \frac{N^n}{n!} ,$$

is the same one we used in considering the number of overall states of bosons and fermions when the number of single-particle states far exceeds the number of particles. And the limit here means the same thing. The choose coefficient appears in the binomial distribution because it counts the number of different ways of getting n raindrops in N of the time intervals dt. But when $N \gg n$, there is very little chance of getting two raindrops in the same interval of length $dt = T/N$, so one counts the number of different ways of getting n raindrops by saying each raindrop can occupy one of N intervals, giving N^n total configurations, and then dividing by $n!$ to account for the fact that permutations of the raindrops are the same configuration.

The only remaining question about Eq. (1) is what happens to the term $(1 - RT/N)^N$. It is a standard result of the calculus that this limits to e^{-RT}. The official way to see why this is true is to consider the logarithm:

$$\ln\left(1-\frac{RT}{N}\right)^N = N\ln\left(1-\frac{RT}{N}\right) \underset{N\to\infty}{\longrightarrow} -RT .$$

The limit follows from applying l'Hopital's rule in the following way:

$$\lim_{x\to 0}\frac{\ln(1-RTx)}{x} = \lim_{x\to 0}\frac{d\ln(1-RTx)/dx}{dx/dx} = \lim_{x\to 0}\frac{-RT}{1-RTx} = -RT .$$

The result is

$$\lim_{N\to\infty}\left(1-\frac{RT}{N}\right)^N = e^{-RT} .$$

An exactly equivalent, somewhat more straightforward way to proceed is to notice that when N gets very large, $1 - RT/N$ is the same as $e^{-RT/N}$, so the limit is

$$\lim_{N\to\infty}\left(1-\frac{RT}{N}\right)^N = \lim_{N\to\infty}\left(e^{-RT/N}\right)^N = e^{-RT} .$$

Our final result, after taking the limit $N \to \infty$, is the Poisson distribution:

$$P_T(n) = e^{-RT}\frac{(RT)^n}{n!} .$$

It is easy to calculate the mean value,

$$\bar{n} = \sum_{n=0}^{\infty} nP_T(n) = e^{-RT}\sum_{n=0}^{\infty} n\frac{(RT)^n}{n!} = RTe^{-RT}\sum_{n=1}^{\infty}\frac{(RT)^{n-1}}{(n-1)!} = RTe^{-RT}\sum_{m=0}^{\infty}\frac{(RT)^m}{m!} = RT ,$$

where we realize that the $n = 0$ term doesn't contribute to the sum and then change summing index to $m = n-1$. This result, which you could guess from the start, allows us to write the Poisson distribution in its standard form,

$$P_T(n) = e^{-\bar{n}} \frac{\bar{n}^n}{n!} \ .$$

If we wanted to calculate any moment, we could get it from

$$\overline{n^k} = \sum_{n=0}^{\infty} n^k P_T(n) = e^{-\bar{n}} \sum_{n=0}^{\infty} n^k \frac{\bar{n}^n}{n!} = e^{-\bar{n}} \left(\bar{n}\frac{\partial}{\partial \bar{n}}\right)^k \sum_{n=0}^{\infty} \frac{\bar{n}^n}{n!} = e^{-\bar{n}} \left(\bar{n}\frac{\partial}{\partial \bar{n}}\right)^k e^{\bar{n}} \ .$$

Let's do the second moment:

$$\overline{n^2} = e^{-\bar{n}} \left(\bar{n}\frac{\partial}{\partial \bar{n}}\right)^2 e^{\bar{n}} = e^{-\bar{n}} \left(\bar{n}\frac{\partial}{\partial \bar{n}}\right) \left(\bar{n}e^{\bar{n}}\right) = e^{-\bar{n}} \left(\bar{n}e^{\bar{n}} + \bar{n}^2 e^{\bar{n}}\right) = \bar{n}^2 + \bar{n} \ ,$$

Thus the variance is

$$\Delta n^2 = \overline{n^2} - \bar{n}^2 = \bar{n} \ .$$

This is another of the direct expressions of the $1/\sqrt{N}$ principle of statistics: the uncertainty $\Delta n = \sqrt{\bar{n}}$ in the number of raindrops falling on an umbrella is the square root of the mean number. As $\bar{n}$ gets big, you can be quite certain that you will not be so lucky as to stand in a place where it just so happens you won't get drenched.

(d) We start with the binomial distribution,

$$P_N(n) = \frac{N!}{n!(N-n)!} p^n q^{N-n} \ .$$

Let's take the natural logarithm of the binomial coefficient and use Stirling's approximation:

$$\begin{aligned}
\ln \frac{N!}{n!(N-n)!} &= \ln N! - \ln n! - \ln(N-n)! \\
&\sim N \ln N - N + \frac{1}{2}\ln(2\pi N) \\
&\qquad - n\ln n + n - \frac{1}{2}\ln(2\pi n) - (N-n)\ln(N-n) + N - n - \frac{1}{2}\ln[2\pi(N-n)]) \\
&= N\ln N - N\ln(N-n) - n\ln n + n\ln(N-n) + \frac{1}{2}\ln\left(\frac{N}{n(N-n)}\right) - \frac{1}{2}\ln 2\pi \ .
\end{aligned}$$

This gives us

$$\ln P_N(n) \sim N\ln N - N\ln(N-n) + N\ln q - n\ln n + n\ln(N-n) + n\ln\frac{p}{q} - \frac{1}{2}\ln[n(N-n)] - \frac{1}{2}\ln(2\pi/N) \ .$$

The first six terms all grow at least linearly with as N gets large, whereas the second-to-last term grows logarithmically with N and the last term is a constant. We can neglect the last two terms in calculating the derivatives of $\ln p_T(n)$, but we will keep them around for the ultimate expansion for reasons that become apparent below.

We now treat n as a continuous variable and take a derivative with respect to n:

$$\frac{d\ln P_N(n)}{dn} \sim \frac{N}{N-n} - \ln n - 1 + \ln(N-n) - \frac{n}{N-n} + \ln\frac{p}{q} = \ln\left(\frac{N-n}{n}\frac{p}{q}\right) \ .$$

We want to expand about the peak of $\ln P_T(n)$, so we need to know where this derivative is zero, and that is when the argument of the logarithm is 1, so it occurs at $n = Np$. Now let's get the second derivative:

$$\frac{d^2\ln P_N(n)}{dn^2} = \frac{d}{dn}[\ln(N-n) - \ln n] \sim -\frac{1}{N-n} - \frac{1}{n} = -\frac{N}{n(N-n)} \ .$$

Evaluated at the peak, the second derivative is

$$\left.\frac{d^2 \ln P_N(n)}{dn^2}\right|_{n=Np} = -\frac{1}{Npq}$$

When evaluated at the peak, function itself is

$$\ln P_N(n = Np) \sim N \ln N - N \ln Nq + N \ln q - Np \ln Np + Np \ln Nq + Np \ln \frac{p}{q} - \frac{1}{2} \ln(2\pi Npq) = -\frac{1}{2} \ln(2\pi Npq) \ .$$

All the large terms cancel in this expression, leaving only the term that we neglected in taking derivatives. This might be thought remarkable, but is not when you realize that the probabilities can't be bigger than 1, so there is guaranteed to be a lot of cancellation. It makes clear why we have to keep the smallest term, because it is the only one that survives the cancellation, even though it has a negligible impact on where the peak of the function is.

Putting all this together, we have our expansion to quadratic order in $n - Np$:

$$\ln p_N(n) \sim \ln p_N(n = Np) + \frac{1}{2} \left.\frac{d^2 \ln p_N(n)}{dn^2}\right|_{n=Np} (n - Np)^2 = -\frac{1}{2} \ln(2\pi Npq) - \frac{1}{2} \frac{(n - Np)^2}{Npq} \ .$$

Our final result for the binomial distribution in the large N limit is

$$p_N(n) \sim \frac{1}{\sqrt{2\pi Npq}} \exp\left(-\frac{1}{2} \frac{(n - Np)^2}{Npq} \right) .$$

You should learn to recognize this form as a normalized Gaussian distribution with mean $\bar{n} = Np$ and variance $(\Delta n)^2 = Npq$.