On the duality of distributions

If a volume V is uniformly filled with particles, then the probability of finding a particle in a smaller volume v $p = \frac{v}{V}$. If the total number of particles is N, the probability of finding exactly m particles in our small volume v depends on p and the relative magnitudes of N and m, yielding the familiar binomial distribution:

$$P(m) = \frac{N!p^m(1-p)^{N-m}}{(N-m)!m!}$$

This is general – true for arbitrary values of p, N and $m < N$.

However, if we make some assumptions about some of these parameters, which are often in fact the case, we can simplify this rather frightening expression. In particular, p is often small (i.e., 0.01 chance of a polymerase falling off the strand, compared to 0.5 chance of getting heads in a coinflip), and the "expected value" m is very small compared to the total particle population N (i.e., a radioactive molecule emitting a single particle of the millions it will emit over its half-life). So, for

$$\frac{N!}{(N-m)!} = N(N-1)(N-2)\cdots(N-m+1),$$

if N is huge and m tiny, there are not going to be very many terms in this series, and they're all going to be relatively close to N. (Take $N = 1,000,000$ and $m = 3$, then the series is $1,000,000 \times 999,999 \times 999,998$.) So this series is just N^m, and

$$\frac{N!p^m}{(N-m)!} \approx N^m p^m = (Np)^m.$$

Moreover, with our small probability p, the average number of molecules $\bar{m}$ that we find in our small volume is going to be the same Np, so this giant term has in fact simplified to $\bar{m}^m$. At this point our distribution looks like this:

$$P(m) = \frac{\bar{m}^m(1-p)^{N-m}}{m!}.$$

e is a pretty "natural" number (ha, ha ha), and it appears a lot in probabilistic physics (Boltzmann factors!), so let's think about it. $e \equiv \lim_{x\to 0}(1+x)^{1/x}$, which looks suspiciously similar to the remaining scary term in the expression. If we rewrite that term:

$$(1-p)^{N-m} = (1-p)^{(N-m)p\frac{1}{p}} = (1-p)^{Np\frac{1}{p}},$$

where I've taken $N - m \approx N$ since $m << N$, we have a similar limit (p is small!), with a $(1-x)$ term instead of $(1+x)$:

$$X = \lim_{x \to 0} (1-x)^{a/x},$$

where $x = p$ and $a = Np$. Taking the natural logarithm of each term,

$$\ln X = \lim_{x \to 0} \frac{a \ln(1-x)}{x}.$$

Since both the numerator and denominator of the fraction in the limit evaluate to zero, we can apply l'Hôpital's rule:

$$\ln X = \lim_{x \to 0} -\frac{a}{1-x}$$

and we see the limit evaluates to $-a$. (I realize the implicit use of e in calculating this might not be strictly kosher, but I'm lazy and ignorant; if a mathematician has a better way of doing this I cede the court in advance.) So $\ln X = -a$, and X, the value of the original limit, is just $e^{-a} = e^{-Np} = e^{-\bar{m}}$.

Inserting this into the partially-simplified probability distribution, we get:

$$P(m) = \frac{\bar{m}^m e^{-\bar{m}}}{m!}$$

which is the Poisson distribution. It is an *approximation* of the binomial distribution under certain conditions, stated above. In particular, as $\bar{m}$ becomes large, the distribution looks increasingly like the binomial. $\bar{m}$ can be the global concentration of particles in a solution, the average number of primer sites per length of DNA (or the exact number of sites in λ-phage DNA), or whatever other average value.

References

Meyer B Jackson, *Molecular and Cellular Biophysics*
John R Taylor, *An Introduction to Error Analysis*

No Microsoft software was used in the production of this document.