Math 302.102 Fall 2010
The Poisson Approximation to the Binomial Distribution

Suppose that $X \sim \text{Bin}(n,p)$ so that $\mathbb{E}(X) = np$ and $\text{Var}(X) = np(1-p)$. We know from the central limit theorem that if n is sufficiently large, then X is approximately normal with mean np and variance $np(1-p)$. The continuity correction makes this precise; that is,

$$\mathbf{P}\{a \le X \le b\} = \mathbf{P}\left\{a - \frac{1}{2} \le X \le b + \frac{1}{2}\right\} = \mathbf{P}\left\{\frac{a-\frac{1}{2}-np}{\sqrt{np(1-p)}} \le \frac{X-np}{\sqrt{np(1-p)}} \le \frac{b+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right\}$$
$$\approx \mathbf{P}\left\{\frac{a-\frac{1}{2}-np}{\sqrt{np(1-p)}} \le Z \le \frac{b+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right\}$$

where $Z \sim \mathcal{N}(0,1)$. This means that the probability in question can be determined using a table of normal probabilities. That is,

$$\mathbf{P}\left\{\frac{a-\frac{1}{2}-np}{\sqrt{np(1-p)}} \le Z \le \frac{b+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right\} = \Phi\left(\frac{b+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)$$

where Φ denotes the standard normal distribution function

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2}\,\mathrm{d}x.$$

However, if p is near 0 or 1, then n needs to be extremely large in order for this approximation to be good. Think about this. If p is near 0, say $p = 0.000001$, then this says that the probability of success is 1 in a million. That means that, on average, if we examined 1 billion trials (say, roughly, the population of China or India), then we would observe only 1000 successes. Although 1000 successes might seems like a lot, it is insignificant in a sample of size 1 billion. This example might seem extreme, but it is meant to point out that care needs to be taken with rare events. The general rule of thumb that is "taught" is that one needs both $np > 10$ and $n(1-p) > 10$ in order for the normal approximation to be good.

If the normal approximation does not apply, then one can use a Poisson approximation. Here it is. We know

$$\mathbf{P}\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}.$$

We can make this look like a Poisson by letting $\lambda = np$ so that

$$\begin{aligned}
\frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} &= \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} \cdot \left(1 - \frac{\lambda}{n}\right)^{n} \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \cdot \frac{n!}{n^k (n-k)!} \\
&= \frac{\lambda^k}{k!} \cdot \left(1 - \frac{\lambda}{n}\right)^{n} \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \cdot \left[\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right) \cdots \left(\frac{n-k+1}{n}\right)\right].
\end{aligned}$$

If we now take $n \to \infty$ in such a way that $\lambda = np$ remains constant, then since k is fixed, we find

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}, \quad \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k} = 1$$

and

$$\lim_{n\to\infty}\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-k+1}{n}\right) = \lim_{n\to\infty}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right) = 1$$

so that

$$\lim_{n\to\infty}\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}.$$

Hence, if $X \sim \text{Bin}(n,p)$ and at least one of np and $n(1-p)$ is small so that the normal approximation does not apply, then

$$\mathbf{P}\{X = k\} \approx \mathbf{P}\{Y = k\}$$

where $Y \sim \text{Poisson}(np)$. Thus,

$$\mathbf{P}\{X = k\} \approx \frac{(np)^k}{k!}e^{-np}.$$

Example. Suppose that 25 families are surveyed at random. Suppose further that each family surveyed has 4 children. What is the probability that exactly 2 of these families have 4 children all of whom are boys? You may assume that children are equally likely to be either a boy or a girl.

Solution. Let X denote the number of families with 4 boys so that

$$X \sim \text{Bin}(n = 25, p = 1/16).$$

We know exactly that

$$\mathbf{P}\{X = 2\} = \binom{25}{2}(1/16)^2(15/16)^{23} \doteq 0.2836.$$

For comparison, the Poisson approximation with $\lambda = np = 25/16$ gives

$$\mathbf{P}\{X = 2\} \approx \frac{(25/16)^2}{2!}e^{-25/16} \doteq 0.2556.$$