

Approximations to the Binomial Distribution

The [Binomial distribution](#) is the most fundamental distribution in probability theory. $\text{Binomial}(p, n)$ models the number of successes s in n trials, where each trial is independent of others and has the same probability of success p . The probability of failure $(1-p)$ is often written as q to make the equations a bit neater.

Normal approximation to the Binomial

When n is large, and p is neither very small nor very large, the following approximation works very well:

$$\text{Binomial}(p, n) \approx \text{Normal}(np, (npq))$$

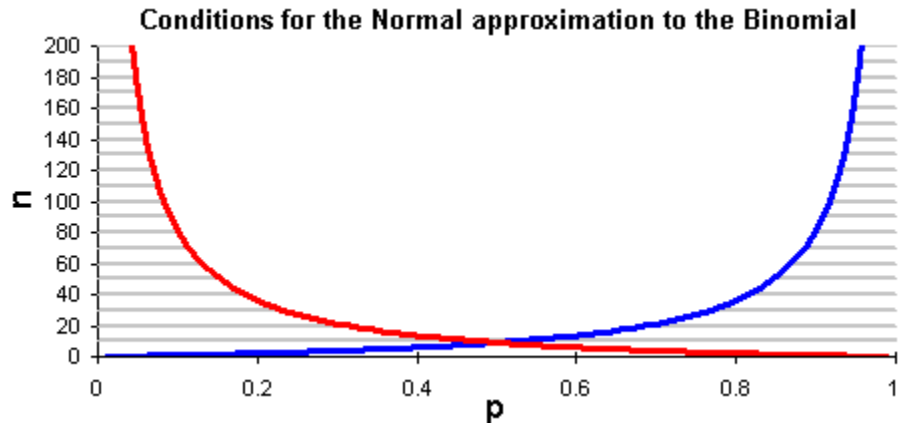
The mean and standard deviation of a Binomial distribution are np and (npq) respectively, so this approximation is quite easy to accept. It also fits nicely with [Central Limit Theorem](#), because the $\text{Binomial}(p, n)$ distribution can be thought of as the sum of n independent $\text{Binomial}(p, 1)$ distributions, each with mean p and standard deviation (pq) .

The difficulty lies in knowing whether, for a specific problem, the values for n and p fall within the bounds for which the Normal distribution is a good approximation. A $\text{Binomial}(0.5, 1)$ is symmetric, so we can intuitively guess that one needs a fairly low value for n for the Normal approximation to be reasonable when $p = 0.5$. On the other hand, a $\text{Binomial}(0.95, 1)$ is very highly skewed and we would reasonably expect that n would need to be large for the Normal approximation to work for such an extreme value of p . An easy way to judge this is to think about the range that a Normal distribution takes: almost all of its probability is contained within a range ± 3 standard deviations from the mean. Now, we know that the Binomial distribution is contained within the range $[0, n]$. It would therefore be reasonable to say that the Normal distribution is a good approximation if its tails stay well within these limits, i.e.:

$$np - 3\sqrt{np(1-p)} > 0 \text{ and } np + 3\sqrt{np(1-p)} < n$$

which simplify to: $n > \frac{9p}{(1-p)}$ and $n > \frac{9(1-p)}{p}$

A more stringent condition (using four instead of three standard deviations for the Normal) would be to use 4 and 16 instead of 3 and 9 in the above equations. The graph on the right shows how these two conditions work together symmetrically to show the $\{p, n\}$ combinations which will work well. [Decker and Fitzgibbon \(1991\)](#) advise using the Normal approximation when $n0.31p > 0.47$, which is discussed more below. At larger values of n , which is when one might wish to use an approximation, their rule of thumb is somewhat more conservative than that presented here even using a range ± 3 standard deviations from the mean.



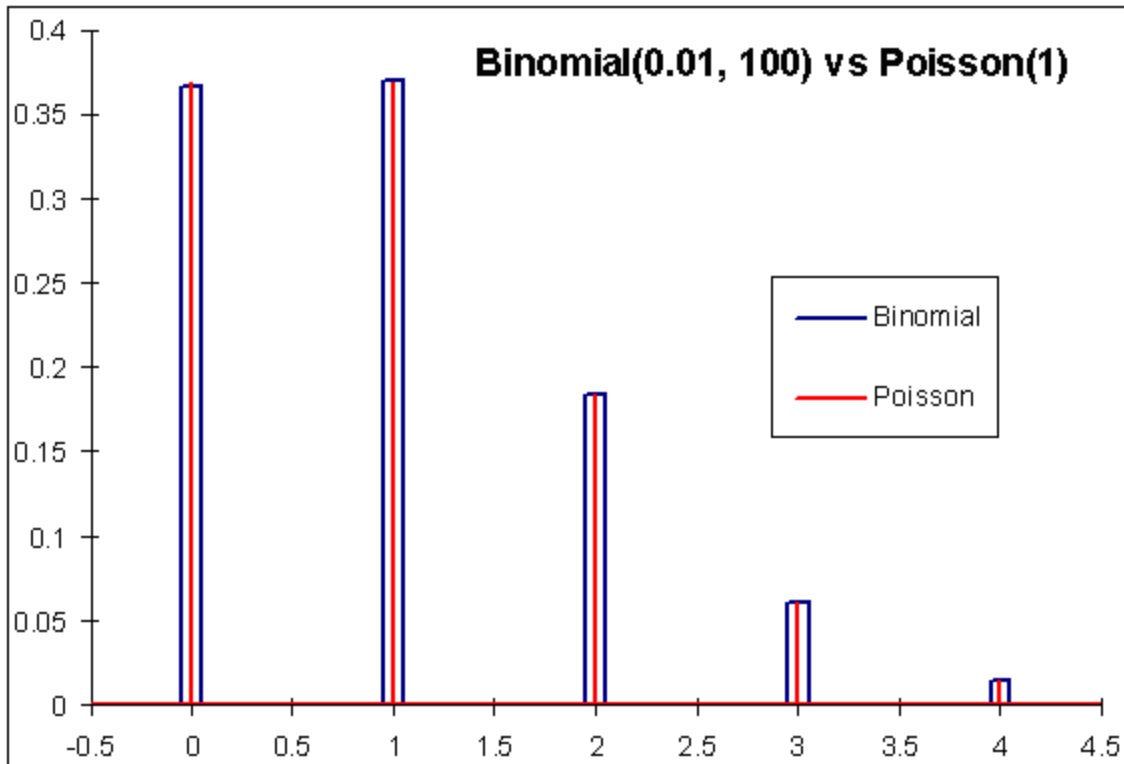
The Normal distribution is continuous while the Binomial distribution is discrete, and this approximation leads to an additional error that can be avoided. Rather than use $f(x)$ from the Normal distribution to give the Binomial distribution probability $p(x)$, it is more accurate to use $F(x+0.5) - F(x-0.5)$ where $F(x)$ is the cumulative distribution function for the Normal distribution. If one is simulating the distribution with Normal, the equivalent is to use a $\text{ROUND}(\dots, 0)$ function around the Normal distribution.

Poisson approximation to the Binomial

The probability mass function for the Poisson distribution can be derived from the Binomial probability mass function by making n extremely large while p becomes very small, but within the constraint that np remains finite. Thus, the following approximation can be made to the Binomial:

$$\text{Binomial}(p, n) \approx \text{Poisson}(np) \quad \text{when } n \rightarrow \infty, p \rightarrow 0, np \text{ remains finite}$$

In fact the Poisson approximation works very well for relatively small values of n and large values of p . For example, a Binomial(1%, 100) is very well approximated by a Poisson(1):



The Poisson approximation tends to overestimate the tail probabilities at both ends of the distribution. Decker and Fitzgibbon (1991) advise using this approximation when $n^{0.31}p < 0.47$ and the Normal approximation otherwise.

We can use a [Poisson](#) approximation to the Binomial when p is close to 1, i.e. as $(1-p) \ll 0$ by simply reflecting the formula. In this case, $\text{Binomial}(p, n) \approx n$ - $\text{Poisson}(np)$ and the Decker and Fitzgibbon condition is then $n^{0.31}(1-p) < 0.47$.
