Approximations to the Negative Binomial distribution

The Negative Binomial distribution NegBinomial(p,s) models the total number of trials (n trials = s *successes plus* n-$s$failures *)* it takes to achieve s successes, where each trial has the same probability of success p.

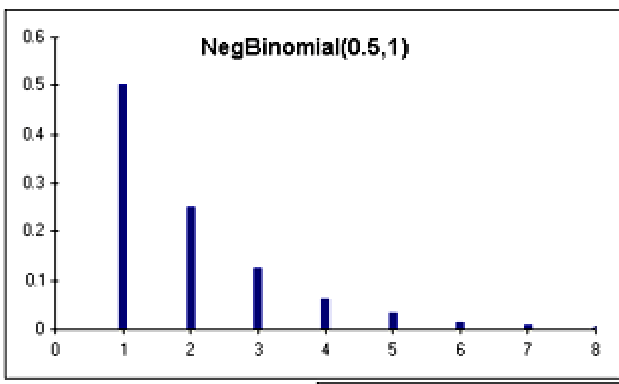
Normal approximation to the Negative Binomial

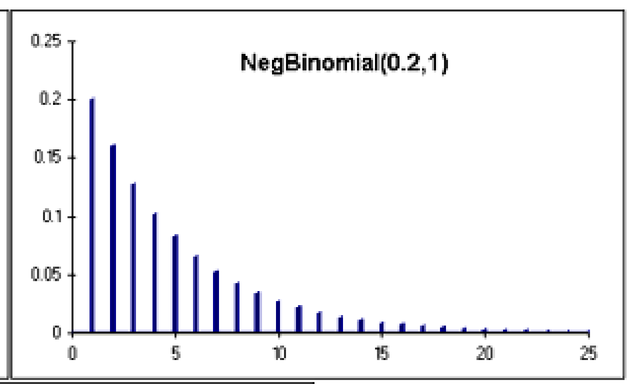
When the number of successes s required is large, and p is neither very small nor very large, the following approximation works pretty well:

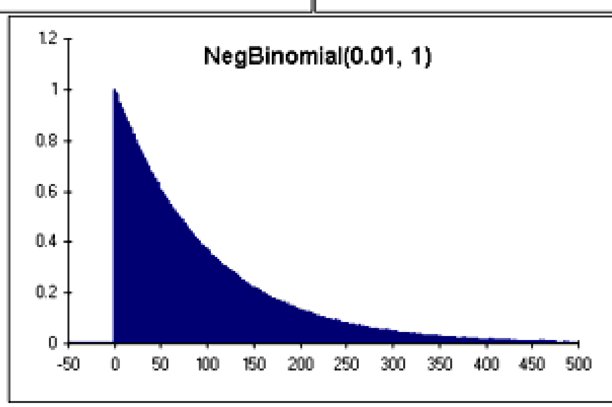
NegBinomial(p, s) » Normal	$\left(\frac{s}{p}, \sqrt{s(1-p)/p^2}\right)$

The approximation can be justified via Central Limit Theorem, because the NegBinomial(p,s) distribution can be thought of as the sum of s independent NegBinomial(p, 1) distributions, each with mean $\frac{1}{p}$ and standard deviation $\sqrt{\frac{1-p}{p^2}}$.

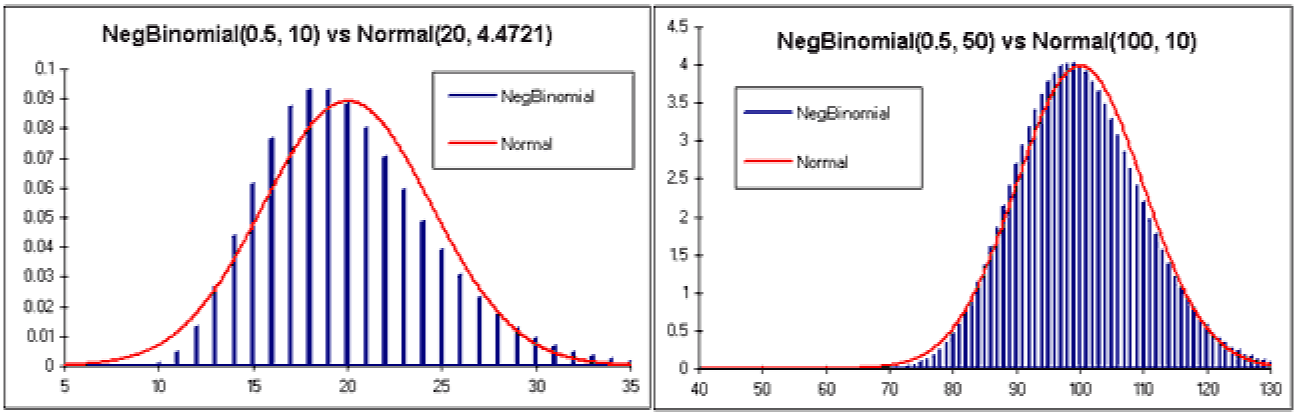
The difficulty lies in knowing whether, for a specific problem, the values for s and p fall within the bounds for which the Normal distribution is a good approximation. The smaller the value of p, the longer the tail of a NegBinomial(p,1) distribution:



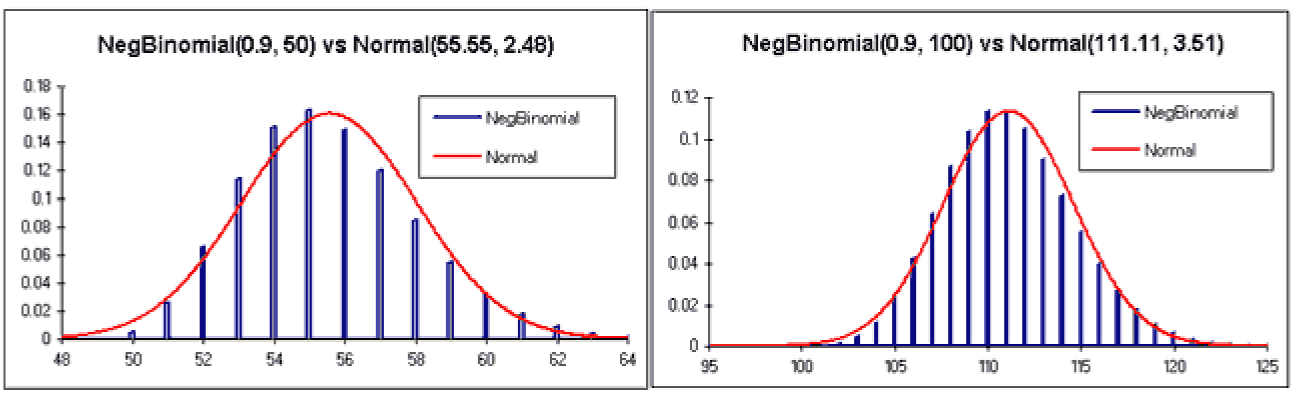




As *p* gets very small, the NegBinomail(*p*,1) becomes an Exponential distribution (see below), and so we can use a Gamma approximation to the NegBinomial instead of a Normal. On the other hand, as *p* is large, so the NegBinomial(*p*,1) distribution gets more skewed, so *s* would need to be much larger for a Normal approximation (which has to overcome this skewness) to be appropriate:



NegBinomial(0.5,s) distributions and their corresponding Normal distribution approximations



NegBinomial(0.9,*s*) distributions and their corresponding Normal distribution approximations, showing that when *p* is large, *s* needs to be higher for the Normal approximation to work well.

Gamma approximation to the Negative Binomial

The Poisson process can be derived from the Binomial process by making *n* extremely large while *p* becomes very small, but within the constraint that *np* remains finite. In a Poisson process, the Gamma(0,*b*,*a*) distribution models the 'time' until observing *a* events where *b* is the mean time between events. The NegBinomial distribution is the binomial equivalent, modeling the total number of trials to achieve *s* successes where [(1/*p*)-1] is the mean number of failures per success. The NegBinomial in Crystal Ball includes the s successes which in terms of a Poisson process are not included in the waiting time because each event is assumed to be instantaneous. To make the two approaches more comparable, we subtract the (non-random) number of successes from the NegBinomial(p,s) distribution to obtain the number of failures only (i.e. shift the distribution s to the left). The remaining distribution models the number of failures, with mean (1/p-1) failures for each success. Then, we can make the following approximation:

$$\text{NegBinomial}(p,s) - s \approx \text{Gamma}(0,1/p-1,s) \qquad \text{when} \quad p \to 0$$

Or equivalently, using the shift parameter for the Gamma distribution:

NegBinomial(p,s) » Gamma(s,1/p,s) when *p ® 0*

For *s* = 1, we also have the special case:

Geometric(*p*) -1 » Exponential(*p/(1-p)*) when *p ® 0*

When the Exponential distribution is a good approximation to the "Geometric(p) - 1" (p<0.05 is usually good, see below), the Gamma is a good approximation to the NegBinomial.