Geometric

Linsey Moddelmog and Paul Johnson

February 9, 2006

1 Introduction

The geometric distribution is used to represent the probabilities of discrete events. It is very similar to the negative binomial distribution, in that it is measuring the probability of success or failures of an event; and therefore uses interval or discrete data. Additionally, the data represent independent events, or events that are not influenced by other events; think coin flips or rolling of the dice.

There are two ways of thinking about the Geometric distribution and hence two different ways in which it can be operationalized:

"If the probability of success on each trial is p, then the probability that n trials are needed to get one success is"

$$Geometric Version\,1 : Prob(x = n) = (1 - p)^{n-1}p$$

for n=1,2,3, [non-negative integers]...

Equivalently the probability that there are m failures before the first success is"[1]

$$Geometric Version\,2 : Prob(y = m) = (1 - p)^{m}p$$

The difference between the two formulations is that one describes a success on the n'th event while the other describes m successive failures before the n'th event on which success occurs. The only real difference between the two models is that the lowest possible value of x that can be observed is 1, while the lowest value of y that can be observed is 0.

The second approach is used in R.

2 Moments

The expected value of a geometrically distributed variable with probability of success p is

$$E(x) = \frac{1}{p}$$

[1] http://en.wikipedia.org/wiki/Geometric_distribution

The variance is

$$Var(x) = \sigma^2 = \frac{1-p}{p^2}$$

The skewness is

$$Skewness(x) = \frac{2-p}{\sqrt{1-p}}$$

The kurtosis excess is

$$Kurtosis(x) = \frac{p^2 - 6p + 6}{1-p}$$

3 Illustrations

In Figure 1, the Geometric probability mass function is illustrated for probabilities of success ranging from 0.1 to 0.9. Note that when the probability of success on an individual trial is small, the probability mass is spread more evenly across the interval being considered (which ranges from 0 to 19). On the contrary, when the probability of success is high, then it is not very likely that the process will continue for more than a few steps before it is terminated by a single success. The graphs for the high probability models are a bit deceptive because they make it appear as though the process continuing for more than a few steps drops to 0.0 and never changes. That is an optical illusion. The value is slowly droping with each increase.

Figure 1: Geometric Distribution