Handout 17: Distribution Applications

In reviewing for the exam, it seemed like a good idea to go back and (re)describe the basic applications of the probability distributions we have seen.

Binomial: Numerous applications of counting the number of events that occur when the total number of events that could occur is fixed. Can often be approximated by the Poisson (small probabilities with large sample size) or by the normal distribution (large n and non-extreme p).

Geometric: Used in engineering to estimate the time until a unit fails.

Poisson: Number of events occurring in a given period of time or over a fixed length of space. Also used to approximate the Binomial when we can estimate its mean better than n or p in particular.

Negative Binomial: Its flexibility makes it a good choice for modeling discrete variables that do not fit into one of the prior distributions. For example, the length of stay of a patient in the hospital.

Beta: Can be used, possibly by modifying by an additive and multiplicative scale, to model any continuous outcome that occurs between two fixed values. One common example is the time to complete an task, such as an exam.

Gamma: Often used in applications where the outcome is a continuous, positive, unbounded value. For example, the size of an insurance claim or the amount of rain that falls in a given storm.

Normal: Numerous applications, include those that involve the combination of many independent processes, such as height and weight, which depend on the interaction of many genetic and environmental effects. Can also be used to measure any continuous outcome that should be (approximately) symmetric around a mean value. It is also used to approximate the binomial distribution.

Pareto: Models continuous outcomes which are bounded from below, often by zero, where it is assumed that the density is monotonically decreasing. Examples include the size of cities or the distribution of wealth.

In real life, no observed values are truly bounded, so how do we decide between the latter four? Usually, we use the Beta if we believe there will be a reasonable amount of density near the endpoints. The normal is used if the distribution is known to be nearly symmetric, and the Pareto if it should be monotonic. The gamma is used in cases where one side of the distribution may be "longer" than the other (such as with money).