Chapter 5 - Discrete Probability Distributions

Previously

1. The events described by these distributions are discrete (countable).
2. $P(x = k) > 0$ for any k in the set of discrete events.
3. Shape of the function can best be described as a jump function and best represented graphically using a barchart.
4. $P(x \leq k) = \sum_{i=1}^{k} P(x = i)$
5. $P(x \geq k) = \sum_{i=k}^{N} P(x = i)$
6. $P(a \leq x \leq b) = P(x \leq b) - P(x < a)$
7. $\sum_{i=1}^{N} P(x = i) = 1.0$

Theoretical Probability Distributions

1. Provides probability values for "common" types of distributions, unimodal symmetric, unimodal distributions skewed to the left, etc..
2. Well-proven, advanced statistical tools.
3. Same set of tools can be used with different data distributions.

Researcher's options

1. Create a new set of statistical tools for each data distribution, accept the condition that the tools are based on data with systematic errors.

2. Determine if data distribution can be represented by one of the theoretical probability distributions.

Probability distribution (density) function Cumulative probability distribution (density) function

1. Probability density function (pdf) - probability that x is a given value - similar to the information obtain from a relative frequency value.

2. Cumulative density function(cdf) - probability that x is less than or greater than a value - similar to a cumulative relative frequency.

Discrete distributions described in this chapter

1. Binomial

2. Poisson

3. Hypergeometric

How are they similar

1. All work with problems based on complementary events (A, B)
   A is treated as a success.
   B a failure.

How to they differ

1. Relative sample size
2. Knowledge of probability of success.
3. Knowledge of average occurrence of success.

Binomial Probability Distribution

1. States of the world consist of two states (Failure/Success)
2. Events are repeated for n trials.
3. Trials are independent, that is, what happened in the past does not affect the outcome of the present or future events.
4. For each trial, there is a probability of success (p) and failure(q). These probabilities are constant across trials.
5. The question that a researcher may have is, what is the probability that, after n trials, a person has k successes.

Function-pdf

$$P(x = k) = C_k^n p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

Function-cdf

$$P(x \leq j) = \sum_{i=0}^{j} P(x = i)$$

Use of a table of values
Examples - Question 5.8
Mathematica Example
Mean, Variance, Standard Deviation

$$\mu = np$$
$$\sigma^2 = nqp$$
$$\sigma = \sqrt{nqp}$$

Restriction - relative size of sample
Size is 5% of the population size or less.
Poisson Distribution

1. Suppose one knows that, in a given time period, it is expected that a given event will occur μ times (Another way of saying this is that it will occur on the average μ times.) What is the probability that the number of events is k, not μ?

2. With the previous distribution, the researcher needed to know the probability of success and failure in each time period. The number of trials was also important. Here one only needs to know the average number of times an event is expected to occur.

   The number of trials is not important.

Function - pdf

$$P(x = k) = \frac{\mu^k e^{-u}}{k!}$$

Function-cdf

$$P(x \leq j) = \sum_{i=0}^{j} P(x = i)$$

Use of a table of values
Mathematica Example
Mean, Variance, Standard Deviation

$$\mu = \mu$$

$$\sigma^2 = \mu$$

$$\sigma = \sqrt{\mu}$$

Note!!! Poisson distribution can be used as a replacement for the binominal - the guideline for determine this is:

$$np < 7$$

If this is true, the poisson can be used in place of the binomial distribution.