Theoretical Probability Distributions

Last time we discussed probabilities of events
Today we will describe random variables
- A variable is used to describe any characteristic that can be measured or categorized (discrete variables, continuous variables)
- Random variable: a variable whose actual observed value is governed by chance
- Probability distribution: describes how likely various outcomes of a random variable (or intervals) are
- Discrete today; continuous next week

Discrete Probability Distributions

- If X is a discrete random variable; X = number of individuals exposed to high levels of ozone per day, then
- P(X= x) is the probability that X takes on the value x;

- Empirical probability distributions are based on observed relative frequencies

Theoretical Distributions

- Rather than describing a probability distribution by probabilities of all outcomes, we can sometimes use theoretical distributions that apply under certain assumptions
  - Binomial Distribution
  - Poisson Distribution

Binomial Distribution

Assumptions:
- A fixed number n trials or experiments are carried out; the outcome of each trial can be classified in precisely one of two mutually exclusive events
  - "success" or "failure"
- The probability of a success p, stays the same from trial to trial. 1 - p = probability of a failure
- The trials are independent; the outcome on one trial has no bearing on the outcome of any other trial

The random variable X, the number of success in the n trials, has a Binomial distribution

Bin(n,p)
**Air Pollution**

Monitor air pollution in the LA basin for a one week period. Let \( X \) be the number of days out of the 7 on which the concentration of ozoner exceeds the maximum levels set by the NAAQS.

*Does \( X \) have a Binomial Distribution?*

---

**Blood Lead**

Childhood lead poisoning is a public health concern in most urban areas of the US. In a certain area, 1 in 10 children has high blood lead levels (greater than 30 micrograms/deciliter)

*In a random sample of 3 children from this area, do the number of children with high blood lead levels follow a Binomial distribution?*

---

**Binomial Probabilities**

Binomial Probability distribution

\[
P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}
\]

Binomial Coefficient: number of ways to select \( x \) objects out of \( n \) without regard to order

\[
\binom{n}{x} = \frac{n!}{x!(n-x)!}
\]

0! = 0, 1! = 1, 2! = 2*1, 3! = 3*2*1

\[n! = n*(n-1)*...*1\]
Expected Number of Successes

Expected Value: Theoretical average or mean of the distribution

\[ \sum x \, P(X = x) \]

For the Binomial distribution the expected number of successes is np (proof...)

Variance = np(1-p)
standard deviation = sqrt(variance)

Large n

Number of elderly individuals admitted to hospital due to asthma (induced by high levels of ozone?) in a large metropolitan area on a given day.

- X= {0,1,2,..... n}
- n is very large (and possibly unknown)
- p is small

use a Poisson Distribution

Poisson Distribution

- X represents the number of events in a fixed interval (often time or space or both)
  - the probability that a single event occurs within an interval is proportional to the length of the interval
  - an infinite number of occurrences of the event are possible (no fixed number of trials)
  - the events occur independently both within the same interval and successive intervals

\[ P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \]

\[ \lambda = \text{Expected number of occurrences of the event in an interval; approximately np} \]