More Discrete Distributions

Poisson

Suppose X counts occurrences of some observable event in an interval $[0, T]$ of time, and that we partition this interval into n subintervals. Let us assume further that

- the event count in any subinterval is independent of those in others.
- by making time subintervals small enough (i.e., large enough n), the number of occurrences in any subinterval is limited to 0 or 1. This assumption is tantamount to saying we can take X_j, the number of occurrences in subinterval j, to be a Bernoulli random variable.
- the probability of an occurrence in subinterval j is proportional to the length of the subinterval. Let us denote this probability by λ/n, making it basically true that each $X_j \sim \text{Binom}(1, \lambda/n)$.

Under these assumptions, then the total count X of occurrences is roughly

$$X = X_1 + X_2 + \cdots + X_n \sim \text{Binom}(n, \lambda/n),$$

which means

$$P(X = x) \approx \binom{n}{x}\left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \underbrace{\frac{n!}{x!(n-x)!\,n^x}}_{\to 1}\lambda^x \underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{e^{-\lambda}}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-x}}_{\sim 1}$$

$$= \boxed{e^{-\lambda}\frac{\lambda^x}{x!},} \quad \text{for large } n.$$

potential pmf

What we have done here is "invent" a possible pmf, and state conditions under which counts might be accurately modeled by it.

Claim 1: The function $f_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}$, defined for nonnegative integers x, is a pmf.

Two things to check

(1) Sum $\sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!} = 1$

(2) $f_X(x) \geq 0$.

$X \sim \text{Pois}(\underline{\lambda})$

λ = rate parameter

Claim 2: (actually Lemma 2.7.2) If $X \sim \text{Pois}(\lambda)$, then

(a) $\boxed{E(X) = \lambda.}$

(B) $\text{Var}(X) = \lambda$.

NCAA Fumbles. Do NCAA fumble counts seem to follow a Poisson model?

There are an average of 1.75 fumbles per game during Week 1

```
mean(~ week1, data=Fumbles)

[1] 1.75
```

so it seems like the Poisson model with $\lambda = 1.75$ would make for the closest fit.

```
myTab <- tally(~week1, data=Fumbles)
xcoords <- as.numeric( row.names( myTab ) )
ycoords <- as.numeric( myTab ) / 120
gf_point( ycoords ~ xcoords ) %>%
  gf_dist( "pois", params = list(lambda=1.75), color="blue" ) %>%
  gf_segment(0 + dpois(0:10, lambda=1.75) ~ (0:10) + (0:10), color="blue" )
```

Assume X = # of arrivals ~ Pois

Example: Arrivals at the bank.
Suppose that, during the noon hour (noon - 1 pm), the average number of customers coming inside for help with a bank teller is 60. What is the probability that

(a) a 10-minute period elapses during the noon hour without customers?

count ~ Pois($\lambda = 10$)

$$P(\text{count} = 0) = e^{-10} \cdot \frac{10^0}{0!}$$
$$= \text{dpois}(0, \text{lambda} = 10)$$

(b) 100 or more customers come in a single day during the noon hour?

Still assuming count $X \sim \text{Pois}(\lambda = 60)$

$$1 - \text{sum}(\text{dpois}(0{:}99, \text{lambda} = 60))$$

$$= 1 - \text{ppois}(99, \text{lambda} = 60)$$