Today

• Homework Solutions

• Expected Values in More Detail

Ch. 1 #17) A purchaser samples 4 items from a lot of 100 and rejects the lot if one or more are defective. Find the probability that the lot is accepted as a function of the percentage of defective items.
As a binomial – let \( p \) be the percentage of the entire population that is defective.

\[
P[\text{lot is accepted}] = P[0 \text{ defectives chosen}] = \binom{4}{0} p^0 (1 - p)^{4-0}
\]

For \( p = 0.2 \) this probability is 0.4096

As a hypergeometric – let \( r = np = 100p \) be the number of defectives out of the lot of 100.

\[
P[0 \text{ defectives chosen}] = \binom{100p}{0} \binom{100 - 100p}{4-0} \binom{100}{4}
\]

For \( p = 0.2 \) this probability is 0.4033

\[
\text{ps<-c(0,0.05,0.10,0.15,0.2,0.25)}
\]
\[
\text{binacc<-dbinom(0,4,ps)}
\]
\[
\text{hypacc<-dhyper(0,100*ps,100*(1-ps),4)}
\]
\[
\text{plot(0,0,xlim=c(0,0.25),ylim=c(0.2,1),type="n",main="Black=Binomial, Green=Hypergeometric",xlab="Percent Defective",ylab="Prob of Accept")}
\]
\[
\text{lines(ps,hypacc,col="Green",lwd=2)}
\]
\[
\text{lines(ps,binacc,col="Black",lwd=2)}
\]
Chapter 2 – RVs (continued…)

A discrete random variable $X$ is defined by its probability mass function $\rho(x_i) = P(X=x_i)$.

The cumulative distribution function (cdf) is $F(x) = P(X \leq x)$.
The mean or expected value of a discrete random variable \(X\) is
\[\mu_X = E(X) = \sum_i x_i p(x_i)\]

The variance of a discrete random variable \(X\) is
\[\sigma_X^2 = \text{Var}(X) = \sum_i (x_i - \mu_X)^2 p(x_i)\]

We have already seen the Binomial Distribution
\[p(x) = \binom{n}{x} p^x (1 - p)^{n-x} \text{ for } x \in 0, \ldots, n\]

And Hypergeometric Distribution
\[p(x) = \binom{r}{x} \binom{n-r}{m-x} \binom{n}{m} \text{ for } x \in 0, \ldots, \min(m, r)\]

Notice that calculating the mean and variance of these distributions appears to be very unpleasant!

For example
\[E(X) = \sum_{x=0}^n xp(x) = \sum_{x=0}^n \binom{n}{x} p^x (1 - p)^{n-x}\]
Expected Values for Discrete RVs
(from Chapter 4)

Definition (pg. 111): If X is a discrete RV with p.m.f. p(x) then

\[ E(X) = \sum x \cdot p(x) \]

when it exists.

Expected value of a function:

Theorem (Pg. 116): If X is a discrete RV with p.m.f. p_X(x) then

\[ E(g(X)) = \sum g(x) \cdot p_X(x) \]

Proof: Let Y be the random variable where for any \( \omega \in \Omega \), \( Y(\omega) = g(X(\omega)) \).

Let A_y be all the x's that correspond to y.

Note that this gives \( \rho_Y(y) = \sum_{x \in A_y} p_X(x) \)

And by definition

\[ E(g(X)) = E(Y) = \sum_{y} \rho_Y(y) \]
Note that this gives \( p_Y(y_i) = \sum_{x \in A_i} p_X(x) \)
So... \( E(g(X)) = E(Y) = \sum_i p_Y(y_i) \)
\[ = \sum_i \{ \sum_{x \in A_i} p_X(x) \} \]
\[ = \sum_i \sum_{x \in A_i} \lambda_i p_X(x) \]
\[ = \sum_i \sum_{x \in A_i} g(x) p_X(x) \]
\[ = \sum_i g(x) p_X(x) \]
\[ \square \]

One place this is used is to get the formula for variance:

\[ \text{Var}(X) \equiv E[(X - \mu_X)^2] \]
\[ = \sum (x_i - \mu_X)^2 p(x) \]

The theorem also allows us to prove two results about a linear function of a random variable:

\[ g(X) = a + bX \]

The constant \( a \) represents a shift and the multiplier \( b \) represents a change of scale.
\[ E(a + bX) = \Sigma_x (a + bx)p_X(x) \]
\[ = \Sigma_x \{ a p_X(x) + bx p_X(x) \} \]
\[ = a \Sigma_x p_X(x) + b \Sigma_x x p_X(x) \]
\[ = a + b E(x) \]

\[ \text{Var}(a + bX) = E[((a + bX) - \mu_{a+bX})^2] \]
\[ = E[(a + bX - (a + b \mu_X))^2] \]
\[ = E[(bX - b \mu_X)^2] \]
\[ = E[b^2(X - \mu_X)^2] \]
\[ = b^2 \text{Var}(X) \]

Neither of these seem to help us with the finding the expected value and variance of the binomial though.

What could help us there is something that let us find the expectation and variance of a sum of independent random variables.
Theorem: (special case of A on 119 and A on 131)

Let \( X_1, X_2, \ldots X_n \) be mutually independent random variables, then:

\[
\mu_{\Sigma X} = E(\Sigma X_i) = \Sigma E(X_i) = \Sigma \mu_X
\]

\[
\sigma_{\Sigma X}^2 = \text{Var}(\Sigma X_i) = \Sigma \text{Var}(X_i) = \Sigma \sigma_X^2
\]

Sketch of Proof: Consider the case of two random variables \( X \) and \( Y \) with p.m.f.s \( p_X(x) \) and \( p_Y(y) \) respectively.

\[
E(X+Y) = \Sigma_{x,y} (x+y)p(X=x,Y=y)
\]

\[
= \Sigma_{x,y} (x+y)P(X=x)P(Y=y)
\]

\[
= \Sigma_x \Sigma_y x P(X=x)P(Y=y)
\]

\[
= \Sigma_x \Sigma_y x P(X=x)P(Y=y) + \Sigma_y \Sigma_x y P(Y=y)P(X=x)
\]
\[ E[X] = \sum_x x \cdot P(X=x) P(Y=y) + \sum_y y \cdot P(Y=y) P(X=x) \]

\[ = \sum_x \{ x \cdot P(X=x) \cdot \sum_y P(Y=y) \} + \sum_y \{ y \cdot P(Y=y) \cdot \sum_x P(X=x) \} \]

\[ = \sum_x x \cdot P(X=x) + \sum_y y \cdot P(Y=y) \]

\[ = E(X) + E(Y) \]

For a binomial X with sample size \( n \) and probability p,

\[ E(X) = np \]

\[ \text{Var}(X) = np(1-p) \]