STAT 702/J702

September 7th, 2004

-Lecture 6-

Instructor: Brian Habing
Department of Statistics
Telephone: 803-777-3578
E-mail: habing@stat.sc.edu

Today

- Homework Solutions
- Finishing Binomial vs. Hypergeometric
- Discrete Random Variables

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.

$$P[\text{lot is accepted}] = P[0 \text{ defectives chosen}]$$

$$= \binom{4}{0} p^0 (1-p)^{4-0}$$

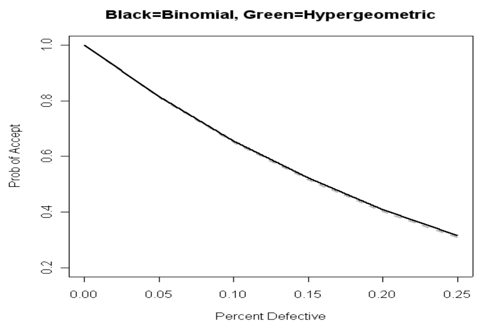
For *p* =0.2 this probability is 0.4096

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

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

For *p* =0.2 this probability is 0.4033


```
ps<-c(0,0.05,0.10,0.15,0.2,0.25)
binacc<-dbinom(0,4,ps)
hypacc<-
  dhyper(0,100*ps,100*(1-ps),4)
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")
lines(ps,hypacc,col="Green",lwd=2)
lines(ps,binacc,col="Black",lwd=2)
```



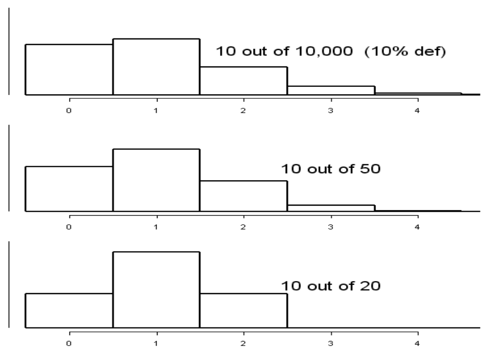
Binomial vs. Hypergeometric (cont…)

When are the binomial and hypergeometric similar?

For a fixed sample size, when the population size goes to infinity the hypergeometric probability converges to the binomial probability.

What do you lose if you sample with replacement instead? (e.g. why not always use binomial?)

Imagine the case where the population size and sample size are nearly equal?



Chapter 2 – Random Variables

In many of the examples we saw before the event we were interested in was a number.

We converted
HHHHHHHHHTTTTTTTTT
to the number 10 for example.

A random variable is a function that assigns each sample point a numerical value.

Depending on whether the sample space is discrete or continuous the random variable can either be discrete or continuous.

We will begin by considering discrete random variables.

A discrete random variable X is defined by its <u>probability mass function</u> $p(x_i) = P(X=x_i)$

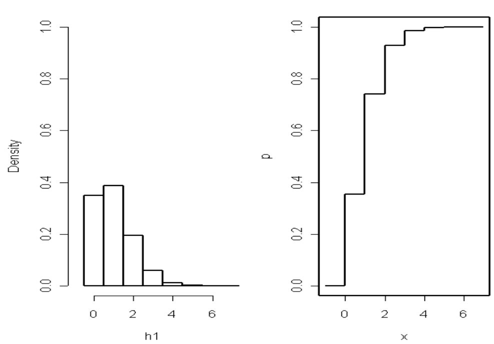
Such that $\Sigma_i \, p(x_i) = 1$

The p.m.f. can be graphed like a histogram.

The cumulative distribution function (cdf) is

$$F(x) = P(X \leq x)$$

This is a step function.



We can also construct numerical summaries of random variables.

The mean or expected value of a discrete random variable X is
$\mu_X = E(X) = \Sigma_i x_i \, p(x_i)$

The variance of a discrete random variable X is
$\sigma_X^2 = Var(X) = \Sigma_i (x_i - \mu_X)^2 \, p(x_i)$

Bernoulli Random Variable

The Bernoulli Random variable is the indicator of an event:

$X = I_A(\omega) =$ 1 if $\omega \in A$
0 otherwise

We can also write the Bernoulli p.m.f.

$p(x) = p^x(1-p)^{1-x}$ if $x = 0$ or 1
$\quad\quad 0$ otherwise

This puts it in the same format regardless of the value of *x*, and also matches the form of the binomial.

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

$$p(x) = \frac{\binom{r}{x}\binom{n-r}{m-x}}{\binom{n}{m}} \text{ for } x \in 0,\ldots \min(m,r)$$