3.7 The Hypergeometric Distribution

- Recall the example of sampling several items from an assembly line and counting the number of defectives among the sampled items.
- The count of defectives followed a ________________ distribution.
- But suppose we were sampling from a collection of items that was not extremely large.

Example 1: We select 10 items (without replacement) from a finite collection of 100 items (2% of which are defective) and count the number of defectives in our sample.

- This count is not a binomial r.v. Why not?

Example 2: Suppose we have an urn containing N balls, r of which are red and $(N-r)$ of which are black. We draw n $(< N)$ balls out (without replacement). Our r.v. of interest Y is the number of sampled balls that are red. Then Y has a hypergeometric distribution.

Defn: If Y follows a hypergeometric distribution as described above [Shorthand: $Y \sim \text{hyper}(r, N, n)$], then its probability function is:

Proof:

Theorem: The hypergeometric distribution is a valid probability distribution.

Outline of Proof:

Theorem: If $Y \sim$ hyper(r, N, n):

- These results can be proved directly via the definitions of expected values and variance, but the proof is tedious.

Example 3: From a set of 20 potential jurors (8 African-American and 12 white), 6 jurors were selected. If the jury selection was random, what is the probability of one or fewer African-Americans on the jury?

- What is the expected number of African-Americans on the jury? What is the standard deviation?

Relationship to Binomial Distribution

- If the "population size" N is large, the probability of "success" on each draw changes very little from draw to draw.
- The experiment will be "close" to a

____________ ____________.

- For N large, the hypergeometric distribution approximately equals a binomial distribution with success probability $p = \frac{r}{N}$.

Note:	Binomial	Hypergeometric
$E(Y)$		
$V(Y)$		

Note: If N is much larger than n, this "finite population adjustment"

Formally, for fixed y,

Back to Example 1 (Collection of 100 items):

Treating Y as hypergeometric:

- See R examples on course web page for calculating hypergeometric probabilities in R.

Proof of Hypergeometric Expected Value

Lemma: $y\binom{r}{y} = r\binom{r-1}{y-1}$ and $n\binom{N}{n} = N\binom{N-1}{n-1}$.

- If $Y \sim hyper(r, N, n)$:

$$E(Y) = \sum_{y=0}^{n} y\, P(Y=y)$$