WMS5 § 3.7 **Hypergeometric** Probability Distribution

Used to describe the Variability of

the Proportion / Count in a random sample drawn, <u>without replacement</u>, from a <u>finite</u> population/universe of N binary elements (0's and 1's); sampling fraction is sizable.

	Choose	Do not choose	Total
"1" elements	y	N1 - y	N1*
"0" elements	n-y	N0 - (n-y)	N0
	n	N-n	N (universe)

(* WMS5 use "r" where we use "N1")

<u>What it is</u>

- **The n+1 probabilities p_0, p_1, ... p_y, ... p_n of observing**

  0 "positive"
  1 "positive"
  2 "positives"
  . ..
  y "positives"
  .
  .
  n "positives"

  in n draws <u>without</u> replacement from the N items

NB **If N1 < n, then the range of y will be less than the full 0 to n, since some of the n+1 possibilities are not possible. e.g., ...**

	Choose	Do not choose	Total
"1" elements	y	11 - y	11*
"0" elements	29-y	20 +y	49
	29	31	60 (universe)

- **Apart from sample size (n), the probabilities p_0 to p_n depend on the 2 parameters N1 and N0 (or equivalently N1 and N)**

<u>How it arises</u>

-Sample surveys of small universes (eg MP's) or with large sampling fractions (even if N large)

-Quality Control Sampling from finite lots

-Psychophysics (tea tasting, water divining, ...)

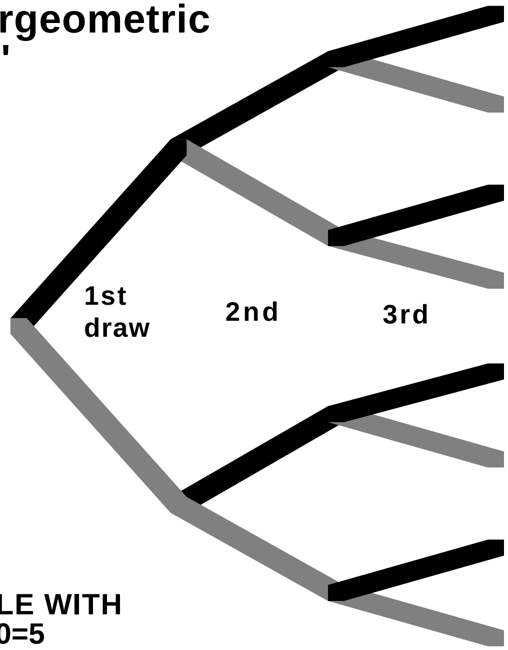
- To evaluate if discrimination in assigning/choosing people (from a pool) for tasks/positions etc..

-Statistical comparison of proportions in small samples (see example of Tamoxifen in preventing recur. of Br Ca)

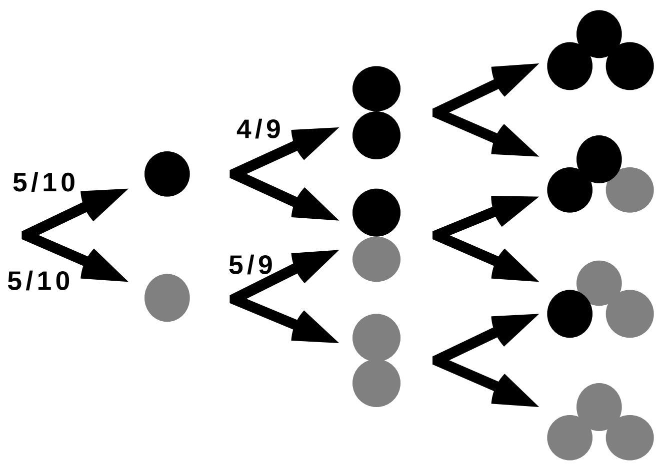
-Lotteries (6/49, Keno, ...)

-To estimate size of wildlife populations (Capture-recapture method)

Hypergeometric "Tree"



EXAMPLE WITH
N1=5, N0=5



(Note that the successive outcomes of draws 1-3 are dependent, so must be careful. Can still multiply and add. Can always turn problem around to make it a tree (see labour dispute example)

Calculations are simplified by fact that all sequences of y +'s & (n–y) –'s have same probabilitysi, in lieu of adding, can multily this prob. by # , i.e. $^{n}C_{y}$, of such sequences

See "Hypergeometric P's, E and V" on web page

Calculating Hypergeometric probabilities

- **Formula (or 1st principles)**

Prob(y out of n)

$$= \frac{[\text{ N1 choose y }] \times [\text{N0 choose (n-y) }]}{\text{N choose n}}$$

- **Calculator / Spreadsheet (see elsewhere on web page)**
- **Approximations to Hypergeometric**

  - Binomial Distribution (n a small fraction of N)

  - Gaussian Distribution (y >> 0 and y << n)

E(Y)	Var(Y)	SD(Y)
$n \times \frac{N1}{N}$	$n \times \frac{N1}{N} \times \frac{N0}{N} \times \frac{N-n}{N-1}$	$\sqrt{VAR}$

WMS5 § 3.7 **Hypergeometric** Probability Distribution

Worked Examples

- **Tea Tasting (small examples, last page of notes on Ch2_1_2_6)**
- **Tamoxifen**
- **6/49 (earlier in Ch 3 notes)**
- **Keno**
- **Banco [loto-québec .. several]**
- **Labour Dispute (wms5 e.g. 2.10, p39 + spreadsheet)**
- **Rhino Politics and small sample sizes (last page of excerpts from notes fron 607)**
- **Example of assessimg food sensitivity (3rd last page)**

Good exercises from text

- **3.75 (jury selection: 6 from pool of N1=8 African Americans and N0=12 white)**
- **3.80 (assume that Y has the value 1, i.e. that Y=1 of the n=3 animals had been tagged previously). Maximize P(Y=1) by trying various "what if" values of N.**
- **Exercise 3.71 and 3.77 have a "quality control" flavour. Can you think of a closer-to-quality-control example?**