Introduction to probability

Stat 134       FALL 2005
Berkeley

Lectures prepared by:
Elchanan Mossel
Yelena Shvets

Follows Jim Pitman’s book:
Probability
Section 2.5
Sampling with replacement

• Suppose we have a population of size $N$ with $G$ good elements and $B$ bad elements. We draw $n$ times with replacement from this population.

• The number $g$ of good elements in the sample will have a binomial$(n,p)$ distribution with $p = G/N$.

$$P(g \text{ good and } b \text{ bad}) = \binom{n}{g} \frac{(G)^g(B)^b}{(N)^n}.$$
Sampling with replacement

• If $n$ is large, this will be well approximated by $N(np, \sqrt{np(1-p)})$.

• The proportion of good elements in the sample $g/n$ will lie in the interval $p \pm \frac{1}{\sqrt{n}}$ with probability 95%.

• If the $p$ is not known, it can be estimated by the method of confidence intervals.
Sampling without replacement

• Let’s now think about drawing without replacement. The sample size has to be restricted to $n \cdot N$.

• Then number of possible orderings of $n$ elements out of $N$ is:

$$ (N)_n = N(N-1)(N-2) \ldots (N-n+1). $$

• $(N)_n$ is called $N$ order $n$
Sampling without replacement

Note that:

\[
\binom{N}{n} = \frac{(N)_n}{n!}.
\]

So:

\[
(N)_n = \binom{N}{n} n!.
\]
Sampling without replacement

• The chance of getting a sample with $g$ good elements followed by $b$ bad ones is:

$$\frac{G}{N} \cdot \frac{G-1}{N-1} \cdot \frac{G-g+1}{N-g+1} \cdot \frac{B}{N-g} \cdot \frac{B-1}{N-g-1} \cdots \frac{B-b+1}{N-g-b+1} = \frac{(G)_g (B)_b}{(N)_n}.$$

• Since there are $\binom{n}{g}$ samples with $g$ good and $b$ bad elements all having the same probability, we obtain:
Sampling with and without replacement

• For sampling without replacement:

\[ P(\text{g good and b bad}) = \binom{n}{g} \frac{(G)_g (B)_b}{(N)_n} = \frac{\binom{G}{g} \binom{B}{b}}{\binom{N}{n}}. \]

• For sampling with replacement:

\[ P(\text{g good and b bad}) = \binom{n}{g} \frac{(G)^g (B)^b}{(N)^n}. \]
Hypergeometric Distribution.

- The distribution of the number of good elements in a sample of
  - size $n$
  - without replacement
- From a population of
  - $G$ good and
  - $N-G$ bad elements

Is called the hypergeometric distribution with parameters $(n,N,G)$. 
Sampling with and without replacement

- When $N$ is large, $(N)_n / N^n \approx 1$.
- When $B$ is large, $(B)_b / B^b \approx 1$.
- When $G$ is large, $(G)_g / G^g \approx 1$.

So for fixed $b,g$ and $n$ as $B,G,N \approx 1$ the hypergeometric distribution can be approximated by a binomial$(n,G/N)$. 