Department of Mathematics
College of William & Mary

MATH 451/551

Chapter 4. Common Discrete Distributions

4.6 Hypergeometric Distribution

GuanNan Wang
gwang01@wm.edu

Motivating Example

If a dozen eggs contains three bad eggs, find the probability of obtaining exactly x defective eggs in a sample of size 5.

Framework as an Urn Model

Framework as an urn model:

- urn contains N balls
- m of which are “successes”
- $N - m$ of which are “failures”
- random sample of n balls is drawn without replacement

HyperGeometric Distribution

- A discrete random variable X with probability mass function

$$f(x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}, \qquad x = 0, 1, 2, \ldots, n$$

  for some nonnegative integer parameter N, $n = 0, 1, 2, \ldots, N$, and $m = 0, 1, 2, \ldots, N$, is a *HyperGeometric*(m, N, n) random variable.

- **Alternative Support:**

$$\mathcal{A} = \{x | x = \max\{0, (m + n - N)\}, \ldots, \min\{n, m\}\}.$$

HyperGeometric Distribution

- **Rationale:**
  - upper bound $\min\{n, m\}$
    - you can't get more successes than n in the sample of size n
    - you can't get more successes than the number of successes m in the population
  - lower bound $max\{0, (m + n - N)\}$
    - you can't get fewer successes than 0 in the sample
    - you are guaranteed to get at least $m + n - N$ successes if the population size is small enough

Mean

Variance

R Functions

R Functions

Function	Returned Value
dhyper(x, m, n, k)	calculates the probability mass function $f(x)$
phyper(x, m, n, k)	calculates the cumulative distribution function $F(x)$
qhyper(u, m, n, k)	calculates the percentile (quantile) $F^{-1}(u)$
rhyper(m, m, n, k)	generates m random variates

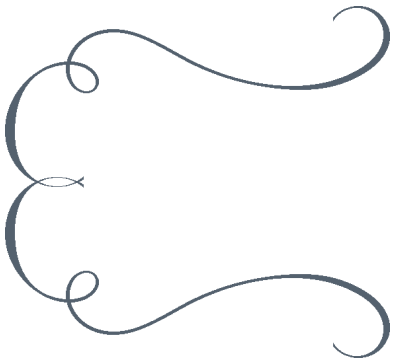
Example 1

A five-card is dealt from a well-shuffled 52-card deck. Let X be the number of diamonds in the hand. Find $P(X = 3)$ and $P(X \leq 3)$.

Example 2

A biologist uses a “catch and release” or “mark and recapture" program to estimate the population size of a particular animal in a region. During the catch phase, 20 animals are captured, tagged and released. Several months later, 30 animals are captured, and 7 of them have tags. What is the most likely population size?

Thank You



THANK YOU!