INTRODUCTION TO PROBABILITY MODELS

Lecture 16

Qi Wang, Department of Statistics

Sep 26, 2018

HYPERGEOMETRIC DISTRIBUTION

HYPERGEOMETRIC DISTRIBUTION

- **The definition of X**: the number of success in n trail without replacement from a finite population of size N that contains exactly M objects with that feature.
- **Support:** $\{0, 1, 2, \cdots, n\}$ or $\{0, 1, 2, \cdots, M\}$
- **Parameters:**
  - N : Population size
  - M : Number of possible successes
  - n : Number of trials
- **PMF:** $P_X(x) = \frac{C_x^M C_{n-x}^{N-M}}{C_n^N}$
- **Expected Value:** $E[X] = n\frac{M}{N}$
- **Variance:** $Var(x) = n\frac{M}{N}(1 - \frac{M}{N})\frac{N-n}{N-1}$
- $X \sim Hyper(N, n, M)$

EXAMPLE 1

There are 100 identical looking 52" TVs at Best Buy in Costa Mesa, California. Let 10 of them be defective. Suppose you want to buy 8 of the aforementioned TVs (at random).

1. What is the probability that you don't get any defective TVs? Identify the distribution parameters and support.
2. Given that we purchase at least one defective TV, what is the probability that you purchase fewer than 3 defective TVs?
3. What is the expected number of defective TVs that you will purchase?
4. Find the standard deviation of the number of defective TVs that you purchase.

EXAMPLE 2

An experiment consists of shuffling a standard deck of 52 cards and then dealing a 5 card hand. Let Y denote the number of diamonds in the hand.

1. Identify the distribution of Y and give its parameter(s) and support. Find the probability that Y is 2.
2. Suppose instead of using 1 deck, we mix together 1,000 decks. The cards are shuffled and 5 are dealt into a hand. Let D denote the number of diamonds in the hand. Find the exact probability that you get 2 diamonds.

THE BINOMIAL APPROXIMATION TO THE HYPERGEOMETRIC

In probability, we can use some distributions to approximate others.

- If $X \sim Hyper(N, n, M)$ AND $N > 20n$, then $X \sim Bin(n, p = \frac{m}{N})$
- With a large enough population, sampling without replacement will also get a Binomial.
- So back to Example 2, is an approximate distribution appropriate for D, why or why not?
- Use that approximation to find P(D = 2). What is the distribution, parameter(s) and support for this approximating distribution?