Discrete Random Variables redux

Module 14

Statistics 251: Statistical Methods

Updated 2019

Discrete Random Variables redux

This supplemental module covers two distributions from the textbook not covered in class (geometric and hypergeometric) plus an additional one not covered in the text

(1) Geometric distribution

There are three assumptions for an experiment to have a geometric distribution
(a) The experiment continues until the first success happens (which is preceded by $n-1$ failures)
(b) Only two outcomes are possible: success with probability p, and failure with probability $q = 1 - p$
(c) The n trials are independent, and since n is not fixed, there needs to be at least one trial

The probability is the probability of $n-1$ failures until the first success (on the last trial)

$$\text{Shorthand notation: } X \sim geom(p) \textit{ or } X \sim G(p)$$

Geometric formulas

$$P(X = x) = q^{x-1}p \quad x = 1, 2, \cdots, \infty$$

$$EX = \frac{1}{p} \;\; VX = \frac{q}{p^2} \;\; SDX = \sqrt{\frac{q}{p^2}}$$

You are allowed to use your calculator to do as much of the calculation as you want. There are instructions on how to use the command if you have a TI graphing calculator on p. 259.

Geometric example

The probability of a defective steel rod is 0.01. Steel rods are selected at random. Find the probability that the first defect occurs on the ninth steel rod; at least nine before the first defect?

$$X \sim geom(0.01)$$

$$P(X = x) = q^{x-1}p$$

$$q = 1 - p = 1 - 0.01 = 0.99$$

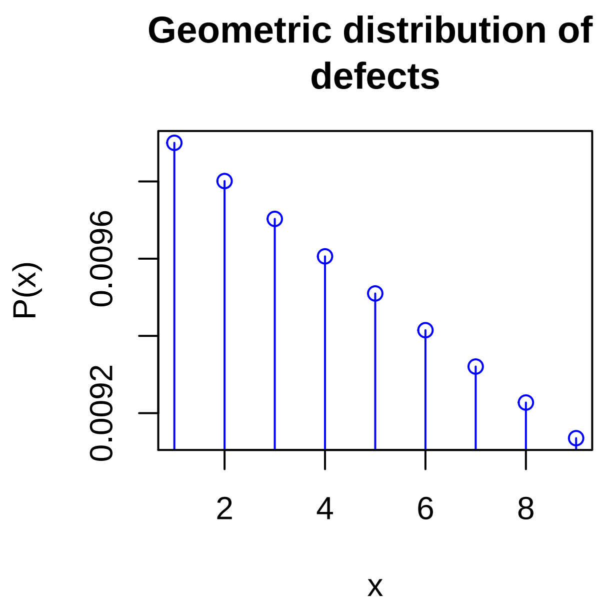
Geometric example continued

$$P(X=9) = (0.99)^{9-1}(0.01) = 0.009135$$

$$P(X \geq 9) = \sum_{9}^{\infty} geom(9:\infty, 20, 0.35)$$
$$= 1 - P(X < 9) = 1 - P(X \leq 8)$$

$= 1 - [P(1) + \cdots + P(8)] = 1 - [(0.99)^{8-1}(0.01) + (0.99)^{7-1}(0.01) \; +(0.99)^{7-1}(0.01) + (0.99)^{6-1}(0.01) + (0.99)^{5-1}(0.01) \; +(0.99)^{4-1}(0.01) + (0.99)^{3-1}(0.01) + (0.99)^{2-1}(0.01) \; +(0.99)^{1-1}(0.01)] = 1 - 0.076483 = 0.923517$

Geometric Graph



Geometric EX, VX, SDX

$$EX = \frac{1}{p} = \frac{1}{0.01} = 100$$

$$VX = \frac{q}{p^2} = \frac{0.99}{0.01^2} = 9900$$

$$SDX = \sqrt{VX} = \sqrt{9900} = 99.498744$$

(2) Hypergeometric distribution

There are three assumptions for an experiment to have a hypergeometric distribution
(a) The population or set to be sampled consists of N individuals (a finite population)
(b) Each individual is either a success or failure, and there are M successes in the population
(c) Sampling is done as swor from the combined groups; each pick is not independent (because of swor)

The probability is the probability of the number of items *from the group of interest*

Hyper Logistics

The probability is the probability of the number of items *from the group of interest*

$$\text{Shorthand notation: } X \sim hyper(M, n, N) \text{ } or \text{ } X \sim H(M, n, N)$$

You are allowed to use your calculator to do as much of the calculation as you want. Use the ${}_nC_r$ function for all three parts of the calculation.

Hypergeometric formulas

$$P(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \quad x = 0, 1, \cdots, n$$

M = sample size of things of interest
x = argument of interest
$N - M$ = second group of things
$n - x$ = number of things from second group (dependent on how many things from the 1st group)
N = the total number of things

Hyper EX, VX, SDX

$$EX = \frac{Mn}{N}$$

$$VX = \left(\frac{N-n}{N-1}\right)\left(\frac{Mn}{N}\right)\left(1 - \frac{M}{N}\right)$$

$$SDX = \sqrt{VX}$$

Hypergeometric example

An animal population thought to be near extinction in a certain region had five individuals that were caught, tagged, and released to mix back into the population. A random sample of 10 of these animals is selected after a time. Let X = the number of tagged animals in the second sample. Say there are actually 25 animals of this type in the region, what is the probability that there are 2, at most 2?

$$X \sim hyper(M = 5, n = 10, N = 25)$$

$$P(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

Hyper example continued

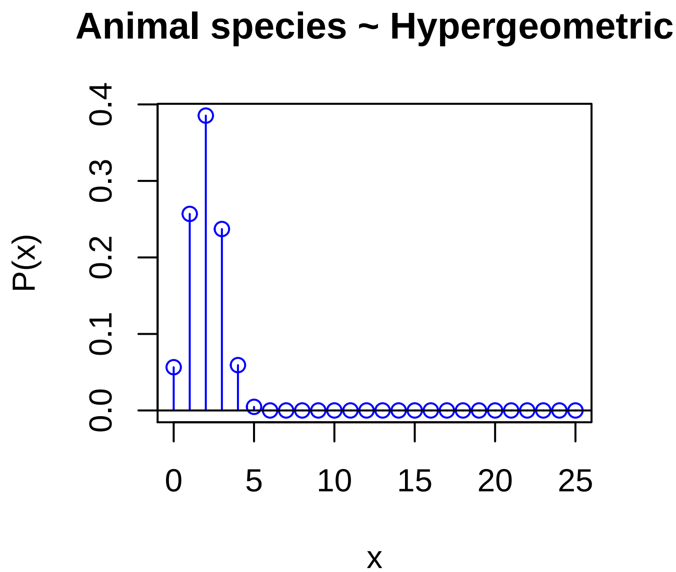
$$P(X = 2) = \frac{\binom{5}{2}\binom{25-5}{10-2}}{\binom{25}{10}} = 0.385375$$

$$P(X \leq 2) = \sum_{0}^{2} hyper(0:2, 5, 10, 25) = P(0) + P(1) + P(2)$$

$$= \sum(0.056522, 0.256917, 0.385375)$$

$= \frac{\binom{5}{2}\binom{25-5}{10-2}}{\binom{25}{10}} + \frac{\binom{5}{1}\binom{25-5}{10-1}}{\binom{25}{10}} + \frac{\binom{5}{0}\binom{25-5}{10-0}}{\binom{25}{10}} = 0.698814$

Hypergeometric Graph



Hyper EX, VX, SDX

$EX = \frac{Mn}{N} = \frac{5(10)}{25} = 2$
$VX = \left(\frac{N-n}{N-1}\right)\left(\frac{Mn}{N}\right)\left(1 - \frac{M}{N}\right) = \left(\frac{25-10}{25-1}\right)\left(\frac{5(10)}{25}\right)\left(1 - \frac{10}{25}\right) = 0.75$
$SDX = \sqrt{VX} = \sqrt{0.75} = 0.866025$

(3) Negative binomial distribution

The assumptions for an experiment to have a negative binomial distribution
(a) The n trials are independent
(b) Only two outcomes are possible: success with probability p, and failure with probability $q = 1 - p$
(c) Probability of success, p, is constant from trial to trial
(d) The experiment continues until a total of r successes have been observed, where r is a *specified* positive integer

The probability of the number of failures which occur before a target number of successes is reached

Example situation for negative binomial distribution

Consider the following statistical experiment[1]. You flip a coin repeatedly and count the number of times the coin lands on heads. You continue flipping the coin until it has landed 5 times on heads. This is a negative binomial experiment because: (1) The experiment consists of repeated trials. We flip a coin repeatedly until it has landed 5 times on heads (2) Each trial can result in just two possible outcomes: heads or tails (3) The probability of success is constant: 0.5 on every trial (4) The trials are independent; that is, getting heads on one trial does not affect whether we get heads on other trials. The experiment continues until a fixed number of successes have occurred; in this case, 5 heads.

$$\text{Shorthand notation: } X \sim nb(r,p) \text{ or } X \sim NB(r,p)$$

[1] https://stattrek.com/probability-distributions/negative-binomial.aspx

Negative binomial logistics

r is the target number of successes
p is the probability of success

Because the formula is basically identical to the binomial, you are allowed to use your calculator to do as much of the calculation as you want. There are instructions on how to use the command if you have a TI graphing calculator on page 254. Most scientific calculators have a ${}_nC_r$ command (same as "n choose x" for binomial and "$x-1$ choose $r-1$" for the negative binomial) for combinations

Negative binomial formulas

$$P(X=x) = \binom{x-1}{r-1} p^r q^{x-r} \quad x = 0, 1, \cdots, \infty$$

$$EX = \frac{rq}{p}$$

$$VX = \frac{rq}{p^2}$$

$$SDX = \sqrt{\frac{rq}{p^2}} = \sqrt{VX}$$

Negative binomial example

A pediatrician wishes to recruit 5 couples, each of whom is expecting their first child, to participate in a new natural childbirth regimen. Let $p =$ the probability that a randomly selected couple agrees to participate. If $p = 0.2$, what is the probability that 15 couples must be asked before 5 are found who agree to participate (i.e. what is the chance that 10 couples refuse before the 5^{th} couple agrees)? What is the probability that the first 5 couples asked agree? What is the probability that at most 10 couples refused before the 5^{th} couple agrees? What is the expected number of refusals (failures) until the 5^{th} couple agrees (the 5^{th} success)? What is the variance and standard deviation?

Negative binomial example continued

$$X \sim nb(r, p) \Rightarrow X \sim nb(5, 0.2)$$

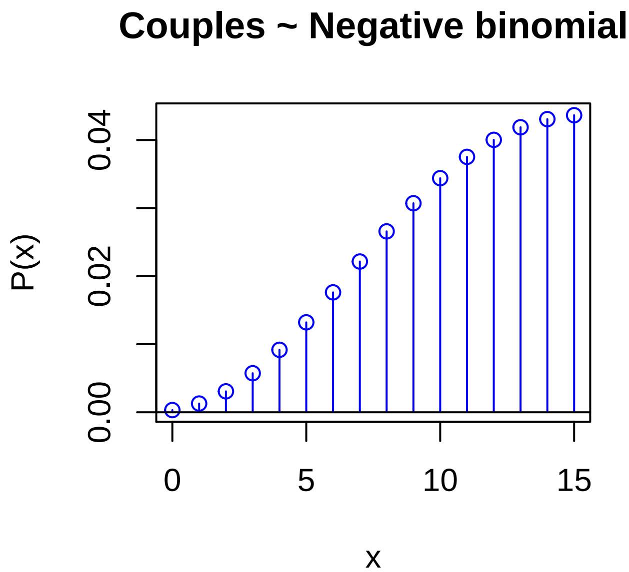
$$P(X=10) = \binom{10+5-1}{5-1}(0.2)^5(0.8)^{10} = \binom{14}{4}(0.2)^5(0.8)^{10} = 0.034394$$

$$P(X=0) = \binom{0+5-1}{5-1}(0.2)^5(0.8)^0 = \binom{4}{4}(0.2)^5(0.8)^0 = 0.00032$$

$$P(X \leq 10) = P(0) + P(1) + \cdots + P(10) = \binom{0+5-1}{5-1}(0.2)^5(0.8)^0 + \binom{1+5-1}{5-1}(0.2)^5(0.8)^1 + \cdots$$

$$+ \binom{10+5-1}{5-1}(0.2)^5(0.8)^{10} = 0.164234$$

Negative binomial graph



Negative binomial EX, VX, SDX

$$EX = \frac{rq}{p} = \frac{5(0.2)}{0.8} = 20$$

$$VX = \frac{rq}{p^2} = \frac{5(0.8)}{(0.2)^2} = 100$$

$$SDX = \sqrt{\frac{rq}{p^2}} = \sqrt{VX} = sqrt100 = 10$$