Discrete Distributions

SOC 512 & CSSS 505

written by Laina Mercer & Jessica Godwin

Aaron Osgood-Zimmerman

Department of Statistics
University of Washington

February 18, 2020

Motivation

A discrete random variable could have a number of different probability distributions. Today we will focus on the following discrete distributions.

- Bernoulli
- Binomial
- Multinomial
- Geometric
- Hyper Geometric
- Poisson
- Negative Binomial

Bernoulli

Some experiments only have two possible outcomes. Often the experiments can be phrased as either having a characteristic or not having a characteristic. Examples

- Flipping a coin - Heads or Tails
- Disease Incidence - Either you have it or you don't
- A test - pass or fail

Bernoulli

Any experiment that can be described as either a 'success' or a 'failure' has a **Bernoulli** distribution. For $0 \leq p \leq 1$

$$X = \begin{cases} 1 \text{ with prob} & p \\ 0 \text{ with prob} & 1-p \end{cases}$$

You need to first determine what outcome is your 'success' - usually the outcome of interest. For example, if you were trying to determine the probability of a specific birth defect, each child born would be assigned a 1 if he/she had the defect, 0 if not. The

probability distribution for a Bernoulli random variable:

$$P(X = x|p) = p^x(1-p)^{1-x} \text{ for } x = 0, 1$$

The expected value of X when $X \sim Bern(p)$ (has a Bernoulli)

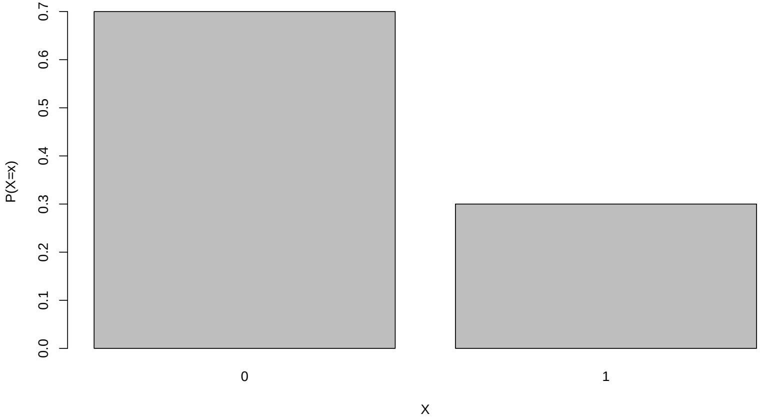
$$
\begin{aligned}
E[X] &= \sum_{i=1}^{n} x_i \cdot P(X = x_i | p) \\
&= 0 \cdot p^0 (1-p)^{1-0} + 1 \cdot p^1 (1-p)^{1-1} \\
&= 0 + 1 \cdot p^1 (1-p)^0 \\
&= p
\end{aligned}
$$

The variance of X when $X \sim Bern(p)$

$$\begin{aligned}
V[X] &= \sum_{i=1}^{n}(x_i - E[X])^2 \cdot P(X = x_i|p) \\
&= (0-p)^2 \cdot p^0(1-p)^{1-0} + (1-p)^2 \cdot p^1(1-p)^{1-1} \\
&= p^2(1-p)^1 + (1-p)^2 \cdot p^1 \\
&= p(1-p)[p + (1-p)] = p(1-p) \cdot 1 \\
&= p(1-p)
\end{aligned}$$

Bernoulli

Distribution when $p = 0.3$.



Binomial

What if we were looking at several Bernoulli random variables? We could ask the question: how many of the random variables or experiments were successes?

Examples

- A couple has 3 children. What is the probability that 2 are redheads?
- We flip 10 coins. What is the probability we get 6 heads?
- We survey 25 teenage drivers. What is the probability that 15 of them have had car accidents?

The binomial distribution requires n independent experiments. We have to randomly sample people, flip fair coins, etc. The success of one random variable or experiment must have no effect on the success of any other random variable or experiment.

Binomial

We say that we have n experiments, each with success probability p and failure probability $1 - p$.

Probability Distribution:

$$P(X = x|n, p) = \binom{n}{x} p^x (1 - p)^{n-x} \text{ for } x = 0, 1, 2, 3..., n$$

This looks very similar to a Bernoulli except for the constant $\binom{n}{x}$.

Binomial

$\binom{n}{x}$ counts how many combinations of x successes and $n - x$ failures we can have out of n experiments.

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Example: To find the probability of 2 redheads out of 3 children, we need to consider all combinations of children that result in 2 redheads.

Out of the 8 possible combinations $S = \{RRR, RRR^C, RR^CR, R^CRR, RR^CR^C, R^CRR^C, R^CR^CR, R^CR^CR^C\}$ three of the combinations (RRR^C, RR^CR, R^CRR) result in 2 redheads.

$$\binom{3}{2} = \frac{3!}{2!(3-2)!} = \frac{3 \cdot 2 \cdot 1}{(2 \cdot 1)(1)} = \frac{6}{2} = 3$$

Binomial

Mean & Variance

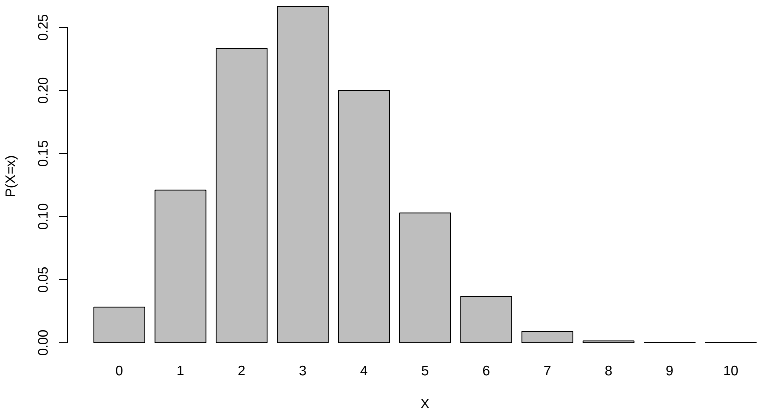
The easiest way to find the mean and variance of X when $X \sim Bin(n, p)$ is to think of $X = Y_1 + ... + Y_n$ where $Y_i \sim Bern(p)$.

$$\begin{aligned} E[X] &= E\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} E[Y_i] \\ &= \sum_{i=1}^{n} p = np \end{aligned}$$

$$\begin{aligned} V[X] &= V\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} V[Y_i] \\ &= \sum_{i=1}^{n} p(1-p) = np(1-p) \end{aligned}$$

Binomial

Distribution when $p = 0.3$ and $n = 10$.



Multinomial

We can expand the binomial to the case where we have n experiments each with 3 possible outcomes.

For example: asking someone to check a box from (agree, couldn't care less, do not agree). We can ask 'out of ten people, what is the probability that three check agree, four checked couldn't care less, and three checked do not agree.' Here our outcome is a vector with a *multinomial* distribution. This can be expanded to include m outcomes.

$$P(X = x_1, ..., x_m | n, p_1, ..., p_n) = \frac{n!}{x_1! \cdots x_m!} p_1^{x_1} \cdots p_m^{x_m}$$

where $p_1 + ... + p_m = 1$ and $x_1 + ...x_m = n$.

Multinomial

Mean, Variance, & Covariance

The expected number of times we observe outcome i, where $i = 1, ..., m$.

$$E[X_i] = np_i$$

The variance of X_i

$$V[X_i] = np_i(1 - p_i)$$

The covariance of X_i and X_j

$$cov[X_i, X_j] = -np_i p_j.$$

Geometric

Sometimes we are interested in how long it will take us to get to a success. That is, how many failure do we have to see before we get a success?

Examples

- What is the probability that we have to lose 10 games before a win?
- What is the probability that we see 4 tails before a head?
- What is the probability that you will take your qualifying exams 3 times before you pass?

Geometric

Mean & Variance

These are examples of a *geometric* distribution. If X=the number of the trial with the first success, the probability distribution of X is:

$$P(X = x|p) = p(1-p)^{x-1} \text{ for } x = 1, 2, ...$$

The expected value of X is

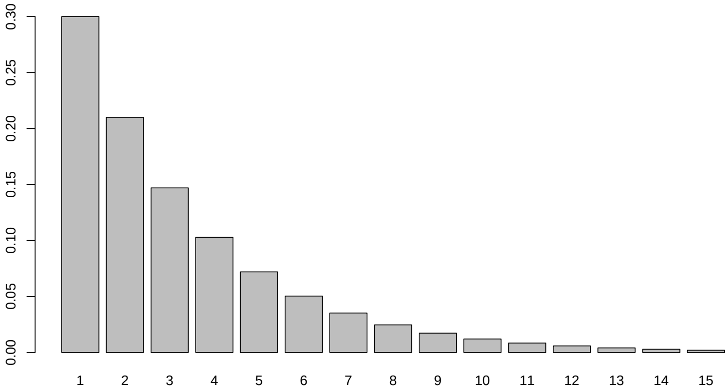
$$E[X] = 1/p$$

The variance of X

$$V[X] = \frac{1-p}{p^2}.$$

Geometric

Distribution when $p = 0.3$, $X =$the trial with the first success.



Geometric

Alternative parameterization

The geometric distribution can also be parameterized such that $X =$ the number of failures before a success

$$P(X = x|p) = p(1-p)^x \text{ for } x = 0, 1, 2, ...$$

The expected value of X is

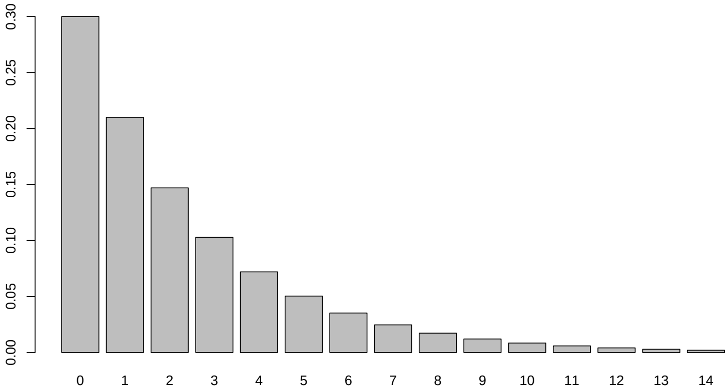
$$E[X] = \frac{1-p}{p}$$

The variance of X

$$V[X] = \frac{1-p}{p^2}$$

Geometric

Distribution when $p = 0.3$, $X =$the number of failures before the first success.



Hyper Geometric

Often we are taking samples without replacement and are interested in the probability of k success out of the n samples taken from a population of size N. When you draw from a finite population, the probability of success changes on each draw.

For example, if you are in a room with 5 men and 5 women. Initially the probability of selecting a woman is $1/2$. If the first draw is a man, then there are 9 people left with 5 women. The probability of drawing a woman after a man has been selected is $5/9$.

Hyper Geometric

Mean & Variance

The hypergeometric distribution requires three parameters. The size of the population N, the total number of 'successes' $K \in \{0, 1, .., N\}$ and the size of the sample $n \in \{0, 1, ..., N\}$.

The probability distribution of X is

$$\frac{\binom{K}{x}\binom{N-x}{n-x}}{\binom{N}{n}}$$

The expected value is

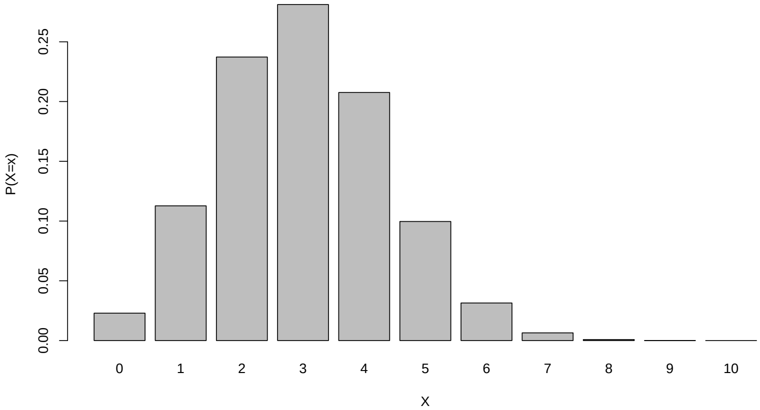
$$E[X] = n\frac{K}{N}$$

The variance is

$$V[X] = n\frac{K}{N}\frac{(N-K)}{N}\frac{(N-n)}{(N-1)}.$$

Hyper Geometric

Distribution when $N = 100$, $K = 30$, and $n = 10$.



Poisson

The *Poisson* distribution is often used when we are interested in how many events will occur in fixed amount of time or space.
Examples

- What is the probability that three people walk into a store in the next ten minutes?
- What is the probability that we get two phone calls in the next hour?
- What is the probability that 10 people are diagnosed with leukemia in King county in 2015?

Poisson

Mean & Variance

The probability distribution for $X \sim \mathit{Poisson}(\lambda)$ is

$$P(X = x|\lambda) = \frac{e^{-\lambda} \cdot \lambda^x}{x!} \text{ for } x = 0, 1, 2, 3....$$

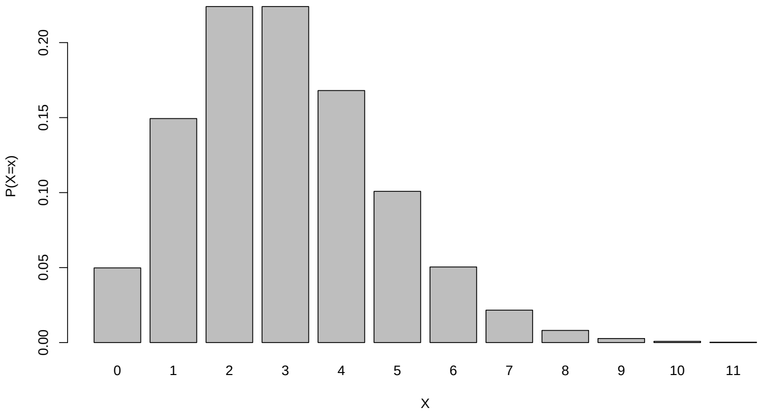
where λ is the average number of occurrences in a given time period (rate).

The mean and variance of a $X \sim \mathit{Poisson}(\lambda)$

$$E[X] = \lambda, \; V[X] = \lambda$$

Poisson

Distribution when $\lambda = 3$.



Negative Binomial

Suppose we are interested in observing Bernoulli random variables until a predefined number of failures has occurred. Then the number of successes we observe will have the *negative binomial* distribution.

The negative binomial distribution requires two parameters, $p =$ the probability of success and $r =$the predefined number of failures.

In practice, the concept of the 'failures' and 'successes' can be flipped. For example we could model then number of free throws that need to be taken for a total of 10 to be made, assuming the probability of any shot going in is $p = 0.7$. Here r=10 would be the predefined number of 'failures'. Our X would be the number of times the free throw is missed before we make the shot 10 times.

Negative Binomial

Mean & Variance

If X =the number of successes with $X \sim NegBin(r, p)$ the probability distribution is

$$P(X = x|r, p) = \binom{x + r - 1}{x}(1 - p)^r p^x \text{ for } x = 0, 1, 2, 3....$$

The mean of a $X \sim NegBin(r, p)$

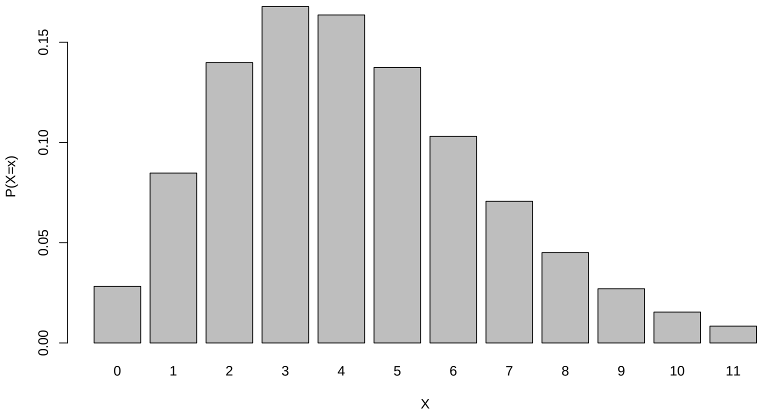
$$E[X] = \frac{pr}{1 - p}$$

and the variance is

$$V[X] = \frac{pr}{(1 - p)^2}.$$

Negative Biniomial

Distribution when $r = 10$ and $p = 0.7$.



Which Distribution?

- A couple has four pets. X =number of dogs.
  $X \sim Bin(n = 4, p = 1/2)$
- A local Seattleite just wants 2 days of sun. X = number of rainy days before two sunny days.
  $X \sim NegBin(r = 2, p = 0.05)$
- A student takes a qualifying exam until they pass, with a 0.9 probability of passing each time. X =the number of times they take the exam. $X \sim Geometric(p = 0.9)$
- A summer camp is trying to buy sunblock for summer camps and needs to know if there will be more than 10 redheads in the camp if the incidence of red hair is about 1 in 100 people (in the Pacific Northwest). X =number of redheads.
  $X \sim Poisson(\lambda = 0.01)$

Which Distribution?

- There are 200 people with diabetes in a census block of 600 people. We randomly sample 50 people. $X =$number sampled with diabetes.
  $X \sim HyperGeometric(N = 600, K = 200, n = 50)$
- Kids play 10 rounds of twister. $X_1 =$number spinner on green, $X_2 =$number spinner on red, $X_3 =$number spinner on yellow, and $X_4 =$number spinner on blue. $X \sim Multinomial(p_1 = 1/4, p_2 = 1/4, p_3 = 1/4, p_4 = 1/4, n = 10)$

The End

Questions?