Module C: Understanding the Data-Generating Process

Slide Deck C6:

Negative Binomial Distributions

The section in which we learn about another named discrete distributions in which the random variable is the number of trials. A random variable that follows a Negative Binomial distribution will model the number of failures until the rth success.

Today's Objectives

By the end of this slidedeck, you should

1. determine what types of random variables follow a Negative Binomial distribution using its definition
2. calculate the expected value and variance
3. identify the two parameters defining the Negative Binomial distribution
4. calculate probabilities associated with a Negative Binomial random variable

Definition of a Negative Binomial Experiment

Definition

The **Negative Binomial distribution** describes the probability of obtaining the rth success after X failures.

The difference between a Binomial and a Negative Binomial distribution is that the Binomial specifies the number of trials, while the Negative Binomial treats that as the random variable.

Definition of a Negative Binomial Experiment

Definition

The **Negative Binomial distribution** describes the probability of obtaining the rth success after X failures.

Examples

- number of times you need to draw a card (with replacement) before you get the third heart
- number of people you need to ask before you meet your 17th Sophomore at Knox College (with replacement)
- number of experiments run until the fifth one fails (where each has a 5% chance of failing)

Negative Binomial pmf

Recall that the probability mass function (pmf) provides the probability of each element of the sample space. For a Negative Binomial random variable, there are an infinite number of possible outcomes (failures until rth success):

$$S = \{0, 1, 2, \infty\}$$

Note that the pmf of a Negative Binomial random variable is easily devised from the definition. Recall for the Binomial distribution:

$$\mathbb{P}[\ X = x\ ] = \underbrace{\binom{n}{x}}_{\text{combinatons}} \quad \underbrace{p^x}_{\text{successes}} \quad \underbrace{(1-p)^{n-x}}_{\text{failures}}$$

How can we use this to create the pmf for the Negative Binomial distribution?

Negative Binomial pmf

For the Negative Binomial, we know there are r successes — but the last is fixed at the end. Thus, the pmf is

$$\mathbb{P}[\ X = x\ ] = \underbrace{\binom{n}{x}}_{\text{combinatons}} \quad \underbrace{p^x}_{\text{successes}} \quad \underbrace{(1-p)^{n-x}}_{\text{failures}} \qquad \text{Binomial}$$

$$\mathbb{P}[\ X = x\ ] = \underbrace{\binom{(x+r-1)}{x}}_{\text{combinatons}} \quad \underbrace{p^r}_{\text{successes}} \quad \underbrace{(1-p)^{x}}_{\text{failures}} \qquad \text{Negative Binomial}$$

In the Negative Binomial, x is the number of failures until the success $\#r$.

Negative Binomial pmf

From their definitions, it is clear that the Negative Binomial is a generalization of the Geometric. Where the Geometric is the number of failures until the *first* success ($r = 1$), the Negative Binomial is the number of failures until the rth success.

Because of this, we can focus on only the Negative Binomial distribution in much the same way that we can focus on the Binomial and ignore the Bernoulli.

The two main parameters are

$$\mathbb{E}\left[X \right] = r \, \frac{1}{p}$$

$$\mathbb{V}\left[X \right] = r \, \frac{1}{p^2}$$

Example 1: The Ace of Spades

Example

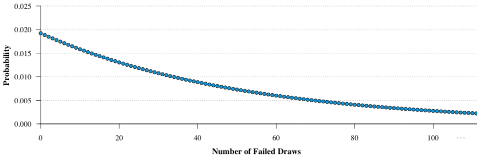
Let X be the number of times I draw one card (with replacement) from a standard deck of cards *before* I get an Ace of Spades. If the deck is fair, then it is clear that

$$X \sim \mathcal{N}eg\mathcal{B}in(r = 1, p = 1/52)$$

1. What is the probability of getting the first Ace of Spades on the first draw?
2. What is the probability of getting the first Ace of Spades on the 10th draw?
3. What is the probability of getting the first Ace of Spades on the 100th draw?

Example 1: The Ace of Spades

A graphic of the probability mass function (pmf) of $X \sim \mathcal{N}eg\mathcal{B}in(r = 1, p = 1/52)$:



Example 1: The Ace of Spades

Example

Let X be the number of times I draw a card (with replacement) from a standard deck of cards until I get the Ace of Spades. If the deck is fair, then it is clear that

$$X \sim \mathcal{N}eg\mathcal{B}in(r = 1, p = 1/52)$$

1. What is the probability of getting the Ace of Spades on the first draw?

Remember that the Negative Binomial models the number of *failures* until the rth success. So, since we are asked to calculate a probability about getting the Ace of Spades on the first draw, we are asked to calculate the probability of having 0 failures. **That is**, we are asked to calculate $\mathbb{P}[\ X = 0\ ]$.

Example 1: The Ace of Spades

In R, this is
`dnbinom(0, 1, 1/52)` $= 0.019\,230\,770$

Example 1: The Ace of Spades

Example

Let X be the number of times I draw a card (with replacement) from a standard deck of cards until I get the Ace of Spades. If the deck is fair, then it is clear that

$$X \sim \mathcal{N}eg\mathcal{B}in(r = 1, p = 1/52)$$

2. What is the probability of getting the Ace of Spades on the tenth draw?

That is, we are asked to calculate $\mathbb{P}[\ X = 9\ ]$. This is a simple application of the pmf. In R, this is just
`dnbinom(9, 1, 1/52)` $= 0.016\,147\,230$

Example 1: The Ace of Spades

Example

Let X be the number of times I draw a card (with replacement) from a standard deck of cards until I get the Ace of Spades. If the deck is fair, then it is clear that

$$X \sim \mathcal{N}eg\mathcal{B}in(r = 1, p = 1/52)$$

3. What is the probability of getting the Ace of Spades on the 100th draw?

That is, we are asked to calculate $\mathbb{P}[\ X = 9\ ]$. This is a simple application of the pmf. In R, this is just

`dnbinom(99, 1, 1/52)` $= 0.002\,812\,633$

Today's Objectives

Now that we have concluded this lecture, you should be able to

1. determine what types of random variables follow a Negative Binomial distribution using its definition
2. calculate the expected value and variance
3. identify the two parameters defining the Negative Binomial distribution
4. calculate probabilities associated with a Negative Binomial random variable

Useful R Functions

In this slide deck, we covered (or hinted) on the following four R functions related to the Negative Binomial distribution:

- `dnbinom(x, size, prob)` is the pmf, p for: $\mathbb{P}[\ X = x\ ] = p$
- `pnbinom(x, size, prob)` is the CDF, p for: $\mathbb{P}[\ X \leq x\ ] = p$
- `qnbinom(q, size, prob)` is the quantile function, x for: $\mathbb{P}[\ X \leq x\ ] = p$
- `rnbinom(n, lambda)` generates n random values from: $\mathcal{N}eg\mathcal{B}in(\texttt{size}=r, \texttt{prob}=p)$

Please do not forget to access the `allProbabilities` document that provides all of the important probability functions in R.

Supplemental Activities

The following are supplements for the topics covered today.

- SCA 5 is for some discrete distributions

Note that you can access all Statistical Computing Activities here:
`https://www.kvasaheim.com/courses/stat200/sca/`

In addition to the SCA, **Laboratory Activity B** is helpful for learning how to handle discrete distributions. The lab actually shows the connection between sampling and discrete distributions. It uses three named distributions. While it does not use the Negative Binomial distribution, understanding it will help you improve the lab activity by adding yet another distribution.
`https://www.kvasaheim.com/courses/stat200/labs/`

Supplemental Readings

The following are some readings that may be of interest to you in terms of understanding these two discrete distributions:

- Hawkes Learning: None
- Intro to Modern Statistics: None
- R for Starters: Appendix A.5
- Wikipedia: Negative Binomial Distribution