Binomial and Negative Binomial Random Variables

Probability Theory: Foundation for Data Science

with **Anne Dougherty**

Data Science

UNIVERSITY OF COLORADO **BOULDER**

Random Variables

At the end of this module, students should be able to

- Define a discrete random variable and give examples of a probability mass function and a cumulative distribution function.
- Calculate probabilities of Bernoulli, **Binomial**, Geometric, and **Negative Binomial** random variables.
- Calculate the expectation and variance of a discrete rv.

Binomial random variable

Examples:

- Suppose you toss a fair coin 12 times. What is the probability that you'll get 5 heads?
- Suppose you pick a random sample of 25 circuit boards used in the manufacture of a particular cell phone. You know that the long run percentage of defective boards is 5%. What is the probability that 3 or more boards are defective?
- Suppose 40% of online purchasers of a particular book would like a new copy and 60% want a used copy. What is the probability that amongst 100 random purchasers, 50 or more used books are sold?

These three situations, and many more, can be modeled by a binomial random variable.

Properties of a binomial random variable;

- Experiment is n trials (n is fixed in advance)
- Trials are identical and result in a success or a failure (i.e. Bernoulli trials) with $P(\text{success}) = p$ and $P(\text{failure}) = 1 - p$.
- Trials are independent (outcome of one trial does not influence any other)

If X is the number of successes in the n independent and identical trials, X is a binomial random variable.
Notation: $X \sim Bin(n, p)$

Find the pmf, expectation, and variance for a binomial random variable, $X \sim Bin(n, p)$.

What is the sample space for a binomial experiment?
$S =$

$P(X = 0) =$

$P(X = 1) =$

$P(X = 2) =$

$P(X = k) =$

Definition: The expected value of a discrete random variable, $E(X)$, is given by

$$E(X) = \sum_k kP(X = k)$$

$X \sim Bin(n, p)$

Definition: The **variance** of a random variable is given by $\sigma_X^2 = V(X) = E[(X - E(X))^2]$.

Computational formula: $V(X) = E(X^2) - (E(X))^2$.

Negative binomial random variable

Examples:

- Suppose you toss a fair coin until you obtain 5 heads. How many tails before the fifth head?
- Suppose you randomly choose circuit boards until you find 3 defectives. You know that the long run percentage of defective boards is 5%. How many must you examine?
- Suppose 40% of online purchasers of a particular book would like a new copy and 60% want a used copy. How many new books are sold before the fiftieth used book?

These three situations can be modeled by a **negative** binomial random variable.

Definition: Repeat independent Bernoulli trials until a total of r successes is obtained. The negative binomial random variable Y counts the number of failures before the r^{th} success. Notation: $Y \sim NB(r, p)$.

- The number of successes r is fixed in advance.
- Trials are identical and result in a success or a failure (i.e. Bernoulli trials) with $P(\text{success}) = p$ and $P(\text{failure}) = 1 - p$.
- Trials are independent (outcome of one trial does not influence any other)

Compare to $X \sim Bin(n, p)$: X is the number of successes in the n independent and identical trials and n is fixed in advance.

Example: A physician wishes to recruit 5 people to participate in a medical study. Let $p = .2$ be the probability that a randomly selected person agrees to participate. What is the probability that 15 people must be asked before 5 are found who agree to participate.

Y is the number of failures before the 5 people are found.

If $Y \sim NB(r, p)$, we have

$P(Y = k) = \binom{k+r-1}{r-1} p^r (1-p)^k$ for $k = 0, 1, 2, \ldots$

$$E(Y) = \frac{r(1-p)}{p} \text{ and } V(Y) = \frac{r(1-p)}{p^2}$$

Relationship between geometric and negative binomial random variables?