Probability Theory

Applications for Data Science
Module 3 Discrete Random Variables

Anne Dougherty

February 14, 2021

TABLE OF CONTENTS

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.

Discrete random variables can be categorized into different types or classes. Each type/class models many different real-world situations.

Bernoulli rv

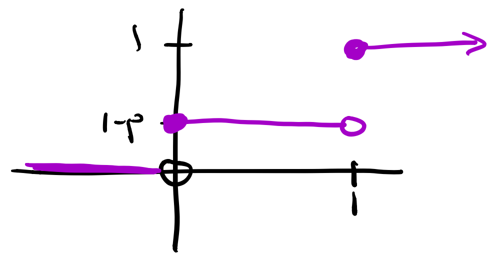
Bernoulli rv, sometimes called a binary rv, is any random variable with only two possible outcomes: 0 or 1.

The probability mass function (pmf) is given by:

$$P(X=1) = p$$

$$P(X=0) = 1-p$$

cdf $F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1-p & \text{if } 0 \le x < 1 \\ 1 & \text{if } 1 \le x \end{cases}$



$$F(x) = P(X \le x)$$

has the distribution of

Notation: We write $X \sim \text{Bern}(p)$ to indicate that X is a Bernoulli rv with success probability p.

Geometric rv

Motivating Example A patient needs a kidney transplant and is waiting for a matching donor. The probability that a randomly selected donor is a suitable match is p.

What is the sample space? What is an appropriate rv? What is the pmf?

$S = \{1, 01, 001, 0001, \ldots\}$

Let X = # of donors tested until a match is found

$X \in \{1, 2, 3, 4, \ldots\}$

$P(X=1) = p$

$P(X=2) = P(\{01\}) = (1-p)p$

$P(X=3) = P(\{001\}) = (1-p)^2 p$

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

Geometric series

$$a + ar + ar^2 + ar^3 + \dots = \sum_{k=1}^{\infty} a r^{k-1} = \begin{cases} \dfrac{a}{1-r} & \text{if } |r|<1 \\ \text{diverges} & \text{if } |r| \geq 1 \end{cases}$$

pmf for a geometric r.v.

$$P(X=k) = (\underbrace{1-p}_{r})^{k-1} \underbrace{p}_{a} \quad \leftarrow \text{pmf for a geometric r.v.}$$

Verify $\displaystyle\sum_{k=1}^{\infty} P(X=k) = 1$

$$\sum_{k=1}^{\infty} (\underbrace{1-p})^{k-1} p = \frac{p}{1-(\underbrace{1-p}_{r})} = 1$$

note: $r = 1-p < 1$

Geometric rv - continued

A **geometric rv** consists of independent Bernoulli trials, each with the same probability of success p, repeated until the first success is obtained.

- Each trial is identical, and can result in a success or failure.
- The probability of success, p, is constant from one trial to the next.
- The trials are independent, so the outcome on any particular trial does not influence the outcome of any other trial.
- Trials are repeated until the first success.

Geometric rv - continued

Summary

- Sample space for a geometric rv:

  $S = \{1, 01, 001, \ldots\}$

- Probability mass function for a geometric rv with probability of success p

  $P(X=k) = (1-p)^{k-1} p, \quad X \in \{1, 2, 3, \ldots\}$

- Notation: We write $X \sim Geom(p)$ to indicate that X is a geometric rv with success probability p.