Probability Theory

Applications for Data Science
Module 3 Discrete Random Variables

Anne Dougherty

February 17, 2021

TABLE OF CONTENTS

Discrete Random Variables

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.**

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.

Let X be the number of potential donors tested until a match is found.
pmf: $P(X = k) = (1 - p)^{k-1} p$ for $k \in \{1, 2, 3, \ldots\}$

Question: How many potential donors must be tested before there is a successful match? In other words, what is the expected value (also known as the average or mean) of the random variable?

Notation: $E(X)$ or μ_X is the expected value of a random variable X.

Example: 5 exams 70, 80, 80, 90, 90

$$\text{Avg} = \frac{70+80+80+90+90}{5} = \frac{1}{5}(70) + \frac{2}{5}(80) + \frac{2}{5}(90) = 82.5$$

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

$$E(X) = \sum_k k \underbrace{P(X = k)}_{\text{fraction of the population with value } k.}$$

If $X \sim \text{Bern}(p)$, what is $E(X)$?

$$P(X=0) = 1-p, \quad P(X=1) = p$$

$$E(X) = 0\,P(X=0) + 1\,P(X=1) = p$$

If $Y \sim \text{Geom}(p)$, what is $E(Y)$?

Recall: $P(Y = k) = p(1-p)^{k-1}$ for $k = 1, 2, \ldots$

$$E(Y) = \sum_{k=1}^{\infty} k\, P(Y=k)$$

$$= \sum_{k=1}^{\infty} k\, p\,(1-p)^{k-1}$$

$$= \frac{p}{(1-(1-p))^2}$$

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

$$= \frac{1}{p}$$

If $p = \frac{1}{10}$, $E(Y) = \frac{1}{\frac{1}{10}} = 10$

Recall from geometric series:

$$\sum_{k=1}^{\infty} a r^{k-1} = \frac{a}{1-r}, \quad |r|<1$$

Differentiate with respect to r

$$\sum_{k=1}^{\infty} a(k-1) r^{k-2} = \frac{a}{(1-r)^2}$$

$$\sum_{k=2}^{\infty} a(k-1) r^{k-2} = \frac{a}{(1-r)^2}$$

Reindex $k-1 = j$

$$\sum_{j=1}^{\infty} a j r^{j-1} = \frac{a}{(1-r)^2}$$

Expected Value - continued

Useful properties of the expected value definition,

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

- If c is a constant, then $E(c) = c$

  $P(X=c)=1$

- If a and b are constants and X is a rv, then
  $E(aX + b) = \sum_k (ak+b)P(X=k) = \underbrace{a\sum_k kP(X=k)}_{E(X)} + \underbrace{b\sum_k P(X=k)}_{1}$

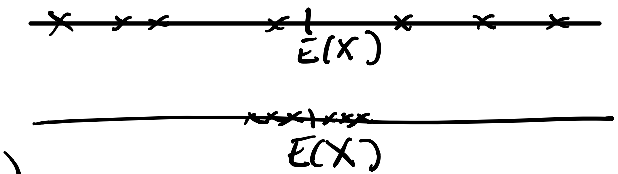
  $= aE(X)+b$

- If $h(X)$ is any function of X, then
  $E(h(X)) = \sum_k h(k)P(X=k)$

Variance

The variance of a random variable X, denoted $V(X)$, measures how far we expect our random variable to be from the mean.

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



$$= \sum_k (k - \mu_X)^2 P(X=k)$$
$$= \sum_k (k^2 - 2\mu_X k + \mu_X^2) P(X=k)$$
$$= \underbrace{\sum_k k^2 P(X=k)} - 2\mu_X \underbrace{\sum_k k P(X=k)}_{\mu_X} + \mu_X^2 \underbrace{\sum_k P(X=k)}_{1}$$
$$= E(X^2) - 2\mu_X^2 + \mu_X^2$$
$$= E(X^2) - \mu_X^2$$

Computational formula: $V(X) = E(X^2) - (E(X))^2$.
(Aside: Standard deviation, $\sigma_X = \sqrt{\sigma_X^2} \geq 0$.)

Examples: Find the variance for $X \sim \text{Bern}(p)$ and for $Y \sim \text{Geom}(p)$.

$X \sim \text{Bern}(p)$ $\quad P(X=0) = 1-p$, $\quad P(X=1)=p$, $\quad E(X)=p$

$V(X) = E(X^2) - (E(X))^2$ $\qquad E(X^2) = \sum_k k^2 P(X=k) = 1^2 \cdot p = p$

$= p - p^2 = p(1-p)$

$Y \sim \text{Geom}(p)$ $\quad P(Y=k) = (1-p)^{k-1} p$, $\quad k=1,2,3,\ldots$

$E(Y) = \frac{1}{p}$

$E(Y^2) = \sum_k k^2 P(Y=k) = \sum_k k^2 (1-p)^{k-1} p = \frac{2-p}{p^2}$

$V(Y) = E(Y^2) - (E(Y))^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}$

Example: if $p = \frac{1}{10}$, $E(Y) = 10$, $V(Y) = 90$

$\sigma_Y = \sqrt{90} \simeq 9.5$

Example: Suppose you have 10 folded pieces of paper, labeled $0, 1, 2, \ldots, 9$. Draw one paper at random. Define the rv U to be the number drawn. Find the pmf, expectation, and variance for U. (Aside: this is a discrete, uniform rv.)

pmf: $P(U=k) = \frac{1}{10}$, $k = 0, 1, 2, \ldots 9$

$$E(U) = \sum_{k=0}^{9} k\, P(U=k) = 0 \cdot \frac{1}{10} + 1 \cdot \frac{1}{10} + \cdots + 9 \cdot \frac{1}{10} = 4.5$$

$$E(U^2) = \sum_{k=0}^{9} k^2 P(U=k) = \sum_{k=0}^{9} k^2 \left(\frac{1}{10}\right) = 28.5$$

$$V(U) = E(U^2) - (E(U))^2 = 28.5 - (4.5)^2 = 8.25$$