- TA Office Hours on Course Page
- Homework adventure continues (HW #1 is still with me :( )
- Homework #2 is in progress
  - Thursday morning office hours gone forever

STAT 430: Lecture 8

More Discrete Random Variables

Continuing Chapter 2

Course page: imouzon.github.io/stat430

Random Variables

Idea: we define a probability on sets of outcomes for a given experiment

Problem: - The sample spaces are context specific
- not much can be learned from studying a specific sample space
- Sample spaces hard to describe and possibly hard to work with.

What Random variables do is connect the samples space to the set of real numbers

i.e. Random variables are functions that have the sample space as their domain and the reals as their range.

Since the actual outcome, ω, is unknown, the the value that will result from a function of ω is also unknown.

We say that the probability on Ω is *induced* on X.

Notation and Properties of discrete random variables

We use capital letters from the end of the alphabet to to represent random variables (i.e., X, Y, Z)

We use lower case to represent specific values the random Variable can take

$$X = x, \quad Y = y, \quad Z = z$$

For discrete random variables, we describe the probability on X using a "probability mass function"

pmf: $p(x)$ means $P(X = x)$

Properties of pmfs

1. $p(x) \geq 0$ for all x
2. $\sum_{\text{all } x} p(x) = 1$

Expectation and Variance

notation for Expected Value:

$E(X)$ or μ

other terms: "mean of the RV"
"expectation of the RV"

notation for Variance:

$\text{Var}(X)$ or $V(X)$ or σ^2

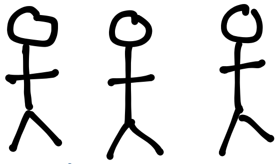
Variance: How spread out are the possible values of X.

What these two properties do is:

Expectation: "what is the center of the possible values of X if we take the probability into account"

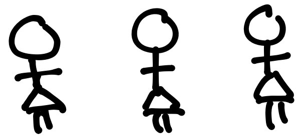
2 groups

Group 1



6'0" 6'0" 6'0"

Group 2



5'6" 6'0" 6'6"

Mean: $\frac{6+6+6}{3} = 6$ $\qquad$ $\frac{5.5+6+6.5}{3} = 6$

Spread: the typical rabbit is 0 ft from the mean $\qquad$ $\frac{0.5+0+0.5}{3} = \frac{1}{3}$

Notice: $\text{mean} = \frac{\text{sum of all values}}{\#\text{ of values}} = \frac{1}{n}\cdot\text{Value}\#1 + \frac{1}{n}\text{Value}\#2 + \dots + \frac{1}{n}\cdot\text{Value}\#n$

$$\text{Expected value} = \sum_{\text{all } x} x\cdot p(x).$$

What about spread?

First idea: absolute deviation

if μ is my mean

then $\sum_{\text{all } x} |x-\mu|\cdot p(x)$

Variance: $\sum_{\text{all } x} (x-\mu)^2 \cdot p(x)$

Ex: if X is a random variable with

$$P(X=-2) = P(X=-1) = P(X=0) = P(X=1) = P(X=2) = \frac{1}{5}$$

then
$$E(X) = \tfrac{1}{5}(-2) + \tfrac{1}{5}(-1) + \tfrac{1}{5}(0) + \tfrac{1}{5}(+1) + \tfrac{1}{5}\cdot(2)$$
$$= -2/5 - 1/5 + 0/5 + 1/5 + 2/5$$
$$= 0$$

The absolute deviation of X from the mean:

$$\sum_{\text{all } x} |x-\mu| \cdot p(x) = |-2-0| \cdot \frac{1}{5} + |-1-0| \cdot \frac{1}{5} + |0-0| \cdot \frac{1}{5} + |1-0| \frac{1}{5} + |2-0| \cdot \frac{1}{5}$$

$$= 2/5 + \frac{1}{5} + 0 \cdot \frac{1}{5} + \frac{1}{5} + \frac{2}{5}$$

$$= 6/5$$

$$Var(X) = \sum_{\text{all } x} (x-\mu)^2 \cdot p(x) =$$

$$= (-2-0)^2 \cdot \frac{1}{5} + (-1-0)^2 \cdot \frac{1}{5} + (0-0)^2 \cdot \frac{1}{5} + (1-0)^2 \cdot \frac{1}{5} + (2-0)^2 \cdot \frac{1}{5}$$

$$= 4/5 + \frac{1}{5} + \frac{0}{5} + \frac{1}{5} + 4/5$$

$$= 2$$

A Sequence of Bernoulli Experiments

Suppose that we don't just have a single Bernoulli expirement, but instead have a sequence of n *independent* Bernoulli experiments and are interested in the *total* number of successful outcomes.

Example We roll a fair six-sided die 5 times and record a success if we observer a 6.

- What is the sample space for Ω?
- If we define $X_1, X_2, \ldots, X_5$ as a Bernoulli random variables for each single experiment, what is the "sample space" related to the vector $(X_1, X_2, \ldots, X_5)$?
- If we define Y as the total number of successful outcomes, what is the "sample space" for Y?
- How can we write the probability function for Y?

So

$$P(X_1=1, X_2=1, X_3=1, X_4=1, X_5=1) = P(X_1=1)\cdot P(X_2=1)\cdot \ldots \cdot P(X_5=1)$$
$$= \left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right)$$
$$= \left(\tfrac{1}{6}\right)^5$$

$$P(Y=5) = \left(\tfrac{1}{6}\right)^5$$

$$P(X_1=1, X_2=1, X_3=1, X_4=1, X_5=0) = P(X_1=1)\cdot P(X_2=1)\cdot P(X_3=1)\cdot P(X_4=1)\cdot P(X_5=0)$$
$$= \left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right)\left(\tfrac{5}{6}\right)$$

By continuing, we get

$$= \left(\tfrac{1}{6}\right)^4\left(\tfrac{5}{6}\right)^1$$

$$P(Y=4) = 5\cdot\left(\tfrac{1}{6}\right)^4\left(\tfrac{5}{6}\right)^1$$

$$P(Y=3) = \frac{5!}{3!\,2!}\left(\tfrac{1}{6}\right)^3\left(\tfrac{3}{6}\right)^2$$

$$P(Y=2) \quad \frac{5!}{2!\,3!}\left(\tfrac{1}{6}\right)^2\left(\tfrac{5}{6}\right)^3$$

$$P(Y=1) = \frac{5!}{1!\,4!}\left(\tfrac{1}{6}\right)^1\left(\tfrac{5}{6}\right)^4$$

$$P(Y=0) \quad \frac{5!}{0!\,5!}\left(\tfrac{1}{6}\right)^0\left(\tfrac{5}{6}\right)^5$$

Discrete Random Variables

Binomial Random Variables

Binomial Random Variables

def: **Binomial Random Variable**

> Suppose that n experiments, or trials, are performed and that
>
> 1. The trials have two possible outcomes ("success" and "failure")
> 2. Each trial has the same probability of success, p, and
> 3. The trials are independent of each other Then the random variable defined as the ~~sum of th~~e number of successful experiments is a Binomial Random Variable and we say that it follows a Binomial Distribution

Discrete Random Variables

Binomial Random Variables

Binomial Random Variables

Motivation:

If I try this experiment n times, how do I figure out how many outcomes will be what I want?

Statement:

Let X be the number of successful outcomes observed by repeating n independent Bernoulli experiments, each with probability of success p.

Notation: "X is distrubuted as"

$$X \sim \text{Binom}(n, p)$$

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

pmf:

The probability of seeing k successful outcomes is given using

probability of failure

n-k = # failures

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, 2, \dots, n$$

Discrete Random Variables

Binomial Random Variables

Binomial Random Variables

Cumulative Probability:

$$P(X \leq k) = p(0) + p(1) + \ldots + p(k), k = 0, 1, 2, \ldots, n$$

Expectation:

$$\mu = n \cdot p$$

Variance:

$$\sigma^2 = n \cdot p \cdot (1 - p)$$

Discrete Random Variables

Geometric Random Variables

Geometric Random Variables

Motivation:

How many times do I have to repeat this experiment until the outcome is what I want it to be?

Statement:

Let X be the trial upon which the first successful outcome is observed in a sequence of independent Bernoulli experiments, each with probability of success p.

Notation:

$$X \sim \text{Geometric}(p)$$

pmf:

The probability of seeing the first successful outcome on trial k

$$p(k) = p(1-p)^{k-1}, k = 1, 2, \ldots$$

Discrete Random Variables

Binomial Random Variables

Geometric Random Variables

Geometric Random Variables

Cumulative Probability:

$$P(X \leq k) = p(1) + p(2) + \ldots + p(k), k = 1, 2, \ldots$$

Expectation:

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

Variance:

$$\sigma^2 = \frac{1-p}{p^2}$$

Discrete Random Variables

Geometric Random Variables

Discrete Random Variables

Binomial Random Variables

Geometric Random Variables

Poisson Random Variables

Geometric Random Variables

Motivation:

What if I have an almost infinite number of independent Bernoulli trials and a very small probability of success?

Notation:

$$X \sim \text{Poisson}(p)$$

pmf:

The probability of seeing k successes is:

$$p(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Discrete Random Variables

Binomial Random Variables

Geometric Random Variables

Poisson Random Variables

Geometric Random Variables

Cumulative Probability:

$$P(X \leq k) = p(0) + p(1) + p(2) + \ldots + p(k), k = 0, 1, 2, .$$

Expectation:

$$\mu = \lambda$$

Variance:

$$\sigma^2 = \lambda$$