Probability and Mathematical Statistics in Data Science

Lecture 06: Section 3.1: Success and Failure
Section 3.2: Random Variables

Success and Failure

- Bernoulli Trials

- We have Bernoulli trials if:
  - there are two possible outcomes (**success and failure**).
  - the probability of success, *p*, is constant.
  - **the trials are independent.**

Counting outcomes of tosses

- Toss a coin 8 times, how many possible outcomes?

- What is the chance of all heads?

- If each student in a class of 250 flip a coin 8 times, what is the chance that at least 1 person gets all heads?

Random Variables

- A random variable is a variable whose value is a numerical outcome of a random phenomenon.

- For a given sample space S of some experiment, a random variable is **any rule** that associates a number with each outcome in S.

- To put it more mathematically, a random variable is a function whose domain is the sample space and whose range is the set of real numbers.

Random Variables

- A **random variable** assumes a value based on the outcome of a random event.

    - We use a capital letter, like *X*, to denote a random variable.

    - A particular value of a random variable will be denoted with the corresponding lower case letter, in this case *x*.

Random Variables vs. Experiments

- An experiment is a physical setup in real world that provides us intuition about randomness.
- A random variable is a mathematical abstraction that describes randomness.
- When the outcome of the experiment can be seen as numerical, e.g., roll a die, we can effectively treat the experiment as a random variable.

Discrete vs. Continuous

- X is a **discrete random variable** if its possible values either constitute a finite set or else can be listed in an infinite sequence in which there is a first element, a second element, and so on ("countably" infinite).

- X is a **continuous random variable** if it takes all possible values in an interval of numbers or all numbers in a disjoint union of such intervals. No possible value of the variable has positive probability, that is, $P(X=c) = 0$ for any possible value c.

Examples of Random Variables

- There are two types of random variables:
  - **Discrete** random variables can take one of a countable number of distinct outcomes.
    - Example: No. of Asthma Attacks, No. of Patients with a disease
  - Continuous random variables can take any numeric value within a range of values.
    - Example: Height, Weight, Cholesterol Levels

Bernoulli Random Variable

- Arguably the simplest probability model is **Bernoulli.** Any random variable whose possible values are only 0 (failure) and 1 (success) is called a Bernoulli random variable.

- Ex. Flip a coin. S = {H, T}. We can define a Bernoulli random variable, X(H) = 1, X(T) = 0. Then the distribution of X is P(X = 1) = .5, P(X = 0) = .5

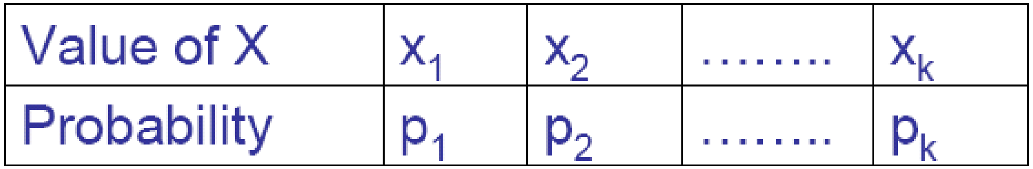
- Ex. Roll a die. S = {1, 2, 3, 4, 5, 6}. We can define a Bernoulli random variable, X(1) = X(2) = 1, X(3) = X(4) = X(5) = X(6) = 0.

  Then the distribution is P(X = 1) = 1/3,P(X = 0) = 2/3

- Notation X(s) = x means that x is the value associated with the outcome s by the random variable X

Probability Mass Function (PMF)

- The probability model for a discrete random variable X, lists its possible values and their probabilities.



Value of X	x_1	x_2	……..	x_k
Probability	p_1	p_2	……..	p_k

- Every probability, p_i, is a number between 0 and 1.

  $p_1+p_2+\ldots+p_k=1$

- The probability distribution or probability mass function (pmf) of a discrete random variable is defined for every number x by $p(x) = P(X=x)$.

- How to check if some function $p(x)$ is a proper PMF?

Random variables and Probability Distribution

- For example: Let X represent the number of heads in 3 tosses.
- We can write down the ***distribution*** of X, which consists of the possible values of X and the probabilities of X taking these values & make a histogram:

outcome: ω	$x = X(\omega)$	$P(X = x)$

- The function describing the distribution is called the ***probability mass function(f(x)),*** $\boldsymbol{f(x) = P(X = x)}$

Example

Ex. Flip three fair coins. (*Binomial*)

S = {HHH, HHT, HTH, HTT, THT, THH, TTH, TTT}. Let's define random variable X to be the number of heads in the experiment, i.e., X(HHH)=3, X(THT)=1, etc.

X	
0	TTT
1	TTH THT HTT
2	THH HTH HHT
3	HHH

Value of X	0	1	2	3
Probability	0.125	0.375	0.375	0.125

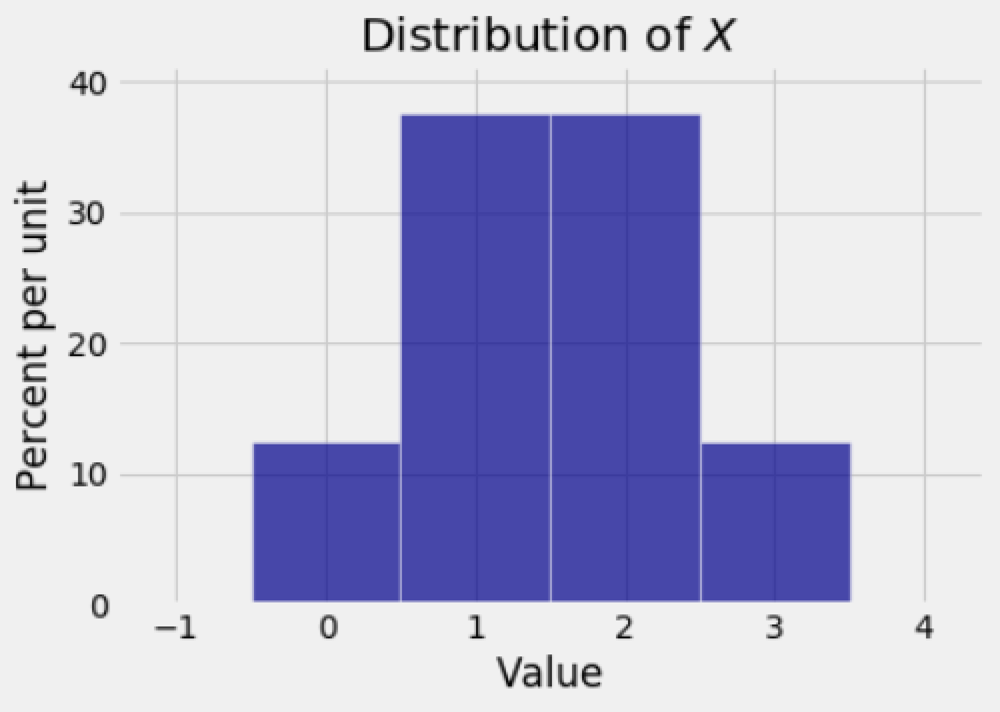
One can calculate the probability of an event by adding the probabilities p_i of the particular values of x_i that make up the event. For example, if we want to know the probability of getting less than 2 heads, we can use

$$P(X<2) = P(X=0) + P(X=1) = 0.125 + 0.375 = 0.5$$

Note: $P(X\leq 2) = P(X=0) + P(X=1) + P(X=2) = 0.875$

Probability Histograms

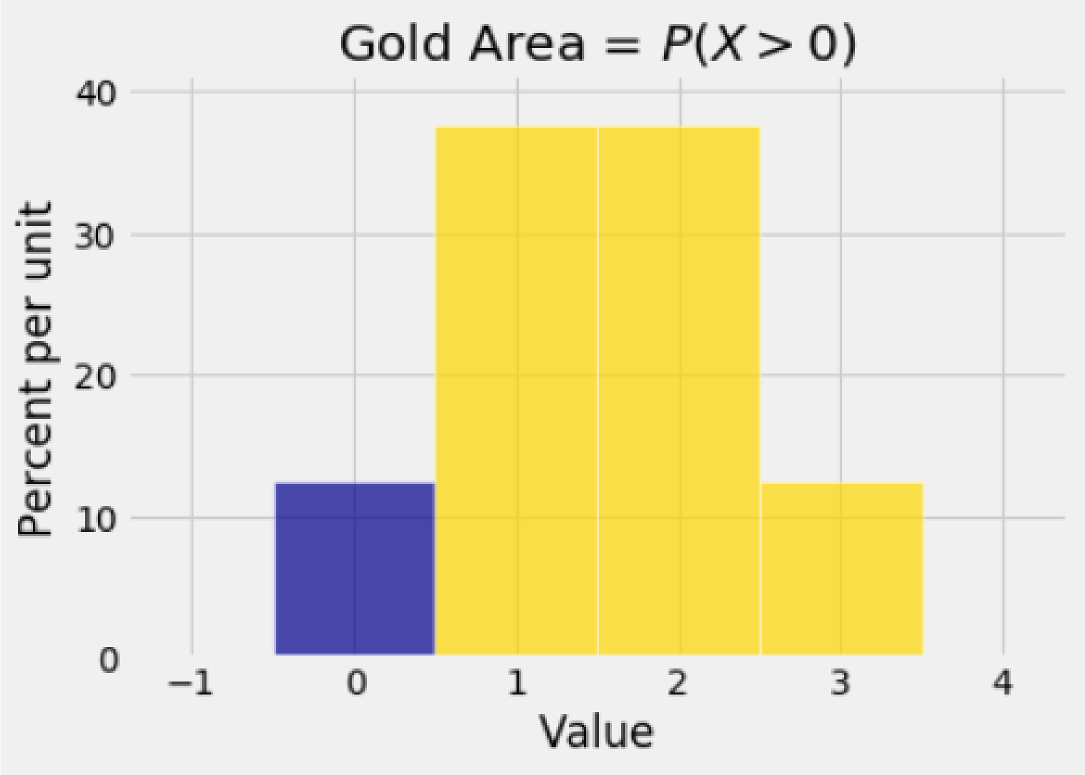
Possible value x	0	1	2	3
$P(X = x)$	1/8	3/8	3/8	1/8



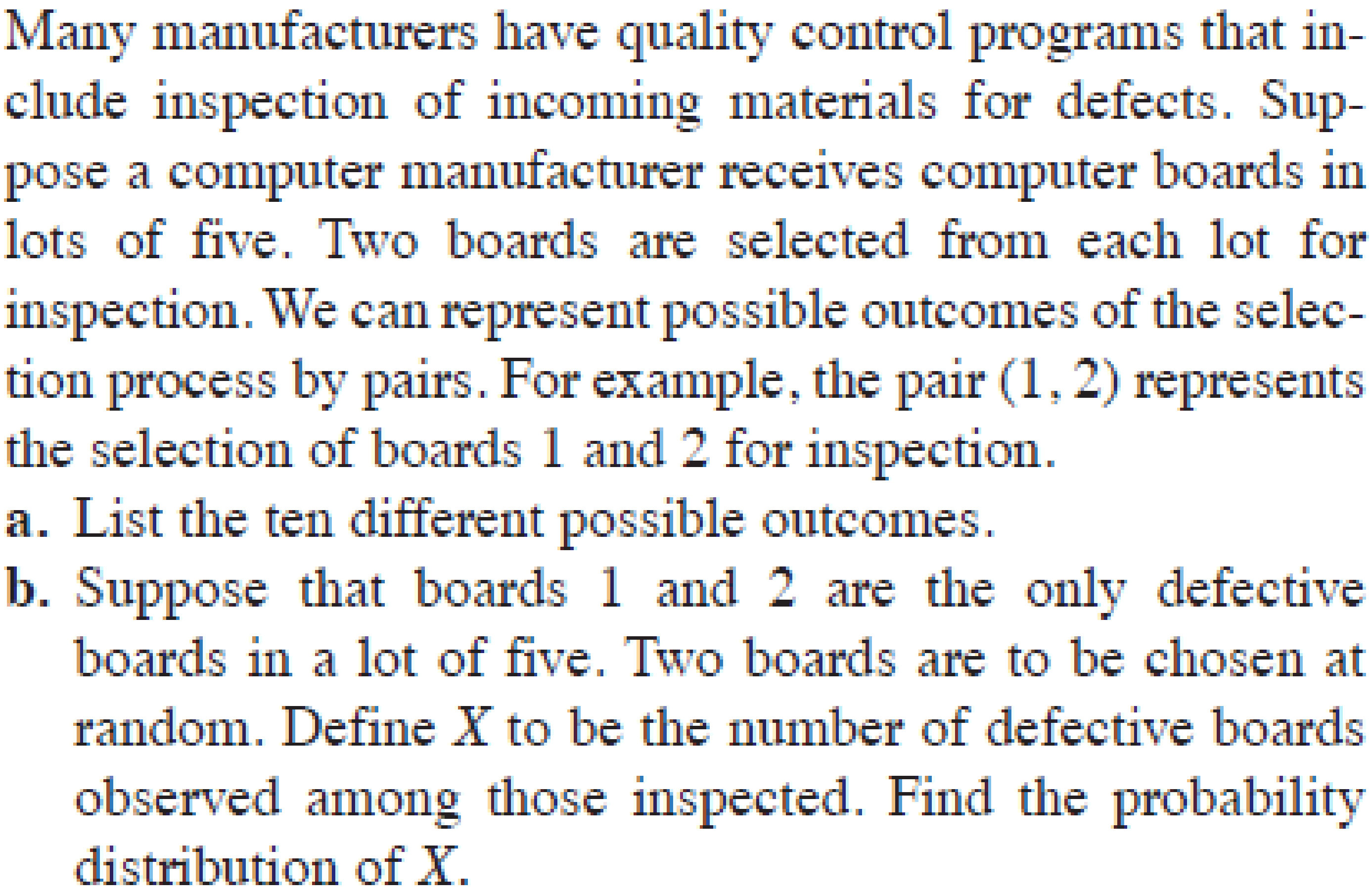
Probability Histograms

$$\begin{aligned} P(X > 0) &= P(X = 1) + P(X = 2) + P(X = 3) \\ &= \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = \frac{7}{8} \end{aligned}$$

$$P(X > 0) = 1 - P(X = 0) = 1 - \frac{1}{8} = \frac{7}{8}$$



Question



Many manufacturers have quality control programs that include inspection of incoming materials for defects. Suppose a computer manufacturer receives computer boards in lots of five. Two boards are selected from each lot for inspection. We can represent possible outcomes of the selection process by pairs. For example, the pair (1, 2) represents the selection of boards 1 and 2 for inspection.

a. List the ten different possible outcomes.

b. Suppose that boards 1 and 2 are the only defective boards in a lot of five. Two boards are to be chosen at random. Define X to be the number of defective boards observed among those inspected. Find the probability distribution of X.

Devore 7th Edition

The Geometric Distribution

- A single Bernoulli trial is usually not all that interesting.
- A **Geometric probability model** tells us the probability for a random variable that counts the number of Bernoulli trials until the first success.
- Geometric models are completely specified by one parameter, p, the probability of success, and are denoted Geom(p).

The Geometric Model

- Independent trials, each having a probability p of being a success, are performed until a success occurs.

Let X = the number of trials required to get a ‘success’

S	p
FS	$(1-p)p$
FFS	$(1-p)^2p$
FFFS	$(1-p)^3p$

$$\mathbf{P(X=x) = q^{x-1}p}$$

Email Example:

If $p=0.20$ $E(X) = 1/.20 = 5$ $P(X=4)$?

The Geometric Model

Geometric probability model for Bernoulli trials: Geom(p)

p = probability of success

$q = 1 - p$ = probability of failure

X = number of trials until the first success occurs

$$\boldsymbol{P(X = x) = q^{x-1}p}$$

Example

Hoops. A basketball player has made 80% of his foul shots during the season. Assuming the shots are independent, find the probability that in tonight's game he

a) misses for the first time on his fifth attempt.
b) makes his first basket on his fourth shot.
c) makes his first basket on one of his first 3 shots.