USC School of Business

Data Driven Decision Making: Probability & Discrete Distributions

GSBA 545, Fall 2021

Professor Dawn Porter

Probability

- Basic Terminology
- Approaches
  - Classical
  - Frequency
  - Subjective
- Random Variables
- Discrete Distributions

Definitions

Probability: Measure of the chance that an experimental outcome will occur when an experiment is carried out. forward looking

Statistics takes observed data & generalizes how the world works, whereas **probability** starts from an assumption about how the world works, and then figures out what data you are likely to see.

Approaches

The three most useful approaches for defining *probability* are:

coin flipping/cards.

→ formulas

1. **Classical approach**

   Gambling and situations where a structure is known

2. **Relative frequency approach**

   Most useful in business situations, as well as many others

   Data is collected over time and compiled to create estimates

stocking decisions.

3. **Subjective approach**

   Used when no reliable data is available

Random Variables

Random variable: A variable that can take on different numerical values determined by the outcome of an experiment.

e.g. people in line at any time
get info about a friend.

Discrete RV: Possible values can be counted

- Number of defective units in a batch of 20
- Listener rating (on a scale of 1 to 5) in a music survey
- Number of children a couple must have in order to have a girl

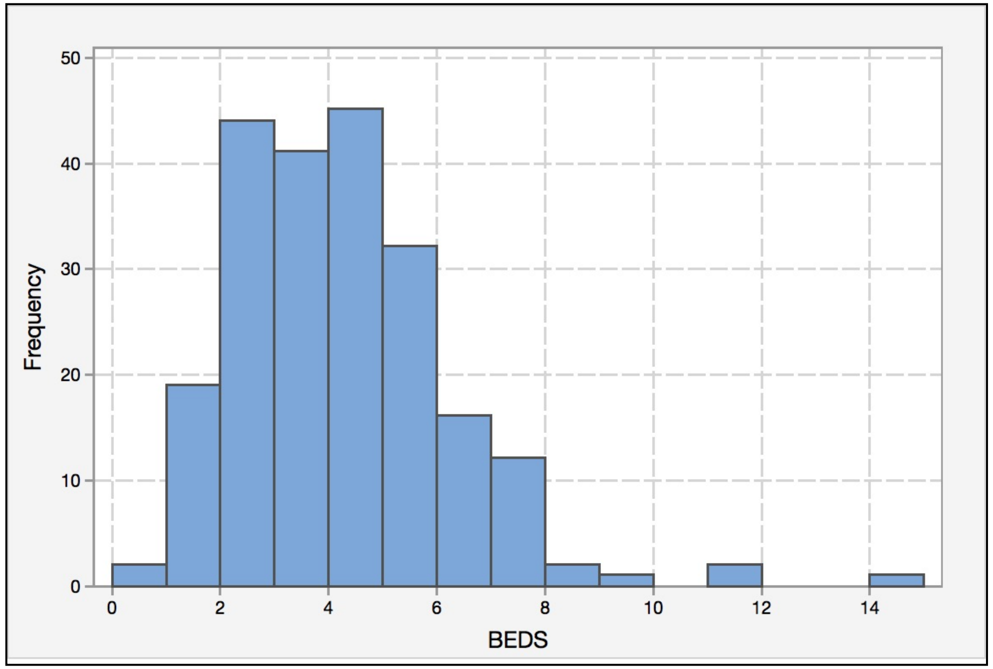
Continuous RV: Number of possible values are almost infinite

- Waiting time for a credit card authorization
- Interest rate charged on a business loan
- Proportion of Super Bowl viewers surveyed who viewed an ad
- Company profits for next year

Discrete Distributions

The **probability distribution** of a discrete random variable is a table or graph giving probabilities associated with each value of the RV.

Eg: Number of Bedrooms in recent LA Real Estate Listings



BEDS	***Count***	***Probability***
0	2	0.009
1	19	0.088
2	44	0.203
3	41	0.189
4	45	0.207
5	32	0.147
6	16	0.074
7	12	0.055
8	2	0.009
9	1	0.005
11	2	0.009
14	1	0.005
	n=	217

Discrete Random Variables (Classical)

How many games will the **World Series** last?

For any "best 4 out of 7" series between two equally matched teams, the duration of the series is a discrete random variable and the probabilities can be calculated using a classical approach*, as the following:

X = Duration	$P(X)$ = Probability
4	0.125
5	0.25
6	0.3125
7	0.3125

Note: Sum of all probabilities MUST = 1 and $0 \leq P(x) \leq 1$.

*The calculations for these are on Blackboard under Misc Docs.

Discrete Random Variables (Frequency)

How many games will the **World Series** last?

Same example as last slide, but this data is the relative frequency of the different options for the series length, collected since 1903.

X = Duration	$P(X)$ = Probability
4	0.174
5	0.239
6	0.220
7	0.367

Note: Sum of all probabilities MUST = 1 and $0 \leq P(x) \leq 1$.

Expected Value

If X is a discrete RV, the **expected value** of X, or $E(X)$, is

$$E(X) = \mu = \sum x \cdot P(x)$$

Weighted average of possible values of X, where the weights are the probabilities of those values occurring.

Long run average value of X over many repetitions.

Mean, or μ, of a population of values of X obtained over many repetitions of the experiment.

Variance & Standard Deviation

If *X* is a discrete RV, the **variance** of *X*, or *Var(X)*, is

$$Var(x) = \sigma^2 = \sum (x - \mu)^2 P(x)$$

Mean Squared Deviation of *X* from its mean, μ, and is rarely, if ever, used in this form.

Standard Deviation of *X* is more common and useful:

$$SD(x) = \sigma = \sqrt{Var(x)}$$

σ measures amount of fluctuation in *X* in the same units as *X*.

Variance & Std Dev: World Series

What are the *expected value* and *standard deviation* of the number of games the **World Series** will last?

x	$P(x)$	$x \cdot P(x)$	$(x-\mu)^2 \cdot P(x)$
4	0.174	$4 \cdot (0.174) = 0.697$	$(4-5.780)^2 \cdot 0.174 = 0.552$
5	0.239	$5 \cdot (0.239) = 1.193$	$(5-5.780)^2 \cdot 0.239 = 0.145$
6	0.220	$6 \cdot (0.220) = 1.321$	$(6-5.780)^2 \cdot 0.220 = 0.011$
7	0.367	$7 \cdot (0.367) = 2.569$	$(7-5.780)^2 \cdot 0.367 = 0.546$

$$E(X) = \mu = \sum x \cdot P(x) = 5.780$$

$$Var(X) = \sigma^2 = \sum (x-\mu)^2 P(x) = 1.254$$
$$SD(X) = \sigma = \sqrt{1.254} = 1.120 \text{ games}$$

Discrete Random Variables: Car Sales

What is the expected number of **automobiles sold** during a day at Dicarlo Motors?

300 days of data with sales and frequencies are calculated:

X = Cars Sold	$P(X)$ = Probability
0	0.180
1	0.390
2	0.240
3	0.140
4	0.040
5	0.010

Variance & Std Dev: Car Sales

What are the *expected value* and *standard deviation* of the number of cars sold in any given day at DiCarlo Motors?

x	$P(x)$		
0	0.180		
1	0.390		
2	0.240		
3	0.140		
4	0.040		
5	0.010		

USC School of Business

Discrete Random Variables: ATM lines

Is the number of ATM machines sufficient at Citicard Banking Centers (CBCs)?

The **waiting times** for an ATM are calculated and the number of customers (*x*) arriving in a one-minute period is distributed as:

X = People waiting	*P(X)* = Probability
0	0.135
1	0.271
2	0.271
3	0.180
4	0.090
5 or more	0.053

*The waiting times are assumed to be distributed as a Poisson RV.

Variance & Std Dev: ATM lines

What are the *expected value* and *standard deviation* of the number of people waiting in line for an ATM machine?

x	$P(x)$		
0	0.135		
1	0.271		
2	0.271		
3	0.180		
4	0.090		
5 or more	0.053		

Example: CheapO Profits

CheapO Profits. CheapO Computers shipped two servers to its biggest client. Four refurbished computers were mistakenly restocked along with 11 new systems.

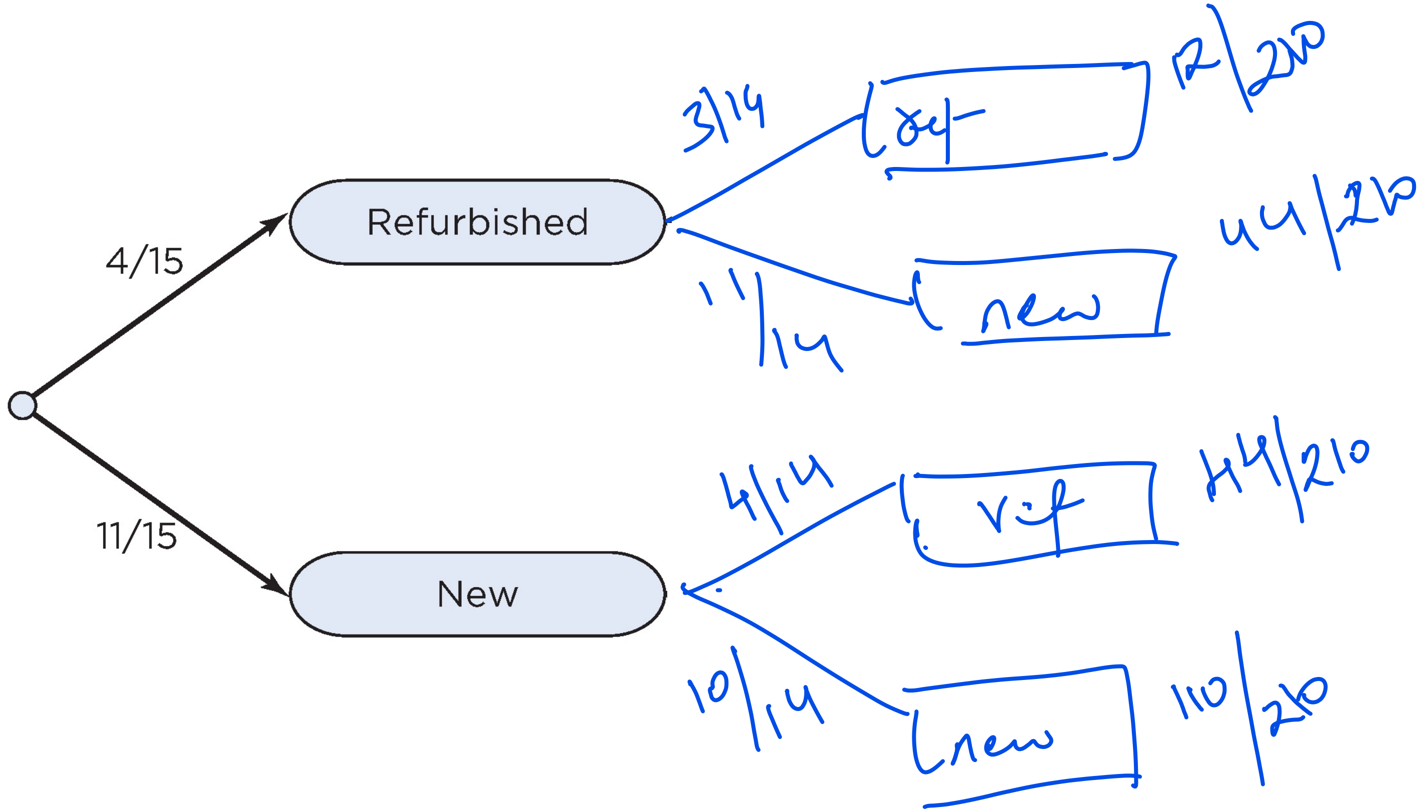
If the client receives two new servers, the profit for the company is \$10,000; if the client receives one new server, the profit is \$4,500. If the client receives two refurbished systems, the company loses \$1000.

What are the expected value and standard deviation of CheapO's profits?

USC School of Business

Example: CheapO Profits

CheapO Profits. CheapO Computers shipped two servers to its biggest client. Four refurbished computers were mistakenly restocked along with 11 new systems.



Example: CheapO Profits

CheapO Profits. What are the expected value and std deviation of CheapO's profits?

	x
Both refurbished	–\$ 1000
One refurbished	\$ 4500
New/New	\$10,000

Example: CheapO Profits

CheapO Profits. What are the expected value and std deviation of CheapO's profits?

	x	$P(x)$
Both refurbished	–$ 1000	12/210 = 0.057
One refurbished	$ 4500	88/210 = 0.419
New/New	$10,000	110/210 = 0.534

Example: CheapO Profits

CheapO Profits. What are the expected value and std deviation of CheapO's profits?

	x	$P(x)$	$x \cdot P(x)$
Both refurbished	–$ 1000	12/210 = 0.057	– 57.14
One refurbished	$ 4500	88/210 = 0.419	1885.71
New/New	$10,000	110/210 = 0.524	5238.10

$\boxed{\mu = \$7067}$

Example: CheapO Profits

CheapO Profits. What are the expected value and std deviation of CheapO's profits?

	x	$P(x)$	$(x-\mu)^2 P(x)$
Both refurbished	–\$ 1000	0.057	3,709,359.87
One refurbished	\$ 4500	0.419	2,760,995.81
New/New	\$10,000	0.524	4,593,729.13

$\sigma^2 = 10{,}986{,}032\ \$^2$

$\sigma = \$3315$

On *average,* this is very profitable… but what does the large standard deviation indicate here?

Example: Raffle Tickets

Five thousand raffle tickets are to be sold at $10 each to benefit a local community group. The prizes, the number of each prize to be given away, and the dollar value of winnings for each prize are as follows:

Prize	Number given away	Dollar value
Automobile	1	$ 13,000
Entertainment center	2	$ 3,000 each
DVD recorder	5	$ 400 each
Gift certificate	50	$ 20 each

a) Create a distribution table for the *profits* and probabilities.

b) What are the *expected winnings* from the purchase of one ticket?