BUAN 3500: Data Visualization and Descriptive Analytics

Probability: An Introduction to Modeling Uncertainty

Lauren M. Nelsen, Ph.D.

University of Colorado Colorado Springs

What is probability?

Definition

Probability is the numerical measure of the likelihood that an event will occur.

In business scenarios, we often need to make decisions where the outcome is not known with certainty.

But we can use data that is available to provide information on possible outcomes in order to better understand the likelihood of certain decisions.

Events and Probabilities

Definition

- **random experiment:** a process that generates well-defined outcomes
- **sample space:** set of all possible outcomes

Example: random experiment: flip a coin
$\rightarrow$ What is the sample space?

$$S = \{\text{Head}, \text{Tail}\}$$

Example: random experiment: roll a die
$\rightarrow$ What is the sample space?

$$S = \{1, 2, 3, 4, 5, 6\}$$

Definition

- **event:** collection of outcomes

Events and Probabilities

Example

The California Power & Light Company (CP&L) is starting a project designed to increase the generating capacity of one of its plants in Southern California.

An analysis of similar construction projects indicated that the possible completion times for the project are 8, 9, 10, 11, and 12 months.

- Let C denote the event that the project is completed in 10 months or less:
$$C = \{8, 9, 10\}$$
- Let L be the event that the project is completed in less than 10 months:
$$L = \{8, 9\}$$
- Let M be the event that the project is completed in more than 10 months:
$$M = \{11, 12\}$$

Events and Probabilities

The table below gives the past completion times for 40 CP&L projects.

- Calculate the probability of the different outcomes.
- Calculate the probabilities of the events on the previous slide:
  - $C = \{8, 9, 10\}$
  - $L = \{8, 9\}$
  - $M = \{11, 12\}$

(Feel free to do these in Excel!)

Completion Time (months)	**No. of Past Projects Having This Completion Time**	**Probability of Outcome**
8	6	
9	10	
10	12	
11	6	
12	6	
Total		

Events and Probabilities

Events and Probabilities

Let's Make a Deal (also known as ***LMAD***) is a television game show that originated in the United States in 1963 and has since been produced in many countries throughout the world. The program was created and produced by Stefan Hatos and Monty Hall, the latter serving as its host for nearly 30 years.

The format of *Let's Make a Deal* involves selected members of the studio audience, referred to as "traders", making deals with the host. In most cases, a trader will be offered something of value and given a choice of whether to keep it or exchange it for a different item. The program's defining game mechanism is that the other item is hidden from the trader until that choice is made. The trader thus does not know if they are getting something of equal or greater value or a prize that is referred to as a "zonk," an item purposely chosen to be of little or no value to the trader.

(This is taken from the Wikipedia page for "Let's Make a Deal".)

Events and Probabilities

Monty Hall Problem:

https://youtu.be/4Lb-6rxZxx0?si=vKnK2AwGGi21IWCy

After watching the beginning of the video, let's take a poll to see how many of us would stick with our original choice and how many of us think we should switch:

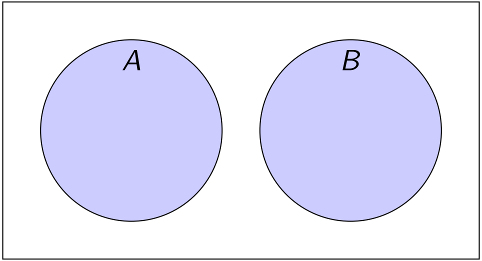
Join by Web **PollEv.com/laurennelsen965** Join by Text Send **laurennelsen965** to **37607**

Some Basic Relationships of Probability

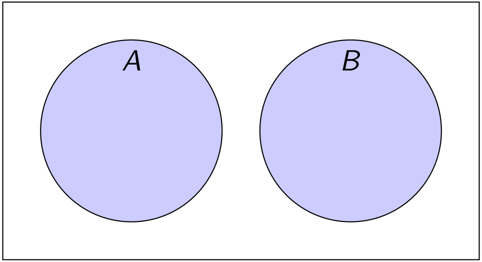
In the California Power & Light Company (CP&L) example, when we found the probability of the event $L = \{8, 9\}$, we added up two probabilities:

$$P(L) = P(\{8\}) + P(\{9\})$$

We could do this because the events $\{8\}$ and $\{9\}$ are **mutually exclusive** (or **disjoint**).



Some Basic Relationships of Probability



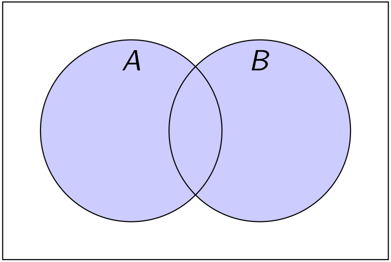
Addition Law for Mutually Exclusive Events: If A and B are mutually exclusive, then

$$\boxed{P(A \cup B) = P(A) + P(B)}$$

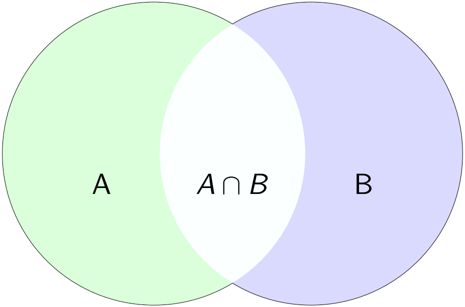
(where the **union of** A **and** B, $A \cup B$, is the event containing all outcomes belonging to A or B or both).

Some Basic Relationships of Probability

What if we have a situation like this? How would we find $A \cup B$?



Some Basic Relationships of Probability



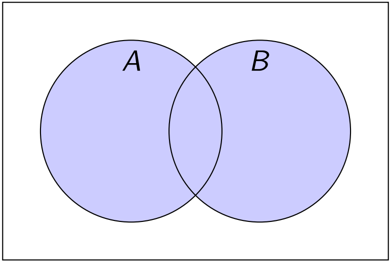
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Some Basic Relationships of Probability

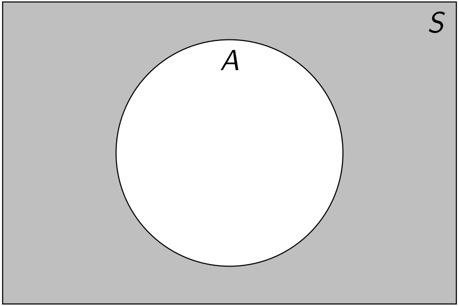
Addition Law:

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

(where the **intersection of** A **and** B, $A \cap B$, is the event containing all outcomes belonging to both A and B).



Some Basic Relationships of Probability



Computing Probability Using the Complement:

$$P(A) = 1 - P(A^C)$$

(where A^C is the **complement** of A).

Some Basic Relationships of Probability

Example

The HR manager of a software company conducted a study and found the following:

- 30% of the employees who left the firm within two years listed dissatisfaction with their salary as one of the reasons
- 20% listed dissatisfaction with their work assignments as one of the reasons
- 12% listed dissatisfaction with both their salary and their work assignments.

What is the probability that an employee who leaves within two years does so because of dissatisfaction with salary, dissatisfaction with the work assignment, or both?

Let

- $S =$ the event that the employee leaves because of salary
- $W =$ the event that the employee leaves because of work assignment

Some Basic Relationships of Probability

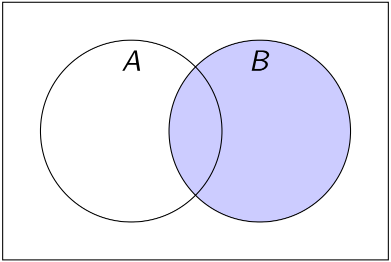
$$\begin{aligned} P(S \cup W) &= P(S) + P(W) - P(S \cap W) \\ &= 0.30 + 0.20 - 0.12 \\ &= 0.38 \end{aligned}$$

Probability an employee will leave for salary or work assignment reasons is 0.38.

Conditional Probability

Definition

The probability of event A *given* that event B has occurred is called the **conditional probability of** A **given** B and is denoted by $P(A|B)$.



Conditional Probability

Example

Open the “mortgagedefaultdata” file in Excel. This data is from a bank that is interested in the mortgage default risk for its home mortgage customers. Some of these customers have defaulted on their mortgages and others have continued to make on-time payments.

The data includes the age of the customer at the time of the mortgage origination, the marital status of the customer, the annual income of the customer, the mortgage amount, the number of payments made by the customer per year on the mortgage, the total amount paid by the customer over the lifetime of the mortgage, and whether or not the customer defaulted on the mortgage.

Create a PivotTable for the marital status and whether the customer defaults on their mortgage.

Conditional Probability

Using the PivotTable you just made, calculate the following probabilities:

1. the probability a customer is single and defaults
2. the probability a customer is married and does not default
3. the probability a customer who is married will default
4. the probability a customer is single if you know they do not default

Conditional Probability

Conditional Probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

or

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Conditional Probability

Independent Events: Two events are independent if $P(A|B) =$....

$$P(A|B) = P(A)$$

Multiplication Law:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(B)P(A|B)$$

or

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \Rightarrow P(A \cap B) = P(A)P(B|A)$$

Multiplication Law for Independent Events:

$$P(A \cap B) = P(A)P(B)$$

Conditional Probability

Example

If we look back at previous example for mortgages, we can see that

$$P(M) = 0.4766$$

and

$$P(D|M) = 0.5524.$$

Using this information, what is the probability that a customer defaults on their mortgage and is married?

$$P(D \cap M) = P(M)P(D|M) = (0.4766)(0.5524) = 0.2633$$

The multiplication law is helpful if we know conditional probabilities but do not know the joint probabilities (probability of intersection).

Conditional Probability

Bayes' Theorem:
Revising probabilities when new information is obtained is an important aspect of probability analysis.

- We often begin our analysis with initial or **prior probability** estimates for specific events of interest.
- Then from sources (such as a sample survey or product test) we obtain additional information about the events.
- Given this new information, we update the prior probability values by calculating revised probabilities, referred to as **posterior probabilities**.

Bayes' Theorem gives us a way to make these probability calculations.

Conditional Probability

Example

Consider a manufacturing firm that receives shipments of parts from two different suppliers.

- A_1: the event that a part is from supplier 1
- A_2: the event that a part is from supplier 2

Currently 65% of the parts purchased by the company are from supplier 1 and the remaining 35% are from supplier 2.

The quality of the parts varies according to their source.

- G: the event that a part is good
- B: the event that a part is bad

Historical data suggests the following:

$$P(G|A_1) = 0.98 \qquad P(B|A_1) = 0.02$$

$$P(G|A_2) = 0.95 \qquad P(B|A_2) = 0.05$$

Conditional Probability

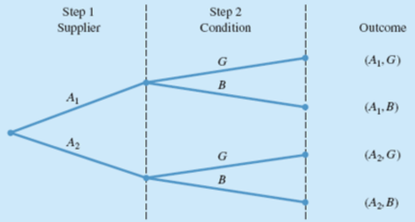
Example

(Continued)

$$
\begin{aligned}
P(A_1, G) = P(A_1 \cap G) &= P(A_1)P(G|A_1) \\
&= (0.65)(0.98)
\end{aligned}
$$

Conditional Probability

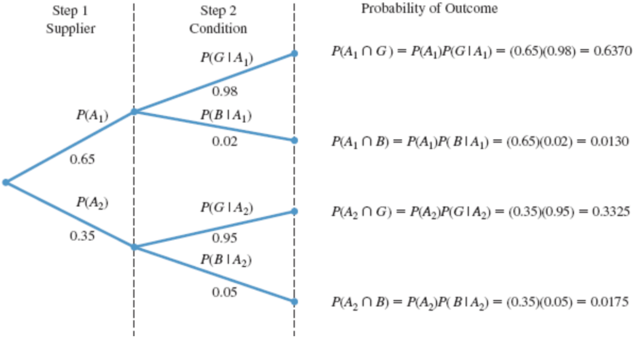
Figure 4.7 **Diagram for Two-Supplier Example**



Note: Step 1 shows that the part comes from one of two suppliers and Step 2 shows whether the part is good or bad.

Conditional Probability

Figure 4.8 **Probability Tree for Two-Supplier Example**



Conditional Probability

Theorem

Bayes' Theorem (Two-Event Case):

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)}$$

$$P(A_2|B) = \frac{P(A_2)P(B|A_2)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)}$$

Conditional Probability

Example

(Continued)
Now suppose that the parts from the two suppliers are used in the firm's manufacturing process and that a machine breaks down while attempting the process using a bad part. Given the information that the part is bad, what is the probability that it came from supplier 1 and what is the probability that it came from supplier 2?

- $P(A_1|B)$:

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)}$$
$$= \frac{(.65)(.02)}{(.65)(.02) + (.35)(.05)}$$
$$\approx 0.426$$

- $P(A_2|B)$: Should get ≈ 0.574

Conditional Probability

Theorem

Bayes' Theorem:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)}$$

Random Variables

Definition

random variable: numerical description of the outcome of a random experiment

- **discrete random variable:** random variable that can take on only specified discrete values
- **continuous random variable:** random variable that can take on any numerical value in an interval or collection of intervals

Random Variables

Examples of Discrete Random Variables:

- **Random Experiment:** Contact 5 customers
  - **Random Variable:** Number of customers who place an order
  - **Possible Values for the RV:** $0, 1, 2, 3, 4, 5$
- **Random Experiment:** Offer a customer a choice of two products
  - **Random Variable:** Product chosen by customer
  - **Possible Values for the RV:** 0 (if none chosen), 1 (if choose product A), 2 (if choose product B)

Random Variables

Examples of Continuous Random Variables:

- **Random Experiment:** Customer visits a website
  - **Random Variable (x):** Time customer spends on the website in minutes
  - **Possible Values for the RV:** $x \geq 0$
- **Random Experiment:** Invest $\$10,000$ in the stock market
  - **Random Variable (x):** value of investment after one year
  - **Possible Values for the RV:** $x \geq 0$

Random Variables

We want to understand *distributions* of random variables.

The way we do this will depend on if our random variables are continuous or discrete.

Discrete Probability Distributions

Definition

- The **probability distribution** for a random variable describes the range and relative likelihood of possible values for a random variable.

- For a discrete random variable, x, the probability distribution is defined by a **probability mass function**.
  - denoted by $f(x)$
  - gives the probability for each value of the random variable

Discrete Probability Distributions

Example

Let's consider a mortgage provider that has information on customers who default on their mortgages. Customers can make monthly payments, two payments per month, or quarterly payments.

Consider the random variable describing the number of payments made per year by the customers.

Number of payments per year	**Number of observations**	**f(x)**
$x = 4$	45	0.15
$x = 12$	180	0.60
$x = 24$	75	0.25

Discrete Probability Distributions

Number of payments per year	Number of observations	f(x)
$x = 4$	45	0.15
$x = 12$	180	0.60
$x = 24$	75	0.25

Probability distribution/"probability mass function":

$$f(x) = \begin{cases} 0.15 & \text{if } x = 4 \\ 0.60 & \text{if } x = 12 \\ 0.25 & \text{if } x = 24 \\ 0 & \text{otherwise} \end{cases}$$

Discrete Probability Distributions

Question: Someone offers you the opportunity to play the following game:

- You pay \$5 to play the game.
- You roll a six-sided die.
  - If you roll a 6, you win \$25.
  - If you roll anything other than a 6, you don't win anything.

Should you play the game?

Discrete Probability Distributions

Definition

Expected Value of a Discrete Random Variable:
The **expected value** (or **mean**) of a random variable is a measure of the central location of the random variable

$$E(x) = \mu = \sum xf(x)$$

Example

Let's go back to the mortgage example. What is the expected value of the number of payments made by a mortgage customer in a year?

Discrete Probability Distributions

Measures of variability in the values of a random variable:

- **Variance:** $\text{Var}(x) = \sigma^2 = \sum (x - \mu)^2 f(x)$
- **standard deviation:** $\sigma = \sqrt{\sigma^2}$

Example

Let's look at the mortgage example and calculate the variance and the standard deviation of the number of payments per year.

Example

A computer company is considering a plant expansion to enable the company to begin production of a new computer product. The company's president must determine whether to make the expansion a medium- or large-scale project. Demand for the new product is uncertain, which for planning purposes may be low demand, medium demand, or high demand. The probability estimates for demand are 0.2, 0.5, and 0.3, respectively. Letting x and y indicate the annual profit in thousands of dollars, the firm's planners developed the following profit forecasts for the medium- and large-scale expansion projects:

(Continued on next slide)

Example

(Continued)

		Medium-Scale Expansion Profit		Large-Scale Expansion Profit	
		x	$f(x)$	y	$f(y)$
Demand	**Low**	50	0.20	0	0.20
	Medium	150	0.50	100	0.50
	High	200	0.30	300	0.30

1. Compute the expected value for the profit associated with the two expansion alternatives. Which decision is preferred for the objective of maximizing the expected profit?
2. Compute the variance for the profit associated with the two expansion alternatives. Which decision is preferred for the objective of minimizing the risk or uncertainty?

Discrete Probability Distributions

Definition

Discrete Uniform Probability Mass Function:

When the possible values of $f(x)$ are all equal, the probability distribution is called a **discrete uniform probability distribution.**

$$f(x) = \frac{1}{n}$$

Discrete Probability Distributions

Now we want to consider other discrete probability distributions that are often involved in real-world problems.

Before we introduce the next distribution, consider the following problem:

Example

Consider a basketball player whose probability of making a free throw is 0.7. If they shoot 5 free throws, what is the probability they make exactly 3 of them?

$$\binom{5}{3}(0.7)^3(0.3)^2$$

This is an example of a binomial probability distribution!

Discrete Probability Distributions

Discrete Probability Distributions Worksheet

Continuous Probability Distributions

Continuous Probability Distributions Worksheet