Set 1 - Basic Terminology and Concepts

January 7, 2023 6:44 PM

Stat 260 Lecture Notes
Set 1 - Basic Terminology and Concepts

Statistics: Is the development and application of methods to collect, analyze, and interpret data.

Stats teaches us how to make intelligent judgments and informed decisions in the the presence of uncertainty and variation.

Definitions:

- **population:** a collection of objects big group we're interested in
- **sample:** a selection from the population take a few from that group
- **parameter:** a descriptive measure of the population (e.g. mean (μ), standard deviation (σ), variance (σ^2), etc.) comes from population
- **statistic:** a descriptive measure of the sample (e.g. mean ($\bar{x}$), standard deviation (s), variance (s^2), etc.) comes from sample
- **random variable (r.v.):** a characteristic that changes from object to object in the population
  - discrete r.v. - set of all possible values are finite or are infinite and countable (they can be listed in a finite or infinite sequence)
  - continuous r.v. - set of all possible values are infinite and it's impossible to list all the possible values (e.g. it's impossible to list all the decimals between 0 and 0.5.)

parameter/statistic: what we're calculating

→ eye colour - answer changes from person to person
→ thing that we're measuring

How many classes? 0, 1, 2, 3, 4, 5, 6 discrete r.v.
↳ b/c can list all of them
0, 0.1, 0.2, 0.3, 0.4, ... discrete

0 0.1 0.2 0.3 0.4 0.5
0.01
0.001
??

continuous r.v.

1

discrete r.v. ⇔ finite decimal places
continuous r.v. ⇔ infinite decimal places

⇒ think about it theoretically

first written assignment posted
(can do first question)

Stat 260 consists of three main topics:

- **Descriptive Stats (Sets 1-3):** Ways to describe the data set (e.g. charts/graphs, mean (average), median, standard deviation, variance, etc.)
- **Probability (Sets 4-21):** How likely events are to occur. We look at the population to see how likely events will be in the sample.
- **Inferential Stats (Sets 22-31):** Use the sample to make generalizations about the population.

Example 1: Suppose we wish to look at the average height of adults who live in Canada. To do this we create a poll on our Stat 260 Brightspace page and ask the students in the class to report their height in cm.

→ socioeconomic, gender, race, age, sample size

- population: adults who live in Canada
- sample: Stat 260 students who answer
  ↳ is subset of population
- parameter: average height of adults who live in Canada
  Notation: μ
- statistic: average height of Stat 260 students who answer the poll
  Notation: $\bar{x}$
- random variable: height (in cm)
  ↳ the thing that's being measured
  ↳ continuous r.v.
  (b/c theoretically height could have

Notation: ...

- random variable: height (in cm)
  ↳ the thing that's being measured
  ↳ continuous r.v.
  (b/c theoretically height could have infinite decimals)

Example 2: Suppose we wish to look at the average number of classes current UVic students are registered in this semester. To do this we randomly select 200 students who are registered this semester and from their UVic profile record the number of classes they are currently registered in. → good set up

- population: current UVic students
- sample: 200 UVic students measured
- parameter: average # of classes of current UVic students
  Notation: μ
- statistic: average # of classes of 200 UVic student sample
  Notation: $\bar{x}$
- random variable: number of classes
  ↳ discrete r.v.
  (all whole numbers, finite decimal)

Example 3: Determine if each of the following are discrete or continuous random variables.

(a) A person's height in *cm*. continuous

(b) The number of courses a student is registered in this semester. discrete

(c) The top running speed of greyhound dogs in km/h. continuous

(d) The price of a cup of coffee in dollars. discrete (only 2 decimal places) → price you have to pay is discrete
↳ but might change based on context

(e) The number of accidents at a particular intersection on a highway in a year. discrete

(f) The lowest temperature in the month. continuous
↳ b/c theoretically infinite

discrete/continuous don't change how you calculate things but changes for probability

Set 2 - Basic Descriptive Statistics

January 7, 2023 6:44 PM

Stat 260 Lecture Notes
Set 2 - Basic Descriptive Statistics

Knowing what type of data we have affects what can be done to analyze the data. Univariate data is collected from single measurements on subjects.

average height vs eye colour

Suppose we take a sample with n observations $x_1, x_2, x_3, \ldots, x_n$. Each x_i represents the single value that was measured in that observation. This is univariate data (i.e. we measured only one number per each observation). We can display univariate data with a frequency table or a histogram or a boxplot.

Example 1: Measurements of lengths of lizards captured in months of August and October.

```
August measurements:
[1] 7.5 7.2 3.0 12.1 15.1 12.1 11.5 11.8 7.2 13.2 13.6 8.2 9.5 8.4 13.3
[16] 12.5 12.4 2.1 10.7 9.4 6.7 6.8 6.1 8.3 7.9 6.0 7.6 13.2 4.5 9.3
[31] 8.1 3.5 9.0 50.0
```

1st observation

31st observation

$n=34$

```
October measurements:
[1] 43.7 37.2 29.0 31.6 47.5 48.3 38.3 19.7 32.5 45.2 36.1 30.5 37.2 50.5 36.9
[16] 44.5 35.9 28.7 37.5 30.2 36.9 43.2 27.0 26.2 41.8 26.4 34.3 28.6 35.9 22.0
[31] 45.4 30.3 29.8 46.1 42.7 31.5 37.4 25.1 27.2 45.0
```

Look at the August measurements. Here we can create a frequency table.

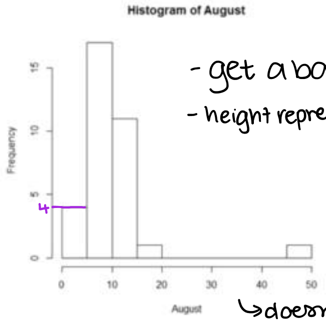
count how many

convert into percentage

* 3-4 decimal places

Interval	Frequency	Relative Frequency
equal → [0, 5)	IIII	4/34 = 0.12 = 12%
[5, 10)	~~IIII~~ ~~IIII~~ ~~IIII~~ II	17/34 = 0.5 = 50%
[10, 15)	~~IIII~~ ~~IIII~~ I	11/34 = 0.32 = 32%
[15, 20)	I	1/34 = 0.03 = 3%
[20, 25)	0	0
[25, 30)	0	0
[30, 35)	0	0
[35, 40)	0	0
[40, 45)	0	0
[45, 50]	I	1/34 = 0.03 = 3%
	$n=34$	1 = 100%

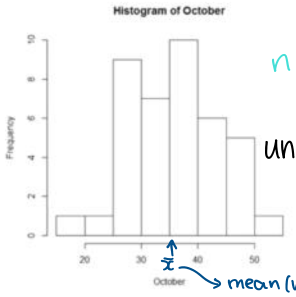
Using this frequency table we can create a histogram. The histogram for the August data looks like this:



- get a bar for each interval
- height represents how many things in interval

↳ doesn't tell you [or)

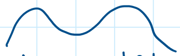
Doing a similar thing for the October data gives a histogram that looks like this:



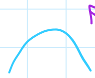
n = sum of heights of all bars

unimodal

↳ mean (vertical weights)



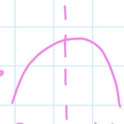
bimodal



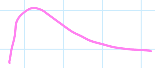
unimodal

modality : which is most common (tallest bar) mode

Skewness



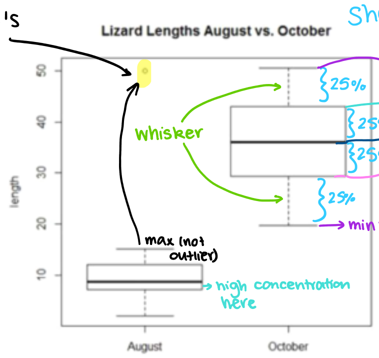
Symmetric



Skewed

Skewed

It can be difficult to compare two data sets like this, so we could also use a boxplot. The boxplots for the August and October data look like this:



With this boxplot it is easy to see that October lengths are larger than August lengths.

Categorical data occurs when our recorded data falls into categories. (e.g. eye colour, program major, customer satisfaction). While we cannot do many of the calculations with categorical data that we can do with univariate data, we can display categorical data with a bar chart.

for categories instead of numbers

Outlier: number that doesn't fit pattern of what other numbers are doing

→ the 50 from August is an outlier

→ outliers are part of data set, so max for August would still be 50

Suppose we have a sample with observations $x_1, x_2, \ldots, x_n$.

- *sample mean*, $\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i$. summation The mean is the same thing as the average.
- *sample median*, $\tilde{x}$, → tilde is the the middle observation. If there is an even number of observations then there are two middle observations so the median is the midpoint (here, the same as the average) of these two values. Remember that the data must be sorted first!
- *mode*, is the most common value. There may be many modes if many values are tied for occurring the most often.

Example 2: Suppose we have the sample:

$$0, 0, 2, 3, 6, 7, 10, 11, 20.$$

Calculate the mean, median, and mode.

mean: $\bar{x} = \frac{0+0+2+3+6+7+10+11+20}{9} = \frac{59}{9}$ leave as reduced fraction

median: $\tilde{x} = 6$

mode = 0

In general: 3-4 decimals

1 more decimal than the data

Example 3: Suppose we have the sample:

$$1, 1, 2, 8, 15, 15, 25, 100.$$

Calculate the mean, median, and mode.

mean: $\bar{x} = \frac{1+1+2+8+15+15+25+100}{8} = \frac{167}{8} = 20.875$ ← exact

median: $\tilde{x} = 11.5 = \frac{8+15}{2}$

mode = 1 and 15

Without 100 : $\frac{67}{7}$ = mean = 9.57

median = 8

mean sensitive to outliers

Mode is useful when we're talking about categories

Notes:

- $\bar{x}$ doesn't have to be a value actually observed in the sample. Neither does $\tilde{x}$. (So don't round your final answers for average or median, even if they physically aren't possible for a single observation.)
- $\bar{x}$ is sensitive to outliers (extreme values), $\tilde{x}$ is not. Look at what happens in Example 3 if the value of 100 is removed. The mean of the new data list changes quite a bit, the median of the new data list changes very little.
- The median splits the data in two: 50% of the observations are larger than $\tilde{x}$, and 50% of the observations are smaller than $\tilde{x}$.
- The mean and median tell us where the "center" of our sample data is.
- The mode tells us which observation is most common. The mode can be used for categorical data, whereas $\bar{x}$ and $\tilde{x}$ cannot be used for categorical data.
- In a histogram, the mean (average) occurs at the "balance point" of the picture. ↳ mean is the balance point (almost like the point of equilibrium)

Example 4: Look at the following two samples. How could we describe the difference in these samples?

sample 1:	**sample 2:**
10	10
20	48
49	49
50	50
51	51
80	52
90	90

further out from the centre

min and max are both same

Both samples have $\overline{x} = \tilde{x} = 50$. Which sample is more spread out?

mean median

We look at *measures of variability* to describe differences between these samples.

variance & standard deviation

- *sample variance*, $s^2 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$
- *sample standard deviation*, $s = \sqrt{\text{variance}}$
- *shortcut formula for sample variance*, $s^2 = \underbrace{\dfrac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \dfrac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]}_{\text{looks longer but easier calculations}}$

Note: The standard deviation s and the mean $\overline{x}$ have the same units as your original data. The variance s^2 uses $(\text{units})^2$.

Standard deviation : same units (kind of like an error measurement)

variance : $(\text{units})^2$

6

How far is data from the mean?

$$S^2 = \frac{\sum_{i=1}^{n} \overbrace{(x_i - \overline{x})^2}^{\text{distance of observation from mean}}}{n-1}$$

(if it wasn't squared you'd always get 0)

→ course has a formula sheet (for exams)

If only 1 data point then variance = 0

Single number doesn't have a variance but outlier contributes to variance

Example 5: Calculate the sample variance for sample 1 and sample 2 in Example 4. Calculate the sample standard deviation for sample 1 in Example 4.

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$$

Sample 1: 10, 20, 49, 50, 51, 80, 90

$n = 7$ $\bar{x} = 50$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
10	-40	1600
20	-30	900
49	-1	1
50	0	0
51	1	1
80	30	900
90	40	1600
	$\Sigma(x_i - \bar{x}) = 0$	5002

will always get 0 here

$$s^2 = \frac{5002}{7-1} = \frac{5002}{6} = 833.7$$

Sample 2: 10, 48, 49, 50, 51, 52, 90 $n = 7$

$$s^2 = \frac{1}{n-1}\left[\Sigma x_i^2 - \frac{1}{n}(\Sigma x_i)^2\right]$$

Shortcut formula

$\Sigma x_i = 350$

$\Sigma(x_i^2) = 10^2 + 48^2 + \ldots + 90^2 = 20{,}710$

$$s^2 = \frac{1}{7-1}\left[20{,}710 - \frac{1}{7}(350)^2\right] = 535$$

Sample 1: $s^2 = 833.7$

Sample 2: $s^2 = 535$

Variance from sample 1 is larger, so sample 1 is more spread out from the mean.

The numbers on their own don't mean much, useful for comparing them

Standard deviation of sample 1: $s = \sqrt{\text{variance}} = \sqrt{833.7} = 28.874$

Sample values are estimates of the population

Note: The **population variance** σ^2 is calculated using a slightly different formula than the sample variance. Suppose our population has N items and we take a sample of n items from it.

- *sample variance*, $s^2 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$ sample mean dividing by n-1 gives the better estimate
- *population variance*, $\sigma^2 = \dfrac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$ population mean for population you divide by N

In the sample, we don't know the whole picture so the sample variance s^2 is just an estimate of the population variance σ^2. As it turns out (due to technical reasons and a definition we won't cover here) that dividing by $n-1$ gives a better estimate of the population when we perform the sample variance calculation. Dividing by n gives an estimate that underestimates.

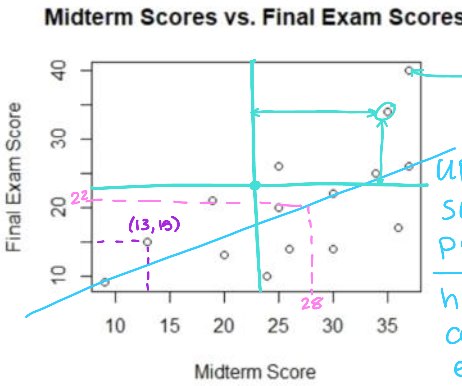
Lastly, the quickest way to calculate standard deviation, variance, and the mean is by using the stats functions on your calculator. On tests and assignments it is preferred that you use the calculator stat functions (so you do not need the formulas from Example 5).

Stat mode (1) → Store values enter #, press M+ → Stat functions above ( ) x ÷ $\boxed{\text{RCL}}$ mean $\overline{x}$ St. dev Sx

Stat 260 Lecture Notes
Set 3 - The Correlation Coefficient

When data arise in pairs, such as (x_1, y_1), (x_2, y_2), (x_3, y_3) ... (x_n, y_n) the structure is called **bivariate**.

Here we may want to see if there is a relationship between the x and y values. To do this we can use a **scatterplot**.



upward trend
straight line
positive relationship,

higher midterm scores correspond to higher final exam scores

can use this model to predict datavalues

Be careful with extrapolating (making predictions). This data only shows what happens on the final exam for students with midterm scores between 9 and 37. The data included here would not be useful for making final exam predictions for midterm scores of, say, 80. (In other words, your data is only useful in making predictions for other data values close to your collection.)

Recall: variance $= s^2 = \frac{\sum(x_i - \overline{x})^2}{n-1} = \frac{\Sigma(x_i - \bar{x})(x_i - \bar{x})}{n-1}$

positive covariance: positive slope

When we work with bivariate data we can calculate the **covariance**, s_{xy}.

$$s_{xy} = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{n-1}$$

can be positive or negative

Example 1: Calculate the covariance for the data (3, 4), (8, 7), (10, 8), (11, 8).

x	y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$
3	4	-5	-2.75	(-5)(-2.75) = 13.75
8	7	0	0.25	0(0.25) = 0
10	8	2	1.25	(2)(1.25) = 2.5
11	8	3	1.25	(3)(1.25) = 3.75

n describes how many observations

$\bar{x} = 8$ $\bar{y} = 6.75$

$S_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{13.75 + 0 + 2.5 + 3.75}{4-1} = \frac{20}{3} = 6.67$

$\bar{x}=8$ $\bar{y}=6.75$

$S_{xy} = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{n-1} = \frac{13.75+0+2.5+3.75}{4-1} = \frac{20}{3} = 6.67$

Kind of meaningless on its own (only tells you positive slope)

The correlation coefficient, r, measures the strength of the linear relationship between x and y values.

$$r = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}} = \frac{s_{xy}}{s_x s_y}$$

where s_{xy} = the covariance of x and y, s_x = the standard deviation of x values, s_y = the standard deviation of y values

Example 2: Calculate the correlation coefficient using the data from Example 1. (3,4),(8,7),(10,8),(11,8)

$S_{xy} = \frac{20}{3}$

$S_x = 3.559$

$S_y = 1.893$

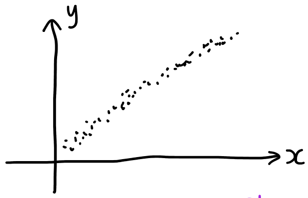
$\frac{\frac{20}{3}}{(3.559)(1.893)} = \frac{20}{3(3.559)(1.893)} = 0.9895$

Scaling covariance measurement into a percentage

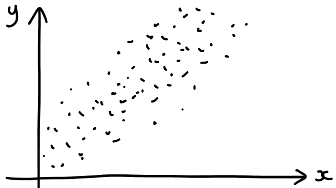
Rule: No matter what we have for x and y values, $-1 \leq r \leq +1$.

The closer that that correlation coefficient r is to $+1$ or -1, the stronger the linear relationship there is. A positive value of r indicates a positive linear relationship, and a negative value of r indicates a negative linear relationship. A value of 0 indicates no linear relationship.

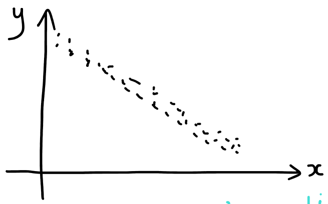
(generally cutoff strong relationship around 0.8)



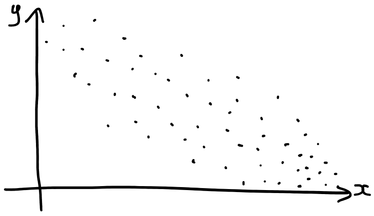
r is positive, strong linear relationship, r is close to +1.



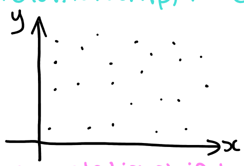
r is positive, weak linear relationship, r is not close to +1.



r is negative, strong linear relationship, r is close to -1.



r is negative, weak linear relationship, r is not close to -1.



no relationship between x and y, r is close to 0

there is a relationship between x and y, just not linear. r is close to 0

An exact linear relationship occurs when $r = 1$ or $r = -1$ and in this case we can represent the data as a straight line in the form $y_i = ax_i + b$.

Be careful! Correlation $\neq$ causation. (spurious correlations)

* r tells us nothing about the value of the slope, other than if it is positive or negative

Set 4

January 17, 2023 11:15 AM

Stat 260 Lecture Notes
Set 4 - Basic Set Theory

The sample space $\mathcal{S}$ of a random experiment is a list of all the possible outcomes of the experiment.

Example 1: The students in our Stat 260 class are asked to report what month they were born in.

S= {Jan, Feb, Mar, ... , Dec}

A simple event is a single outcome from the sample space.

e.g. person is born in March

An event is a set of one or more simple events.

e.g. the event $E = \{\text{Jan}, \text{Feb}, \text{Mar}\}$ = person is born in the first three months of the year

sample space and event both about months

There are multiple ways to assign probabilities to events:

Capital letters for events

- classical approach: used when possible outcomes are equally likely

theoretical approach

$$P(A) = \frac{\#A}{\#\mathcal{S}} = \frac{\#\text{ of ways event } A \text{ can occur}}{\#\text{ of outcomes in } \mathcal{S}}$$

A = born in Jan

P(A) = $\frac{1 \text{ Jan}}{12 \text{ total months}}$

assuming that being born in any month is equal to being born in any other month.

doesn't take into account real life things

- frequency approach: counts the number of times the event occurs in many, many observations

$$P(A) = \lim_{N\to\infty}\left(\frac{\#\text{ of occurrences of } A}{N}\right) \approx \frac{f}{n}$$

N: number of things in population

f: frequency

n: number of observations

P(A) = $\frac{4}{150}$ (4: # w/ Jan birthday; 150: # people asked)

1

In all models, the probability of an event is the sum of its simple events.

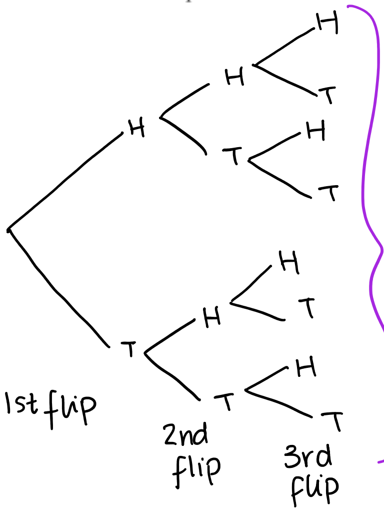
Example 2: Calculate P(born in the first three months) using the classical approach. $E = \{Jan, Feb, Mar\}$

$$P(E) = P(Jan) + P(Feb) + P(Mar) = \frac{1}{12} + \frac{1}{12} + \frac{1}{12}$$
$$= \frac{3}{12}$$
$$= \frac{1}{4} = 0.25 = 25\%$$

$\boxed{a^{b/c}}$ calculator button for reducing fractions

Tree diagrams help us list all possible outcomes in the sample space.

Example 3: Look at the experiment where we flip a fair coin 3 times and we record the sequence of heads and tails.



Outcomes

HHH $H_1 \cap H_2 \cap H_3$
HHT
HTH
HTT
THH
THT
TTH
TTT

# of outcomes = number of endpoints

$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$.

$P(\text{all } H) = \frac{1}{8}$

$P(\text{exactly } 1 H) = \frac{3}{8}$

The union of events A and B, denoted $A \cup B$, is read as "A or B". The set $A \cup B$ = outcomes in A or B or in both.

in either or both

$A \cup B$: glue them together

$A \cap B$: the overlap of the sets

The intersection of events A and B, denoted $A \cap B$, is read as "A and B". The set $A \cap B$ = outcomes that are in both A and B.

In other words, the list of events in $A \cup B$ is the lists from A and B combined together as one larger list. The list of events in $A \cap B$ is the overlap in the lists from A and B.

Sometimes we write AB as a shorter version of $A \cap B$.

The event $\emptyset$ is the empty event / impossible event.

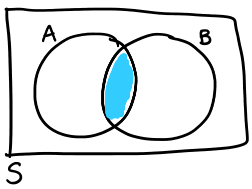
Rule: $P(\emptyset) = 0$.

The events A and B are mutually exclusive (or disjoint) if $A \cap B = \emptyset$. (That is, there is no overlap between events A and B.)

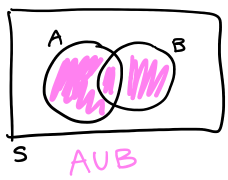
So if A and B are mutually exclusive, then $P(A \cap B) = P(\emptyset) = 0$.

The complement of an event A, denoted $\overline{A}$, is the set of all outcomes in $\mathcal{S}$ that are not in A. "Not A" "$\overline{A}$" "A^c" "A'"

Venn Diagrams help us picture probabilities.

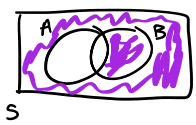


$A \cap B$

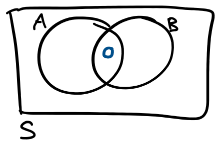


"at least 1 of A or B"

$A \cup B$



$\overline{A}$



A and B mutually exclusive

Rule: DeMorgan's Laws

- $\overline{A \cup B} = \overline{A} \cap \overline{B}$
- $\overline{A \cap B} = \overline{A} \cup \overline{B}$

Note: the phrase "nor" can be translated as "and not". For example "neither A nor B" is the same as "not A and not B".

Stat 260 Lecture Notes
Set 5 - Introduction to Probability

Probability is used to express the likelihood that some event will or will not occur. We measure probability on a scale from 0 to 1, where 0 indicates that it is impossible for the event to occur and 1 indicates that the event is guaranteed to occur.

For an event A, we denote the probability that A will occur by $P(A)$.

Axioms of Probability:

- For any event A, $P(A) \geq 0$. - no negative probability
- $P(\mathcal{S}) = 1$, where $\mathcal{S}$ represents the sample space.
  P(S) = 100%
- If events $A_1, A_2, A_3, \ldots,$ are mutually exclusive (disjoint), then $P(\bigcup_{i=1}^{\infty} A_i) = P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots$.

In the last axiom, the fact that the events are disjoint is important. Below we will see how to find the probability of the union when the events are not disjoint.

A and $\bar{A}$ have no overlap

Notice that an event A and its complement $\overline{A}$ are disjoint and together they comprise all of the sample space $\mathcal{S}$. Therefore we can say that $P(A \cup \overline{A}) = P(A) + P(\overline{A}) = P(\mathcal{S}) = 1$.

Rule: $P(A) + P(\overline{A}) = 1$.

This is often useful in the form $P(A) = 1 - P(\overline{A})$.

Theorem: General Addition Rule

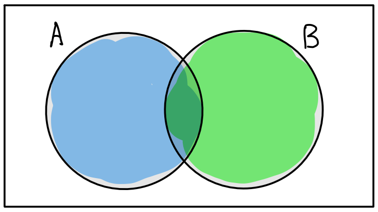
for two sets:

prob of A or B

subtract the overlap

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

→ minus A and B



*we use this formula when we have the word "or"

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

↑
# of elements in $A \cup B$

for three sets:

$$P(A \cup B \cup C) = \underbrace{P(A) + P(B) + P(C)}_{\text{add prob. of single events}} \overbrace{- P(A \cap B) - P(A \cap C) - P(B \cap C)}^{\text{subtract the double}} + \underbrace{P(A \cap B \cap C)}_{\text{add the three}}$$

$$n(A \cup B \cup C) = n(A) + n(B) + n(C) - n(A \cap B) - n(A \cap C) - n(B \cap C) + n(A \cap B \cap C)$$

Can also be done w/ # of elements

Example 1: Suppose that when Michelle is on the computer the only activities she does are:

this is the Sample Space

- event E_1: marking R assignments, occurs with probability p $\quad p$
- event E_2: prepping lectures, 8 times as likely as R assignments $\quad 8p$
- event E_3: answering email, 3 times as likely as R assignments $\quad 3p$
- event E_4: updating course website, 2 times as likely as R assignments $\quad 2p$
- event E_5: creating new assignments, 6 times as likely as R assignments $\quad 6p$

Assume Michelle can only do one activity at a time (i.e. all events are mutually exclusive). A student visits Michelle at a random time and finds that she is working on the computer. What is the probability that she is not marking R assignments?

complement

We know: $E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5 = S$ + all events are disjoint

$$1 = P(S) = P(E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5)$$
$$= P(E_1) + P(E_2) + P(E_3) + P(E_4) + P(E_5)$$
$$= p + 8p + 3p + 2p + 6p$$
$$1 = 20p \Rightarrow p = 1/20$$

Want $P(\bar{E_1}) = 1 - P(E_1)$
$$= 1 - 1/20$$
$$\boxed{= 19/20}$$

Example 2: In a colony of 160 rabbits

$G =$ grey

$E =$ straight ears

$F =$ short fur

single events
- 104 are grey
- 105 have straight ears
- 126 have short fur

double events
- 90 have short fur & are grey $F \cap G$
- 80 have straight ears & short fur $E \cap F$
- 149 are grey **or** have straight ears $G \cup E$
  ↳ use general addition rule
- 5 have none of these qualities

What is the probability that a randomly selected rabbit from the colony has all three qualities? $P(G \cap E \cap F)$

What is the probability that a randomly selected rabbit is grey but has neither of the other two qualities? (i.e. "just grey") $G \cap \bar{E} \cap \bar{F}$

what we want

$$\underbrace{P(G \cup E \cup F)}_{\text{need to find}} = P(G) + P(E) + P(F) - \underbrace{P(G \cap E)}_{\text{need to first}} - P(G \cap F) - P(E \cap F) + P(G \cap E \cap F)$$

$$P(G \cup E \cup F) = 1 - P(\overline{G \cup E \cup F}) = 1 - 5/160 = 155/160$$

none of G or E or F

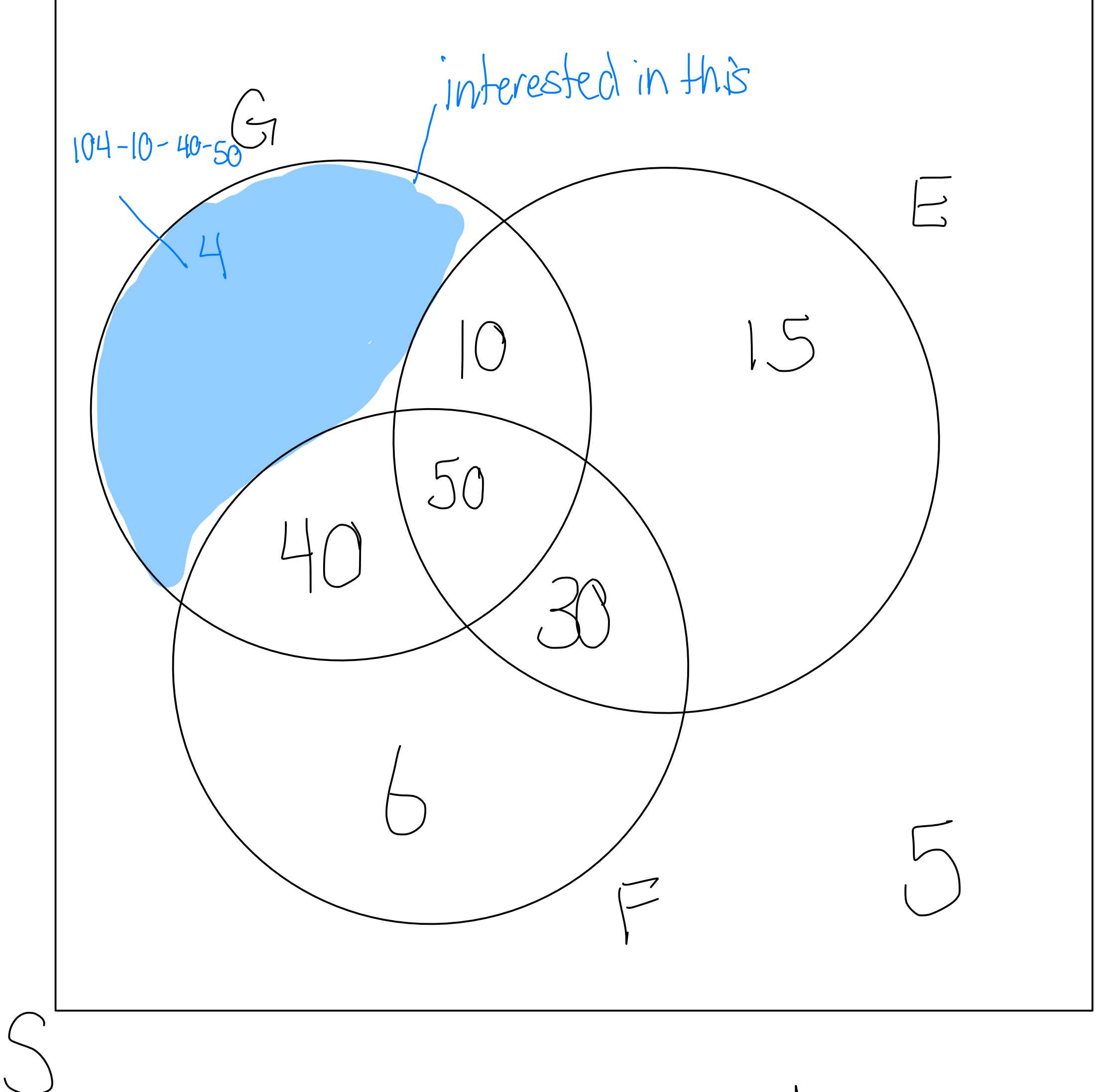
$$P(G \cap E) = P(G) + P(E) - P(G \cap E) =$$

$$\frac{149}{160} = \frac{104}{160} + \frac{105}{160} - P(G \cap E)$$

$$P(G \cap E) = 60/160$$

$$\frac{155}{160} = \frac{104}{160} + \frac{105}{160} + \frac{126}{160} - \frac{60}{160} - \frac{90}{160} - \frac{80}{160} + P(G \cap E \cap F)$$

$$P(G \cap E \cap F) = 50/160 = \boxed{5/16 = 0.3125}$$



Start in the middle & work backwards

$$P(\text{just grey}) = \frac{4}{160} = \frac{1}{40} = 0.025$$

$$P(\text{exactly one}) = \frac{4+15+6}{160} = \frac{25}{160} = \frac{5}{32} = 0.15625$$

Stat 260 Lecture Notes
Set 6 - Conditional Probabilities

Example 1: Rolling a 6-sided die.

Suppose we roll a standard 6-sided die and record the number that is facing up. What is the probability of rolling a 3?

$$P(\text{roll a } 3) = \frac{1 \leftarrow \text{one \# is 3}}{6 \leftarrow \text{six in total}}$$

$$S = \{1, 2, 3, \ldots, 6\}$$

Now suppose we are told that an odd number was rolled. Now what is the probability of rolling a 3?

$$P(\text{roll a 3 if odd \# is rolled}) = \frac{1 \leftarrow \text{one \# is 3}}{3 \leftarrow \text{three \# in sample space } S}$$

new and reduced $S = \{1, \underline{3}, 5\}$ $n=3$

$$P(\text{roll a } 3 \mid \text{odd \#}) = \frac{P(\text{roll a 3 and odd \#})}{P(\text{odd \#})} = \frac{P(\text{roll a 3})}{P(\text{odd \#})} = \frac{1/6}{3/6} = \frac{1}{3}$$

Instead, suppose we were told that an even number was rolled. What is the probability of rolling a 3?

$$P\left(\begin{array}{l}\text{roll a 3 if we know}\\ \text{even \# was rolled}\end{array}\right) = 0 = \frac{0 \leftarrow \text{0's were 3}}{3 \leftarrow \text{3 in sample space } S}$$

new, reduced sample space $= S = \{2, 4, 6\}$

Idea: Knowing extra information can change the probability of an outcome. When we know extra info it is called a *conditional probability*. The "given" part of the event is the extra info we are told.

↑ info we know

Formula: probability of A given B (i.e. the probability that event A will occur given that we know event B has occurred).

$$\underbrace{P(A|B)}_{\text{A given B}} = \frac{\overbrace{P(A \cap B)}^{\text{A and B}}}{P(B)}$$

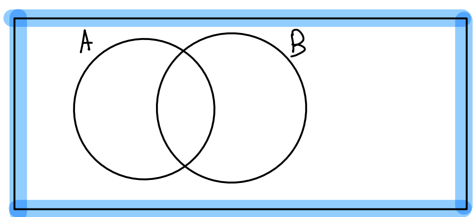
these use original Sample space

Notice we could also talk about the probability of B given A:

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)}. \qquad \frac{\text{always probability of intersect}}{\text{what is given}}$$

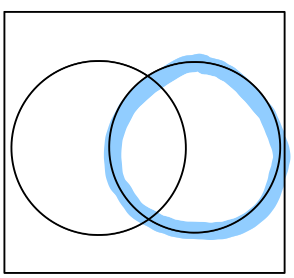
B given A

Where does the formula come from?



S

$$P(A) = \frac{n(A)}{n(S)}$$



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

$$= \frac{n(A \cap B)/n(S)}{n(B)\ /n(S)}$$

Example 2: A maufacturer wants to see if using a clear coating of paint on their product is connected to if the product rusts.

	rust present (R)	no rust present ($\overline{R}$)	
clear coating used (C)	0.03	0.12	0.15
no clear coating used ($\overline{C}$)	0.17	0.68	0.85

given

If we know that a randomly selected component has a clear coating, what is the probability that it has rust present?

$$P(R|C) = \frac{P(R\cap C)}{P(C)} = \frac{0.03}{0.03+0.12} = \frac{0.03}{0.15} = 0.20 = 20\%$$

"If... then..." statements

If we know that a randomly selected component does not have a clear coating, what is the probability that it has rust present?

$$P(R|\overline{C}) = \frac{P(R\cap \overline{C})}{P(\overline{C})} = \frac{0.17}{0.85} = 0.20 = 20\%$$

So here knowing if it clear coating or not did not change the probability of rush

Recall: The formula for conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad P(B) \neq 0 \qquad\qquad P(B|A) = \frac{P(A \cap B)}{P(A)} \qquad P(A) \neq 0$$

From these we get the multiplication rule:

$$P(A \cap B) = P(B) \cdot P(A|B) \qquad\qquad P(A \cap B) = P(A) \cdot P(B|A)$$

We can **partition** an event A by looking at where it overlaps with event B: event A can overlap with event B, or it can overlap with event $\overline{B}$. That is, the events $A \cap B$ and $A \cap \overline{B}$ partition the event A. Note too that $A \cap B$ and $A \cap \overline{B}$ are disjoint events, so we can say

$$P(A) = P(A \cap B) + P(A \cap \overline{B}).$$

If we then use the multiplication rule we can say that

$$P(A) = P(A \cap B) + P(A \cap \overline{B}) = P(B) \cdot P(A|B) + P(\overline{B}) \cdot P(A|\overline{B}).$$

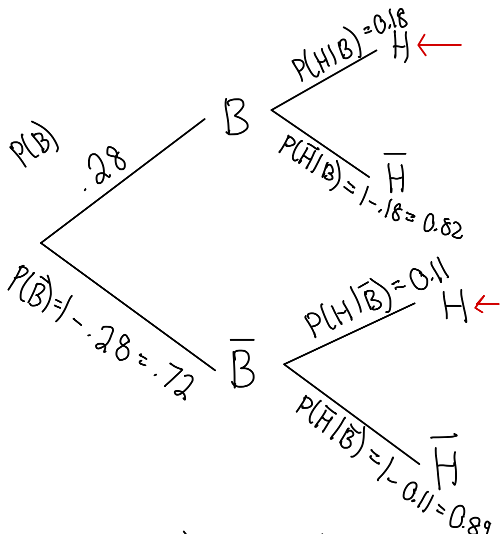
This is best illustrated on a tree diagram.

each split of branches adds up to 100% - complement rule

Example 3: Balding men and heart attacks.

A survey was taken of middle aged men and it was found that 28% of them are balding. Of those who are balding, there is an 18% chance that they will have a heart attack in the next 10 years. For those who are not balding, there is an 11% chance that they will have a heart attack within the next 10 years. What is the probability that a middle aged man will have a heart attack in the next 10 years? - consider a heart attack whether they are balding or not

B = balding, H = heart attack



$$P(H \cap B) = P(B) \cdot P(H|B)$$
$$= 0.28 \cdot 0.18$$
$$= 0.0504$$

$P(\bar{B} \cap H) = (0.72)(0.11) = 0.0792$

Want $P(H) = P(B \cap H) + P(\bar{B} \cap H)$

$$P(H) = 0.0504 + 0.0792$$
$$= 0.1296$$

Idea: multiply going across branch paths to get "and" probability

add cases going down tree

The **law of total probability** says that to find $P(A)$, we add up the probabilities of a partition of A (i.e. add up all disjoint cases that arrive at A). In symbols:

multiply across the branches of the tree

$$P(A) = \sum_{i=1}^{\infty} P(A \cap B_i) = \sum_{i=1}^{\infty} P(B_i) \cdot P(A|B_i)$$

This is the same thing we did when we added together the different cases from the tree branches in the last example.

Bayes' Theorem puts together the conditional probability formula along with the multiplication rule. Using tree diagrams for these questions are very useful!

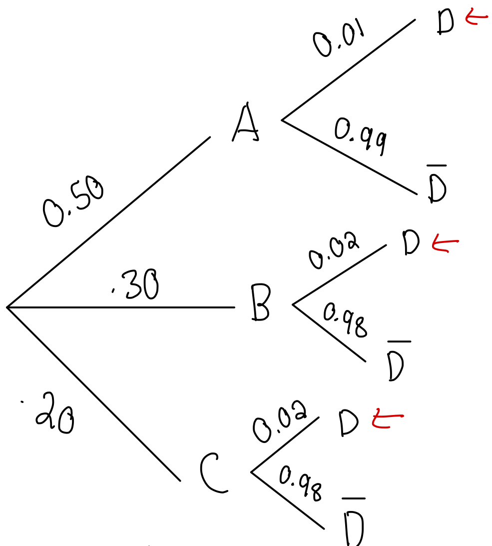
Example 4: TV sets.

TV sets are made at production plants A, B, and C. Suppose 50% are made at plant A, 30% are made at plant B, and 20% are made at plant C. Quality control finds that:

- 1% of plant A TVs are defective.
- 2% of plant B TVs are defective.
- 2% of plant C TVs are defective.

conditional probability

Given that a randomly selected TV is defective, what is the probability that it was produced at plant C?



$$P(C|D) = \frac{P(C \cap D)}{P(D)}$$

$$= \frac{(0.20)(0.02)}{(0.5)(0.01) + (0.30)(0.02) + (0.2)(0.02)}$$

$$= \frac{0.004}{0.015} \approx 0.267$$

$$\frac{P(C) \cdot P(D|C)}{P(A) \cdot P(D|A) + P(B) \cdot P(D|B) + P(C) \cdot P(D|C)}$$

Suppose we have events A_1, A_2, ..., A_n, which are then followed by event B or $\overline{B}$. We can write Bayes' Theorem in symbols as:

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{P(A_1) \cdot P(B|A_1)}{\sum_{i=1}^{n} P(A_i) \cdot P(B|A_i)}$$

You don't need to write the symbolic form in your solution, it is much more important to know how to come up with this from using the conditional probability formula and a tree diagram like we did in Example 4.

Sometimes we perform diagnostic tests and the results shown are wrong.

There are 4 options for outcomes in a diagnostic test:

- The condition actually occurs and the test indicates positive for the condition occurring.
  This is a **true positive**, and no error occurs here.
- The condition actually occurs and the test indicates negative for the condition occuring.
  This is a **false negative**, this is an error.
- The condition does not actually occur and the test indicates positive for the condition occurring.
  This is a **false positive**, this is an error.
- The condition does not actually occur and the test indicates negative for the condition occurring.
  This is a **true negative**, and no error occurs here.

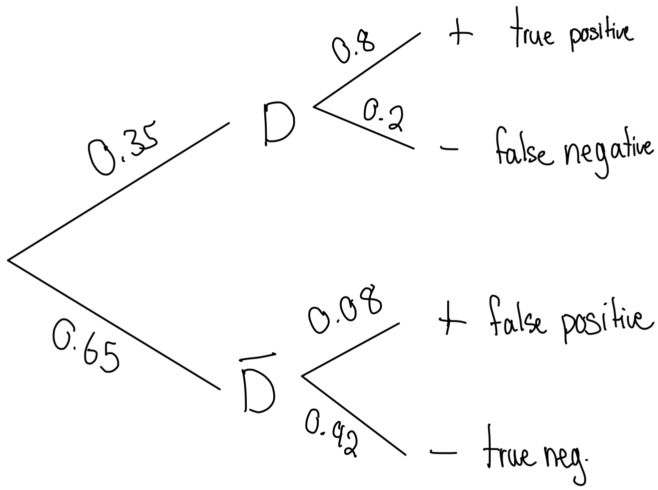
Example 5: Kidney transplants.

A patient receives a kidney transplant. Suppose that 35% of kidney transplants are rejected. During the healing process the patient is tested to see if they are rejecting the kidney. For this particular test the false positive rate is 8% and the false negative rate is 20%. What is the probability that the patient is rejecting the kidney if their test result is positive (i.e. the test indicates they are rejecting the kidney)?

D= actually rejecting kidney

+ = positive result (i.e. rejecting kidney)

− = neg. result (no rejecting)



$$P(D|+) = \frac{P(D \cap +)}{P(+)} = \frac{(0.35)(0.8)}{(0.35)(0.8) + (0.65)(0.08)} = 0.843 = 84.3\%$$

Remember:

- false positive rate $= P(+|\overline{D})$
- false negative rate $= P(-|D)$
- true positive rate $= P(+|D)$. This is also sometimes called the sensitivity.
- true negative rate $= P(-|\overline{D})$. This is also sometimes called the specificity.

Stat 260 Lecture Notes
Set 7 - Independent and Mutually Exclusive Events

Idea: Knowing extra information can change the probability of an outcome. This is what we saw with conditional probabilities. Here we look at the case when knowing the extra information does not change the probability of the outcome. This is the idea of having *independent events*.

Definition: Events A and B are *independent* when $P(A|B) = P(A)$ (or when $P(B|A) = P(B)$). $P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$

Using this definition along with the conditional probability formula we arrive at the alternate definition that events A and B are independent exactly when

$$\boxed{P(A \cap B) = P(A) \cdot P(B)}$$

Only when A and B are independent

Definition: Events A and B are *mutually exclusive* if they cannot occur at the same time (i.e. there is no overlap between A and B).

In the mathematical sense, we have that A and B are mutually exclusive when $\boxed{P(A \cap B) = 0}$ (i.e. when having both event A and B occur together is impossible).

Example 1: Revisiting the clear coating and rust example.

A maufacturer wants to see if using a clear coating of paint on their product is connected to if the product rusts.

	rust present (R)	no rust present ($\overline{R}$)	
clear coating used (C)	0.03	0.12	
no clear coating used ($\overline{C}$)	0.17	0.68	

Is having rust present independent of using the clear coating?

Check: $P(R\cap C) = P(R)\cdot P(C)$

$P(R\cap C) = 0.03$ ← Same

$P(R)\cdot P(C) = (0.20)(0.15) = 0.03$

So $P(R\cap C) = P(R)P(C)$ so R and C are independent events.

Rule: If events A and B are independent, then $\overline{A}$ and B are independent too (and A and $\overline{B}$ are independent, and also $\overline{A}$ and $\overline{B}$ are independent).

So R and $\overline{C}$ are independent too.

Is using the clear coating independent of not using the clear coating?

check: $P(C\cap \overline{C}) = P(C)\cdot P(\overline{C})$

$P(C\cap\overline{C}) = 0$ ← Not Same

$P(C)\cdot P(\overline{C}) = (0.15)(0.85) \neq 0$

So $P(C\cap\overline{C}) \neq P(C)\cdot P(\overline{C})$ so C and $\overline{C}$ are <u>not</u> independent

Careful! "Mutually exclusive" and "independent" are not the same thing. Here "clear coating" and "no clear coating" are mutually exclusive (since they are disjoint), but they are not independent.

$P(C\cap\overline{C}) = 0$

The rule that if A and B are independent then

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

can be extended to more than two events.

Rule: If events $E_1, E_2, E_3, \ldots, E_n$ are independent events, then

$$P(E_1 \cap E_2 \cap E_3 \cap \ldots \cap E_n) = P(E_1) \cdot P(E_2) \cdot P(E_3) \cdot \cdots \cdot P(E_n).$$

Note: When $n \geq 3$ this rule does not work the other way around. That is, just because you have $P(E_1 \cap E_2 \cap \ldots \cap E_n) = P(E_1) \cdot P(E_2) \cdot \cdots \cdot P(E_n)$ does not guarantee that the events E_1, E_2, $\ldots$, E_n are all independent. (To guarantee independence you would have to do this formula check on all pairs, triples, quadruples, etc.)

(So the independence formula is an "if and only if" statement for two sets, and just an "if" statement for three or more sets.)

A, B, C for independent need

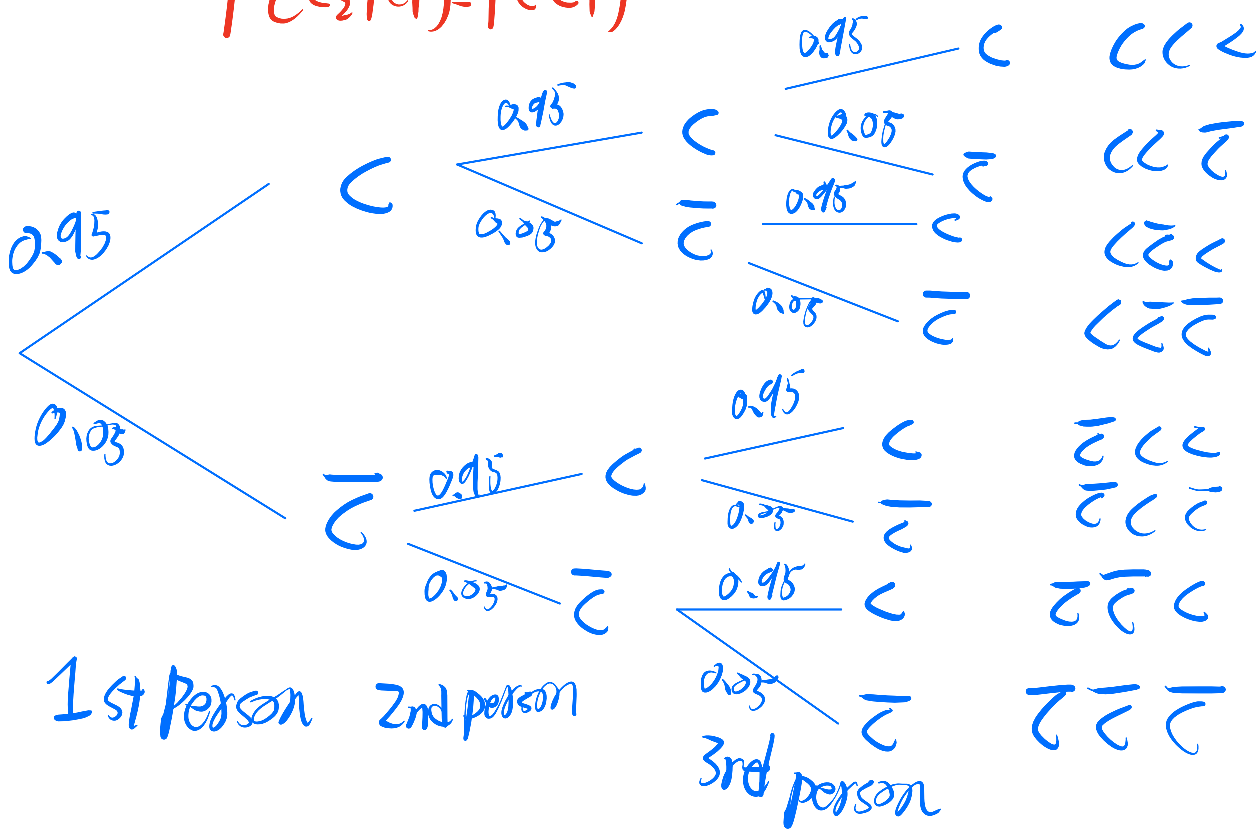
check:

$$\begin{cases} P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C) \\ P(A \cap B) = P(A) \cdot P(B) \\ P(A \cap C) = P(A) \cdot P(C) \\ P(B \cap C) = P(B) \cdot P(C) \end{cases}$$

Example 2: A diagnostic test is correct 95% of the time. Suppose 3 people are independently tested. What is the probability that exactly two of the three receive a correct diagnosis?

C = correct diagnosis

$P(C_2 | C_1) = P(C_1)$



$P(\text{exactly } 2 \text{ correct}) = \boxed{P(CC\bar{C})} + P(C\bar{C}C) + P(\bar{C}CC)$

$= (0.95)(0.95)(0.05) + (0.95)(0.05)(0.95)$

$+ (0.05)(0.95)(0.05)$

$= 0.135375$

$P(CC\bar{C}) = P(C \cap C \cap \bar{C})$

$= P(C) \cdot P(C) \cdot P(\bar{C})$

$= (0.95)(0.95)(0.05)$

Since event are independent

What is the probability that none of the three receive a correct diagnosis?

$$P(\bar{C}\bar{C}\bar{C}) = P(\bar{C}) \cdot P(\bar{C}) \cdot P(\bar{C})$$
$$= 0.05 \cdot 0.05 \cdot 0.05$$
$$= 0.000125$$

What is the probability that at least one of the three receive a correct diagnosis?

$$P(\text{at least one correct}) = 1 - P(\bar{C}\bar{C}\bar{C})$$
$$= 1 - 0.000125$$
$$= 0.999875$$

or add up probs from the first 7 branch path in the tree diagram.

Set 8 - Random Variables

January 24, 2023 12:15 PM

Stat 260 Lecture Notes

Set 8 - Random Variables

HHT

$X = \#$ of H $\quad X=2$

has to be a number

A **random variable (r.v.)** (usually we denote it by X) is a function or a rule that assigns a number to each outcome of the experiment. Back in Set 1 we saw that random variables can be discrete or continuous.

→ can list all possible x values

The **probability mass function (pmf)**, or **probability distribution**, is a table, formula, or graph that describes the possible values of the r.v. and the probability that each value will occur.

Think of the pmf as a function f.

$f(2) = P(X=2)$

r.v. specific values

$$f(x) = P(X = x)$$

function represents a probability

A pmf for a a discrete r.v. X must meet the requirements:

1. $f(x) = P(X = x)$ is defined for all values of x.
2. $f(x) = P(X = x) \geq 0$ for all values of x.
3. $\sum_{\text{all} x} f(x) = \sum_{\text{all} x} P(X = x) = 1$ (the sum of all probabilities is 1).

Continuous probability distributions are studied in a later Set.

1

Example 1: Dominant writing hands.

Suppose 25% of people are left-handed. Suppose we independently sample 3 people and count how many are right handed.

→ so will be an actual number

Let the r.v. X be the number of right-handed people in the 3 sampled. The possible values of X are $0, 1, 2, 3$.

After some work we can find the pmf:

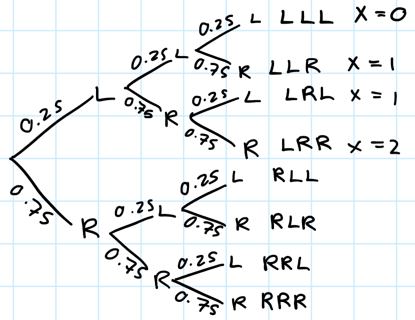
how many right handed

x	0	1	2	3
$f(x) = P(X = x)$	0.015625	0.140625	0.421875	0.421875
	$(0.25)^3$	$3(0.25)^2(0.75)$	$3(0.25)(0.75)^2$	$(0.75)^3$

Notice that all probabilities are ≥ 0 and that

$$\sum_x f(x) = 0.015625 + 0.140625 + 0.421875 + 0.421875 = 1.$$

→ won't always be the same, just because of the specific numbers in this example



(a) Find $P(X = 2)$.

(b) Find $P(X \geq 1)$.

(c) Find $P(X \leq 2 \mid X \geq 1)$.

a) $P(X=2) = 0.421875$

b) $P(X \geq 1) = P(X=1) + P(X=2) + P(X=3)$
$= 0.140625 + 0.421875 + 0.421875$

or $P(X \geq 1) = 1 - P(X=0) = 1 - 0.015625 = 0.984375$

c) $P(X \leq 2 \mid X \geq 1) = \dfrac{P(X \leq 2 \cap X \geq 1)}{P(X \geq 1)} = \dfrac{P(1 \leq X \leq 2)}{P(X \geq 1)}$

overlap

$= \dfrac{P(X=1) + P(X=2)}{P(X \geq 1)}$

$= \dfrac{0.140625 + 0.421875}{0.984375} = 0.5714$

2

The cumulative distribution function (cdf) of a r.v. X is defined as $F(x) = P(X \leq x)$.

So for a value c, $F(c) = P(X \leq c) = \sum_{x \leq c} P(X = x) = \sum_{x \leq c} f(x)$.

$F(2) = P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$

pmf $f(x) = P(X = x)$

cdf $F(x) = P(X \leq x)$

Example 2: Find the cdf for the dominant writing hand example.

pmf

x	$f(x) = P(X=x)$
0	0.015625
1	0.140625
2	0.421875
3	0.421875

⇒ cdf

x	$F(x) = P(X \leq x)$
0	0.015625
1	0.15625
2	0.578125
3	1

→ last entry of cdf table is 1

$F(0) = P(X \leq 0) = P(X=0) = f(0) = 0.015625$

$F(1) = P(X \leq 1) = P(X=0) + P(X=1) = f(0) + f(1) = 0.015625 + 0.140625$
$= 0.15625$

$F(2) = P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = 0.578125$

3

$$cdf \Rightarrow pmf$$

Example 3: The cdf for an experiment is given below. Find the pmf.

x	$F(x)$ $= P(X \le x)$
0	0.15
1	0.38
2	0.74
3	0.92
4	0.98
5	1

pmf

x	$f(x) = P(X = x)$
0	0.15
1	0.23
2	0.36
3	0.18
4	0.06
5	0.02

probabilities sum to 1

$$f(3) = P(X=3) = P(X \le 3) - P(X \le 2)$$
$$= 0.92 - 0.74 = 0.18$$

Example 4: Using the distribution from Example 3, find:

(a) $P(X = 2)$

(b) $P(X \ge 3)$

(c) $P(1 < X \le 4)$

a) $P(X = 2) = 0.36$

pmf

b) $P(X \ge 3) = P(X \le 5) - P(X \le 2)$
$= 1 - 0.74$
$= 0.26$

c) $P(1 < X \le 4) = P(2 \le X \le 4) = P(X \le 4) - P(X \le 1)$
$= 0.98 - 0.38$
$= 0.6$

can also do using the other method

Rules for discrete r.v.s:

- $P(X \ge x) = 1 - P(X < x)$
- $P(X > x) = 1 - P(X \le x)$
- $P(a \le X \le b) = P(X \le b) - P(X < a)$