Conditional Probability and Independence
Chapter 2

Kevin Lutz
PhD Candidate of Statistics
Monday, August 29, 2022

Department of Mathematical Sciences
The University of Texas at Dallas

Table of Contents

Introduction

Joint Probability of Independent Events

$P(A \cap B) = P(A) \cdot P(B)$ if A and B are independent.

- What if A and B are dependent events?
  1. $P(A \cap B) \neq P(A) \cdot P(B)$
  2. Need to compute the conditional probability for dependent events.

- Examples of Conditional Probabilities
  1. Titanic: The probability of survival depends on class.
  2. Travel: The probability of an on time flight depends on the weather, staffing, etc.

Definition: Conditional Probability

Definition: Conditional Probability

The conditional probability of event A given that event B has occurred is notated as

$$P(A|B)$$

and is the $\boxed{\text{probability of } A}$ $\underline{\text{given that event } B \text{ has occurred}}$.

- B is the condition.
- A and B are dependent.
- Note:
  - $P(A|B) \neq P(B|A)$
  - $P(A|B) = P(A)$ when $A \perp B$
  - $P(B|A) = P(B)$ when $A \perp B$

Computing Conditional Probability

- If A and B are dependent events, the Multiplicatin Rule for $P(A \cap B) = P(A) \cdot P(B)$ gets revised as:

Multiplication Rule for Dependent Events

If A and B are dependent events, then their joint probability is given by one of the following:

1. A is the condition:

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

2. B is the condition:

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

Example 1

(a) Two cards are selected from a standard deck of 52 cards without replacement. What is the probability that both cards are 7's?

(b) In a bin, there are 4 red chips and 6 white chips. You select three chips without replacement. What is the probability that the first two are red and the third is white?

See handwritten solutions.

Conditional Probability

1. $P(A \cap B) = P(A) \cdot P(B|A)$
2. $P(A \cap B) = P(B) \cdot P(A|B)$

Solving for the conditional probabilities in either formula above gives

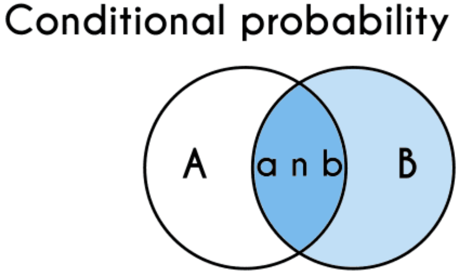
Computing Conditional Probability

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

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

Visualizing Conditional Probability

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



- B becomes the new sample space (denominator term!)
- $P(A|B)$ can be interpreted as the probability of belonging to B but also belonging to A.

Example 2

Given a 6-sided die:

(a) What is the probability of rolling a number less than 5 given that the outcome is odd?

(b) If the outcome is at least 3, what is the probability the outcome is not 6?

See handwritten solutions.

Example 3

- 90% of flights depart on time (A)
- 80% of flights arrive on time (B)
- 75% of flights depart and arrive on time ($A \cap B$)

(a) Interpret and calculate $P(B|A)$.

(b) Interpret and calculate $P(A|B)$.

See handwritten solutions.

Complement of Conditional Probability

Complement of $P(A|B)$

The complement of $P(A|B)$ is $P(A^C|B)$. You can calculate the complement probability using:

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

Example 4

- A company is testing a new medicine for migraine headaches in women. The results of the study are below.

	Improvement	No Improvement	Total
Medicine	132	18	150
Placebo	56	44	100
Total	188	62	250

(a) What is the probability that the headache went away given that the patient was given the medicine?

(b) If a patient felt improvement, what is the probability they were given the medicine?

(c) If a patient felt improvement, what is the probability they were not given the medicine?

See handwritten solutions.

Example 5: Let's Make a Deal

- Bonus round: Choose a door to win a car.

- The host (Monty Hall) then opens one of the goat doors.
- The player is given the choice to stay or switch to the other door.
- Should they stay or switch doors? Does it make any difference? Play the Game

Example 5: Solution if I switch doors

Car Door	First Choice	Host Opens	Door If You Switch	Result
1	1	2 or 3	3 or 2	Lose
1	2	3	1	Win
1	3	2	1	Win
2	1	3	2	Win
2	2	1 or 3	3 or 1	Lose
2	3	1	2	Win
3	1	2	3	Win
3	2	1	3	Win
3	3	1 or 2	2 or 1	Lose

$\Longrightarrow \; P(\text{win car}|\text{switch doors}) = 6/9 = 2/3$

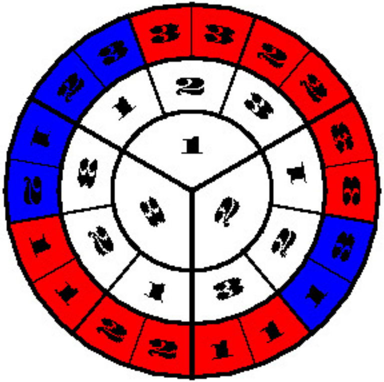
Example 5: Solution if I do not switch doors

Car Door	First Choice	Host Opens	Door If No Switch	Result
1	1	2 or 3	1	Win
1	2	3	2	Lose
1	3	2	3	Lose
2	1	3	1	Lose
2	2	1 or 3	2	Win
2	3	1	3	Lose
3	1	2	1	Lose
3	2	1	2	Lose
3	3	1 or 2	3	Win

$\Longrightarrow$ $P(\text{win car}|\text{do not switch doors}) = 3/9 = 1/3$

Example 5: Continued

Another way to visualize the results in the previous two tables:

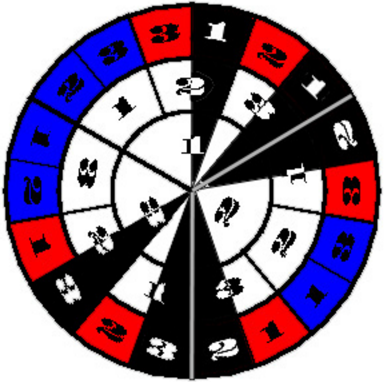


1. Inner wheel: Car door
2. Middle wheel: First door choice
3. Outer wheel: Door Monty opens
   - Red = switch doors to win
   - Blue = do not switch doors to win

Game Simulator

Example 5: No Condition

What if the host does not know which door is the car?



1. If the host accidentally opens the door with the car, the game is over. The contestant can play again.
2. Probability of winning?

$$P(\text{win}|\text{switch}) = 6/12 = 1/2$$

$$P(\text{win}|\text{no switch}) = 6/12 = 1/2$$

These changed since the host does not know the door.

Table of Contents

Comparing Groups Fairly

Suppose the number of failures for CS and SE students last semester were:

	CS	SE
Failed	20	9

- Conclude: CS did worse. (Possibly wrong!)
- Need to consider Ω (whole class) for accurate comparison:

	CS	SE	% CS	% SE
Failed	20	9	20%	90%
Passed	80	1	80%	10%

- Conclude: SE majors did worse. (Correct conclusion)

Comparing Groups Fairly

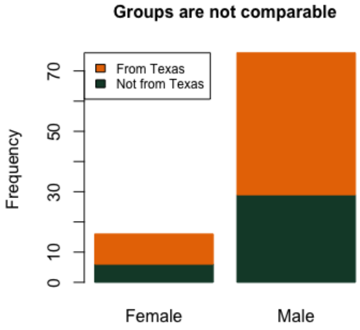
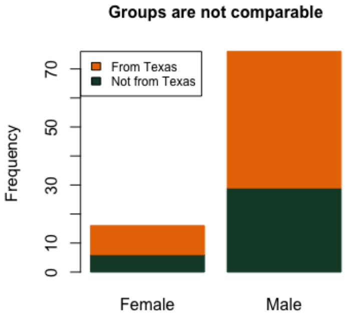


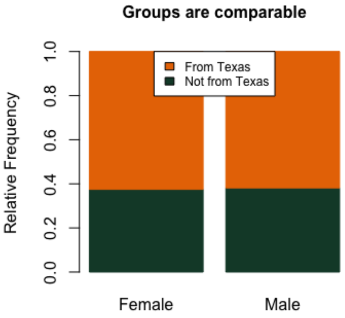
Figure: STAT3341 - Spring 2022

- Beware of different group sizes!
- Groups that are larger in size will always standout.
- Groups need to be "normalized" for fair comparisons.
  - Compare group proportions.

Comparing Groups Fairly

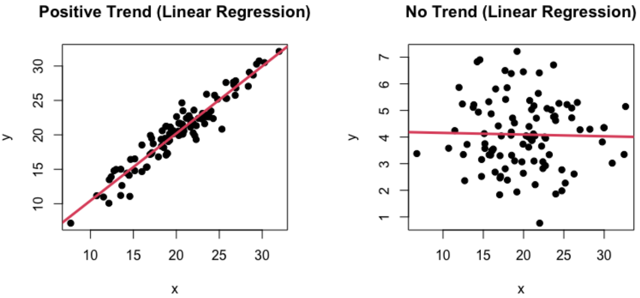
The graph on the right shows the normalized scale to fairly compare the groups.





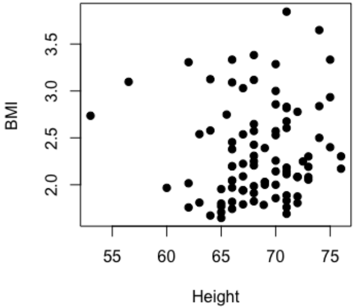
Conclusion: No difference between gender and Texas residency!

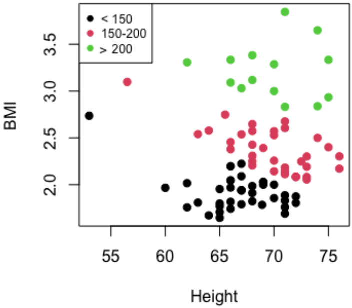
Discovering Trends in Data



- Scatterplots compare two variables.
- Left plot: shows a clear upward (positive) linear trend.
- Right plot: there appears to be no trend.
    - Could there be a trend if we introduce a third variable?

Example





- The first plot shows no clear trend between BMI and Height.
- By accounting for weight, we now see a trend.
- BMI $\propto$ Height/Weight

Definition

Simpson's Paradox

A statistical phenomenon where a trend appears in several groups within a data set but disappears or reverses when the groups are combined into one group.

- Visualize Simpson's Paradox
  - Ignoring the subgroups shows a line with a downward trend through the entire group.
  - Accounting for the subgroups shows a positive trend within each group.
  - The trend reverses!

Avoiding Simpson's Paradox

- Confounding variables: variables that are unknown or ignored.
  - Multiple variables can influence an outcome.
  - Spurious or misleading results will occur by leaving these out.
  - Addressing these reduces error.

- Separate the population into mutually exclusive subgroups called strata.
  - Strata can be gender, age group, major, income group, living situation, etc.

Example 1 - from Wikipedia

Gender bias study among graduate school admissions to University of California, Berkeley in the Fall of 1973.

	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
Total	12,763	41%	8,442	44%	4,321	35%

- Conclusion? Significantly more men were being admitted than women if we consider everyone as a whole.
- Is there enough information to make this conclusion?
- What other variables might we consider?

Example 1 - Consider "Major" as Strata

Department	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
A	933	64%	**825**	62%	108	82%
B	585	63%	**560**	63%	25	68%
C	918	35%	325	37%	**593**	34%
D	792	34%	417	33%	375	35%
E	584	25%	191	28%	**393**	24%
F	714	6%	373	6%	341	7%
Total	**4526**	**39%**	**2691**	**45%**	**1835**	**30%**

Legend:

greater percentage of successful applicants than the other gender

greater number of applicants than the other gender

bold - the two 'most applied for' departments for each gender

1. Top 6 out of 85 majors shown
2. Trend reversal: a greater % of women were admitted than men in some departments
3. More women tended to apply to majors with low acceptance rates.

Example 2: COVID-19

	No Vax	Fully Vax	Total
Hospitalized	214	301	515

Table: Hospitalized cases for COVID-19 in Israel (August 2021); Source: https://mindmatters.ai/2021/09/covid-19-bayes-rule-and-simpsons-paradox/

- Conclusion: The vaccine does not work.
- Two problems:
  1. The values are not normalized. So, the comparison is not fair.
  2. Not accounting for the full Ω (such non-hospitalized cases)

Example 2: COVID-19

Need to compare against "no vax" and "fully vax" group totals!

	No Vax	Fully Vax	Total
Hospitalized	214	301	515
Not Hospitalized	1,302,698	5,634,333	6,937,031
Total	1,302,912	5,634,634	6,937,546

Table: COVID-19 in Israel (August 2021)

1 Calculate $P(\text{Hospitalized}|\text{No Vax})$:

$$= \frac{P(\text{Hosp} \cap \text{No Vax})}{P(\text{No Vax})} = \frac{214/6937546}{1302912/6937546} = 0.00016425$$

Example 2: COVID-19

Need to compare against "no vax" and "fully vax" group totals!

	No Vax	Fully Vax	Total
Hospitalized	214	301	515
Not Hospitalized	1,302,698	5,634,333	6,937,031
Total	1,302,912	5,634,634	6,937,546

Table: COVID-19 in Israel (August 2021)

2 Calculate $P(\text{Hospitalized}|\text{Fully Vax})$:

$$= \frac{P(\text{Hosp} \cap \text{Fully Vax})}{P(\text{Fully Vax})} = \frac{301/6937546}{5634634/6937546} = 0.00005342$$

Example 2: COVID-19 Normalized

3 Compare the results. Use risk ratio:

$$\frac{P(\text{Hospitalized}|\text{No Vax})}{P(\text{Hospitalized}|\text{Fully Vax})} = \frac{0.00016425}{0.00005342} = 3.07$$

- In the whole population, a non-vaccinated person is about 3 times more likely to be hospitalized than someone who is fully vaccinated. (The vaccine works!)
- Trend reversal!

Example 2: COVID-19 Stratified by Age

	No Vax	Fully Vax	Total
Under the Age of 50			
Hospitalized	43	11	54
Not Hospitalized	1,116,791	3,501,107	4,617,898
Total	1,116,834	3,501,118	4,617,962
Probability Hosp	0.00003850	0.00000314	
Risk Ratio			12.25
Over the Age of 50			
Hospitalized	171	290	461
Not Hospitalized	185,907	2,133,226	2,319,133
Total	186,078	2,133,516	2,319,594
Probability Hosp	0.00091897	0.00013593	
Risk Ratio			6.76

No trend reversal, but age-group risk ratios are higher than the population as a whole.