STAT 22000 Lecture Slides
Probability

Yibi Huang
Department of Statistics
University of Chicago

Outline

Coverage: mostly Section 2.1-2.3 in the text.

- Probability and Events (2.1)
- General Addition Rule (2.1.2-2.1.3)
- The Complement Rule (2.1.4-2.1.5)
- Conditional Probability (2.2-2.2.3)
- General Multiplication Rule (2.2.4)
- Independence (2.1.6, 2.2.5)
- Tree Diagrams and Bayes' Theorem (2.2.6-2.2.7)
- Law of Large Number (2.1.1)

Probability and Events

What is Probability?

- People talk loosely about *probability* all the time:
  “What are the chances the Cubs will win this weekend?”
  “What’s the chance of rain tomorrow?”
- For scientific purposes, we need to be more specific in terms of defining and using probabilities

Frequentist Interpretation of Probability

- The probability of heads when flipping a fair coin is 50%
- The probability of rolling a 1 on a 6-sided fair die is 1/6

Everyone agrees with these statements, but what do they really mean?

The *frequentist interpretation of the probability* of an event occurring is defined as the long-run fraction of time that it would happen if the random process occurs over and over again under the same conditions

- Therefore, probabilities are always between 0 and 1

Bayesian interpretation of Probability

- The frequentist interpretation of the probability is limited in application because many interesting random phenomena cannot be repeated over and over again, e.g., weather
- A Bayesian interprets probability as a *subjective degree of belief*: For the same event, two separate people could have different viewpoints and so assign different probabilities
- Nonetheless, both interpretations agree on the probability rules that we will introduce in STAT 220.

Sample Space & Events

- The *sample space* (S) of a random phenomenon is a set of all possible outcomes of the random phenomenon.
- An *event* is is a subset of the sample space.

Example: Flip a coin 3 times and record the side facing up each time.

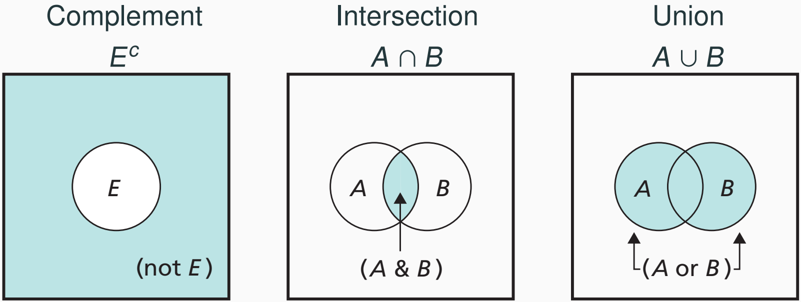
- Sample space
  $S = \{\text{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}\}$
- Events:
  - {all heads} ={HHH}
  - {get one heads} ={HTT,THT,TTH}
  - {get at least two heads} ={HHT,HTH,THH,HHH}

Intersections, Unions, and Complements

- We are often interested in events that are derived from other simpler events:
  - Rolling a 2 or 3
  - Patient who receives a therapy is relieved of symptoms and suffers from no side effects
- The event that A **does not occur** is called the *complement* of A and is denoted A^c or *(not A)*
- The event that **both** A **and** B **occur** is called the *intersection* and is denoted $A \cap B$ or *(A and B)*
- The event that **either** A **or** B **occurs** is called the *union* and is denoted $A \cup B$ or *(A or B)*

Venn Diagrams

Complements, intersections, unions of events can be represented visually using *Venn diagrams*:



General Addition Rule

Disjoint Events (= Mutually Exclusive Events)

Disjoint (mutually exclusive) events cannot be both true.

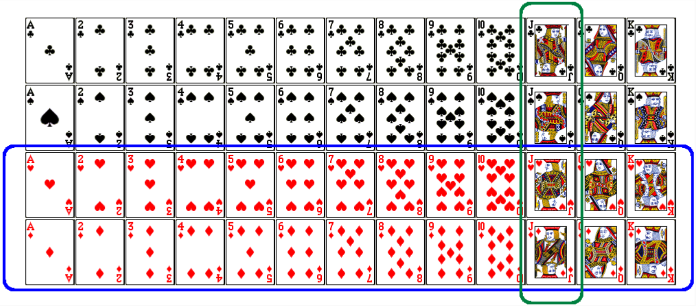
- Tossing a coin once, the events {getting H} and {getting T} are disjoint
- The events {John passed STAT 220} and {John failed STAT 220} are disjoint
- Drawing a card from a deck, the events {getting an ace} and {getting a queen} are disjoint

Non-disjoint events can be both true.

- The events {John got an A in STATs} and {John got an A in Econ} are NOT disjoint

General Addition Rule

What is the probability of drawing a jack or a red card from a well shuffled full deck?



$$P(jack\ or\ red) = P(jack) + P(red) - P(jack\ and\ red)$$

$$= \frac{4}{52} + \frac{26}{52} - \frac{2}{52} = \frac{28}{52}$$

General Addition Rule

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

For disjoint events $P(A \text{ and } B) = 0$, so the above formula simplifies to

$$P(A \text{ or } B) = P(A) + P(B).$$

The Complement Rule

The Complement Rule

- Because an event must either occur or not occur,

$$P(A) + P(A^c) = 1$$

- Thus, if we know the probability of an event, we can always determine the probability of its complement:

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

- This simple but useful rule is called the *complement rule*

Example — The Complement Rule

Question: What is the probability of getting at least one head in 3 tosses of a fair coin?

- Sample space
  $S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$

$$\begin{aligned} \text{Event } A &= \{\text{at least one heads}\} \\ &= \{HHH, HHT, HTH, THH, HTT, THT, TTH\} \end{aligned}$$

- A^c ={not least one heads} ={all tails}= $\{TTT\}$
- $P(A^c)$ =1/8
- $P(A) = 1 - P(A^c) = 1 - 1/8 = 7/8$

Conditional Probability

Example – Poker

A card is drawn from a well-shuffled deck.

- What is the probability that the card drawn is a King?

$$P(\text{got a King}) = \frac{4}{52} = \frac{1}{13}.$$

- If the card drawn is known to be a face card (J, Q, K), what is the probability that it is a K?

$$\frac{4}{12} = \frac{1}{3}$$

Conditional Probabilities

Given two events A and B. We denote the probability of event A happens **given** that event B is known to happen as

$$P(A|B),$$

read as the probability of "*A given B*."

For the example on the previous slide, let

$$A = \text{the card is a King},$$
$$B = \text{the card is a face card (J,Q,K)}.$$

We have

$$P(A|B) = \frac{4}{12} \neq P(A) = \frac{4}{52}.$$

Current knowledge (face card) has changed (restricted) the sample space (possible outcomes).

Exercise

In the previous example, what is

$P(B|A^c) = P(\text{the card is a face card} \mid \text{the card is not a King})$?

- *If the card is not a King, as there are 4 Kings in a deck, then the card must be one of the remaining 48, among which 8 are face cards (4 Jacks and 4 Queens). So the answer is* $8/48$.

$P(B|A) = P(\text{the card is a face card} \mid \text{the card is a King})$?

- $= 1$, *since a King is a face card.*

$P(A|B^c) = P(\text{the card is a King} \mid \text{the card is not a face card})$?

- $= 0$, *since a King is a face card. If it's not a face card, it cannot be a King.*

Another Example

A deck of cards is well-shuffled and the two cards are drawn w/o replacement. What is the probability that second card is a King

given that the first card is a King?

- 3/51

given that the first card is NOT a King?

- 4/51

Example: Age and Rank of Faculty Members

The table below cross-classifies faculty members in a certain university by age and rank.

Age (year)	Full professor	Associate professor	Assistant professor	Lecturer	**Total**
Under 40	54	173	220	23	470
40-49	156	125	61	6	348
50-59	145	68	36	4	253
60+	75	15	3	0	93
Total	430	381	320	33	1164

Marginal Probabilities of Contingency Tables

If a faculty member is selected at random, what is the probability that he/she is a full professor?

Age (year)	Full professor	Associate professor	Assistant professor	Lecturer	**Total**
Under 40	54	173	220	23	470
40-49	156	125	61	6	348
50-59	145	68	36	4	253
60+	75	15	3	0	93
Total	*430*	381	320	33	*1164*

P(full professor) = $\frac{430}{1164} \approx 0.37$.

Probabilities that involve only one of the categorical variables in a contingency table are called *marginal probabilities*.

Joint Probabilities of Contingency Tables

If a faculty member is selected at random, what is the probability that he/she is a full professor and under 40?

Age (year)	Full professor	Associate professor	Assistant professor	Lecturer	**Total**
Under 40	*54*	173	220	23	470
40-49	156	125	61	6	348
50-59	145	68	36	4	253
60+	75	15	3	0	93
Total	430	381	320	33	*1164*

$$\text{P(full professor \& under 40)} = \frac{54}{1164} \approx 0.046.$$

- Probabilities that involve combination of categories of both categorical variables in a contingency table are called *joint probabilities*.
- same as the *overall proportions* for contingency tables introduced in Week 2.

Conditional Probabilities of Contingency Tables

If we know that the selected faculty member is under 40, what is the probability that he/she is a full professor?

Age (year)	Full professor	Associate professor	Assistant professor	Lecturer	**Total**
Under 40	*54*	173	220	23	*470*
40-49	156	125	61	6	348
50-59	145	68	36	4	253
60+	75	15	3	0	93
Total	430	381	320	33	1164

$$P(\text{full professor} \mid \text{under } 40) = \frac{54}{470} \approx 0.11$$

Notice that the *conditional probabilities* of rank given age (column variable given row variable) for a randomly selected faculty member are simply the *row proportions* for contingency tables introduced in Week 2

Conditional Probabilities of Contingency Tables

If we know that the selected faculty member is a full professor, what is the probability that he/she is under 40?

Age (year)	Full professor	Associate professor	Assistant professor	Lecturer	**Total**
Under 40	*54*	173	220	23	470
40-49	156	125	61	6	348
50-59	145	68	36	4	253
60+	75	15	3	0	93
Total	*430*	381	320	33	1164

$$P(\text{under 40} \mid \text{full professor}) = \frac{54}{430} \approx 0.13.$$

Notice that the *conditional probabilities* of age given rank (row variable given column variable) for a randomly selected faculty member are simply the *column proportions* for contingency tables introduced in Week 2

Conditional probability

The conditional probability of the outcome of interest A given condition B can be calculated as

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

Age (year)	Full prof.	Assoc. prof.	Assist. prof.	Lect.	**Total**
< 40	54	173	220	23	470
40-49	156	125	61	6	348
50-59	145	68	36	4	253
60+	75	15	3	0	93
Total	430	381	320	33	1164

$$P(\text{full prof.} \mid \text{under } 40)$$
$$= \frac{54/1164}{470/1164}$$
$$= \frac{P(\text{full prof. \& under } 40)}{P(\text{under } 40)}$$

If only the *joint probabilities* (relative frequencies) are available, but the counts are not,

Age (year)	Full prof.	Assoc. prof.	Assist. prof.	Lect.	**Total**
Under 40	*0.046*	0.149	0.189	0.020	*0.404*
40-49	0.134	0.107	0.052	0.005	0.299
50-59	0.125	0.058	0.031	0.003	0.217
60+	0.064	0.013	0.003	0.000	0.080
Total	0.369	0.327	0.275	0.028	1.000

- the *marginal probabilities* can be computed from the joint probabilities
- the *conditional probabilities* can be computed from the joint probabilities, e.g.,

$$P(\text{full prof.} \mid \text{under 40}) = \frac{P(\text{full prof. \& under 40})}{P(\text{under 40})} = \frac{0.046}{0.404}$$

General Multiplication Rule

General Multiplication Rule

The formula for conditional probability

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

can be used the other way around. Multiplied both sides by $P(A)$, we get the *General Multiplication Rule*:

$$\boxed{P(A \text{ and } B) = P(A) \times P(B|A)}$$

If we want $P(A \text{ and } B)$, and both $P(A)$, $P(B|A)$ are known or are easy to compute, we can use the General Multiplication Rule.

Example: General Multiplication Rule

A deck of cards is shuffled and the two top cards are placed face down on a table. What is the probability that both cards are Kings?

Solution. Let

$$A = \text{ 1st card is a King,}$$
$$B = \text{ 2nd card is a King.}$$

- $P(A) = P(\text{the 1st card is a King}) = 4/52$.
- Given that the 1st card is a King, the conditional probability that the 2nd card is a King =? $P(B|A) = \dfrac{3}{51}$.
- So the probability that both cards are Kings =?

$$P(A \text{ and } B) = P(A) \times P(B|A) = \frac{4}{52} \times \frac{3}{51} = \frac{1}{221} \approx 0.0045.$$

Example: General Multiplication Rule

A deck of cards is shuffled and the two top cards are placed face down on a table. What is the probability that neither card is a K?

Sol. Let

$$A = \text{ 1st card is NOT a K,}$$
$$B = \text{ 2nd card is NOT a K.}$$

- $P(A) = P(\text{the 1st card is not a K}) = 48/52$.
- Given that the 1st card is not a K, the conditional probability that the 2nd card is not a K =? $P(B|A) = \frac{47}{51}$.
- So the probability that neither card is a K =?
  $P(A \text{ and } B) = P(A) \times P(B|A) = \frac{48}{52} \times \frac{47}{51} = \frac{188}{221} \approx 0.851.$
- P(at least one of the two cards is a K) =?
  {at least a K}^{c} ={neither is a K}
  So, P(at least a K)= 1− P(neither is K) $= 1 - 0.851 = 0.149$.

General Multiplication Rule for Several Events

$$
\begin{aligned}
P(ABC) &= P(A) \times P(B|A) \times P(C|AB) \\
P(ABCD) &= P(A) \times P(B|A) \times P(C|AB) \times P(D|ABC) \\
P(ABCDE) &= P(A) \times P(B|A) \times P(C|AB) \times P(D|ABC) \times P(E|ABCD)
\end{aligned}
$$

and so on

Example: General Multiplication Rule for Several Events

Five cards are dealt from a deck of well-shuffled card. What is the chance that none of them are hearts $\heartsuit$?

Sol. Let A_i be the event that the ith card dealt is not a $\heartsuit$.

- $P(A_1) = P(\text{1st card is not a } \heartsuit) = 39/52$
- Given that the 1st card is not a $\heartsuit$, the conditional probability that the 2nd is not a $\heartsuit = P(A_2|A_1) = \dfrac{38}{51}$.
- Given neither of the first two cards is a $\heartsuit$, the condition probability that the 3rd is not a $\heartsuit = P(A_3|A_1A_2) = \dfrac{37}{50}$.
- Likewise, $P(A_4|A_1A_2A_3) = \dfrac{36}{49}$, $P(A_5|A_1A_2A_3A_4) = \dfrac{35}{48}$
- By the General Multiplication Rule,

$$P(A_1A_2A_3A_4A_5) = \frac{39}{52} \times \frac{38}{51} \times \frac{37}{50} \times \frac{36}{49} \times \frac{35}{48} \approx 0.222$$

Continue the previous slide, what is the probability of getting at least one heart $\heartsuit$ among the five cards?

Sol. Since $\{$at least one $\heartsuit\}^c = \{$no $\heartsuit\}$, by the complement rule,

$$P(\text{at least one } \heartsuit) = 1 - P(\text{no } \heartsuit) = 1 - 0.222 = 0.778.$$

Keep in mind:

- The complement of {at least one is...} is {none is...}
- The complement of {all are...} is {at least one is not ...}

Independence

Independence

Two random processes are *independent* if knowing the outcome of one provides no useful information about the outcome of the other.

- Knowing that the coin landed on a head on the first toss <u>does not</u> provide any useful information for determining what the coin will land on in the second toss.
  $\longrightarrow$ Outcomes of two tosses of a coin are independent.
- Knowing that the first card drawn from a deck is an ace <u>does</u> provide useful information for determining the probability of drawing an ace in the second draw.
  $\longrightarrow$ Outcomes of two draws from a deck of cards (w/o replacement) are dependent.

Independence and Conditional Probabilities

In mathematical notation, if $P(A|B) = P(A)$ then the events A and B are said to be independent.

- Conceptually: Giving B doesn't tell us anything about A.

Equivalently, one can also check the independency of the events A and B by check whether $P(B|A) = P(B)$.

Practice – Checking for Independence

Between January 9-12, 2013, SurveyUSA interviewed a random sample of 500 NC residents asking them whether they think widespread gun ownership protects law abiding citizens from crime, or makes society more dangerous. 58% of all respondents said it protects citizens. 67% of White respondents, 28% of Black respondents, and 64% of Hispanic respondents shared this view. Are opinion on gun ownership and race ethnicity independent?

$$P(\textit{protects citizens}) = 0.58$$
$$P(\textit{protects citizens} \mid \textit{White}) = 0.67$$
$$P(\textit{protects citizens} \mid \textit{Black}) = 0.28$$
$$P(\textit{protects citizens} \mid \textit{Hispanic}) = 0.64$$

P(protects citizens) varies by race/ethnicity, therefore opinion on gun ownership and race ethnicity are dependent.

http://www.surveyusa.com/client/PollReport.aspx?g=a5f460ef-bba9-484b-8579-1101ea26421b

Independence of Events v.s. Independence of Variables in a Two-Way Contingency Table

Recall that for a two-way contingency table, the two variables are *independent* if the *row proportions do not change from row to row*. This is consistent with the definition of independence of events since

$$\text{row proportions} = P(\text{column var.}|\text{row var.}),$$

and if the row proportions do not change from row to row, they will be equal to the marginal prob. of the column variable.

For the faculty example, as $P(\text{rank} \mid \text{age}) \neq P(\text{rank})$, the age and rank of faculty members are dependent.

Multiplication Rule for Independent Events

When A and B are independent

$$P(A \text{ and } B) = P(A) \times P(B)$$

- This is simply the general multiplication rule: $P(A \text{ and } B) = P(A) \times P(B|A)$ in which $P(B|A)$ reduce to $P(B)$ when A and B are independent
- More generally,

$$P(A_1 \; and \; \cdots \; and \; A_k) = P(A_1) \times \cdots \times P(A_k)$$

  if $A_1, \ldots, A_k$ are independent.

Exercise: You roll a 6-face die twice, what is the probability of getting two aces in a row?

$$P(\text{ace in the 1st roll}) \times P(\text{ace on the 2nd roll}) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

Abuse of the Multiplication Rule

As estimated in 2012, of the U.S. population,

- 13.4% were 65 or older, and
- 52% were male.

True or False and explain: $0.134 \times 0.52 \approx 7\%$ of the U.S. population were males age 65 or older.

False, age and gender are dependent.
In particular, as women on average live longer than men, there are more old women than old men.

Among those age 65 or older, only 44% are male, not 52%.
Of the U.S. population in 2012, only $0.134 \times 0.44 \approx 5.9\%$ *were males age 65 or older.*

Tree diagrams and Bayes' Theorem

Example – Nervous Job Applicant

Suppose an applicant for a job has been invited for an interview.

The probability that

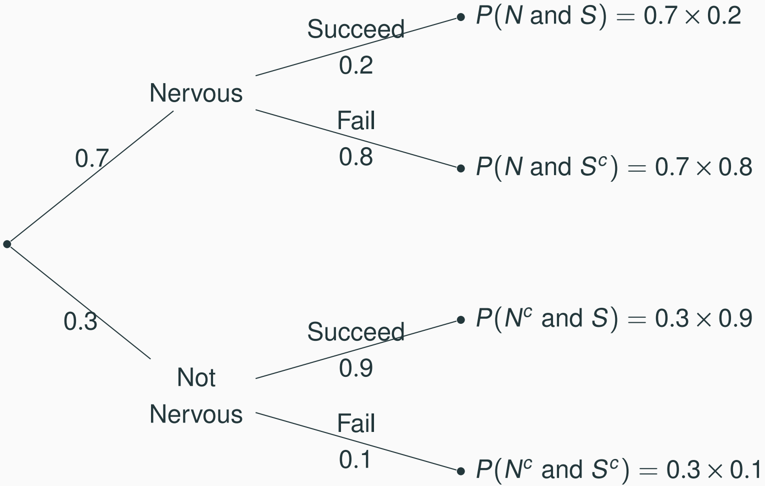
- he is nervous is $P(N) = 0.7$,
- the interview is successful given he is nervous is $P(S|N) = 0.2$,
- the interview is successful given he is not nervous is $P(S|N^c) = 0.9$.

What is the probability that the interview is successful?

$$\begin{aligned} P(S) &= P(S \text{ and } N) + P(S \text{ and } N^c) \\ &= P(N)P(S|N) + P(N^c)P(S|N^c) \\ &= 0.7 \times 0.2 + 0.3 \times 0.9 = 0.41 \end{aligned}$$

Tree Diagram for the Nervous Job Applicant Example

Another look at the nervous job applicant example:



Nervous Job Applicant Example Continued

Conversely, given the interview is successful, what is the probability that the job applicant is nervous during the interview?

$$
\begin{aligned}
P(N|S) = \frac{P(N \text{ and } S)}{P(S)} &= \frac{P(N \text{ and } S)}{0.41} \\
&= \frac{P(N)P(S|N)}{0.41} \\
&= \frac{0.7 \times 0.2}{0.41} = \frac{14}{41} \approx 0.34.
\end{aligned}
$$

in which $P(S) = 0.41$ was found in the previous slide.

Bayes' Theorem

The problem in the previous slide is an example of the **Baysess' Theorem**.

Knowing $P(B|A), P(B|A^c)$, and $P(A)$, is there a way to know $P(A|B)$?

$$
\begin{aligned}
P(A|B) = \frac{P(A \text{ and } B)}{P(B)} &= \frac{P(A)P(B|A)}{P(B)} \\
&= \frac{P(A)P(B|A)}{P(B \text{ and } A) + P(B \text{ and } A^c)} \\
&= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)}
\end{aligned}
$$

Medical Testing

A common application of Bayes' rule is in the area of diagnostic testing

- Let D denote the event that an individual has the disease that we are testing for
- Let $T+$ denote the event that the test is *positive*, and $T-$ denote the event that the test comes back *negative*
- $P(T+|D)$ is called the *sensitivity* of the test
- $P(T-|D^c)$ is called the *specificity* of the test
- Ideally, both $P(T+|D)$ and $P(T-|D^c)$ would equal 1. However, diagnostic tests are not perfect. They may give false positives and false negatives.

Enzyme Immunoassay Test for HIV

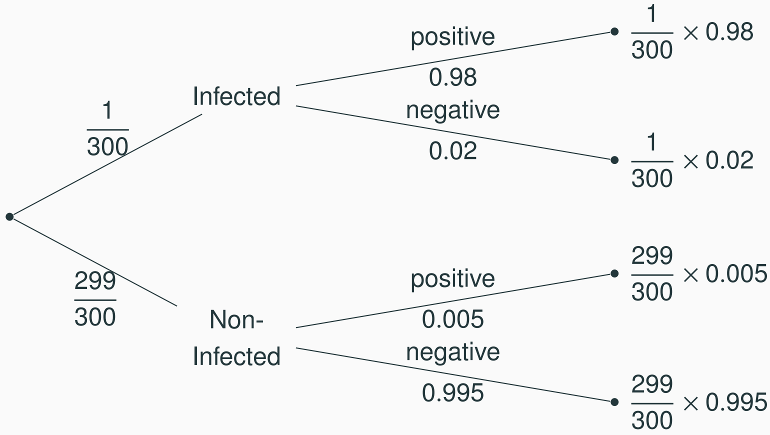
- $P(T+|D) = 0.98$ (sensitivity - positive for infected)
- $P(T-|D^c) = 0.995$ (specificity - negative for non-infected)
- $P(D) = 1/300$ (prevalence in the US: estimated 1 million HIV infected)

What is the probability that the tested person is infected if the test was positive?

$$
\begin{aligned}
P(D|T+) &= \frac{P(D)P(T+|D)}{P(D)P(T+|D) + P(D^c)P(T+|D^c)} \\
&= \frac{(1/300) \times 0.98}{(1/300) \times 0.98 + (299/300) \times 0.005} \\
&= 39.4\%
\end{aligned}
$$

This test is not confirmatory. Need to confirm by a second test.

Tree Diagram for the Enzyme Immunoassay Test for HIV



$$P(D|T+) = \frac{(1/300) \times 0.98}{(1/300) \times 0.98 + (299/300) \times 0.005}$$

Draw a tree diagram!

Don’t try to memorize the formula of Bayes’ Theorem.

Law of Large Numbers

Law of Large Numbers

Law of large numbers states that as more observations are collected, the proportion of occurrences with a particular outcome, converges to the probability of that outcome.

Misunderstanding of Law of Large Numbers I

When tossing a *fair* coin, if heads comes up on each of the first 10 tosses, what do you think the chance is that another head will come up on the next toss? 0.5, less than 0.5, or more than 0.5?

$$\underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{?}$$

- The probability is still 0.5, or there is still a 50% chance that another head will come up on the next toss.

$$P(H \text{ on } 11^{th} \text{ toss}) = P(T \text{ on } 11^{th} \text{ toss}) = 0.5$$

- The coin is not "due" for a tail.
- The common misunderstanding of the LLN is that random processes are supposed to compensate for whatever happened in the past; this is just not true and is also called *gambler's fallacy*

Misunderstanding of Law of Large Numbers II

Let

$$H(n) = \text{\# of heads obtained in } n \text{ tosses of a fair coin.}$$

Generally, we say LLN says that "$H(n) \approx \frac{n}{2}$."

What does "$H(n) \approx \frac{n}{2}$" mean precisely?

- Does $\left|H(n) - \frac{n}{2}\right| \longrightarrow 0$ as n gets large?
- Does $\left|\frac{H(n)}{n} - \frac{1}{2}\right| \longrightarrow 0$ as n gets large?

See Lab 4 and HW3.