STAT 3375Q: Introduction to Mathematical Statistics I

Lecture 3: Bayes' rule

Mary Lai Salvaña, Ph.D.

Department of Statistics
University of Connecticut

January 24, 2024

Outline

1. Quiz Review Exercises
2. Previously...
   - Counting Rules
   - Conditional Probability
   - Independence
3. Bayes' Rule

Quiz Review Exercises

Problem 1

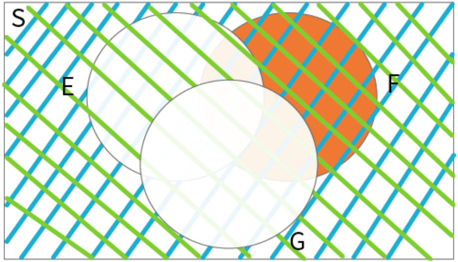
Let E, F, G be three events. Find expressions for the following events using notations for intersection, union, and complement:

a. only F occurs,
b. both E and F but not G occur,
c. at least one event occurs,
d. at least two events occur,
e. all three events occur,
f. none occurs,
g. at most one occurs,
h. at most two occur.

Problem 1

Let E, F, G be three events. Find expressions for the following events using notations for intersection, union, and complement:

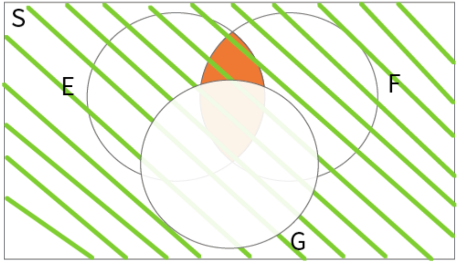
a. only F occurs: $F \cap \overline{E} \cap \overline{G}$



Problem 1

Let E, F, G be three events. Find expressions for the following events using notations for intersection, union, and complement:

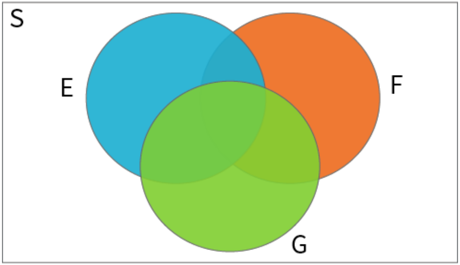
- (b) both E and F but not G occur: $E \cap F \cap \overline{G}$



Problem 1

Let E, F, G be three events. Find expressions for the following events using notations for intersection, union, and complement:

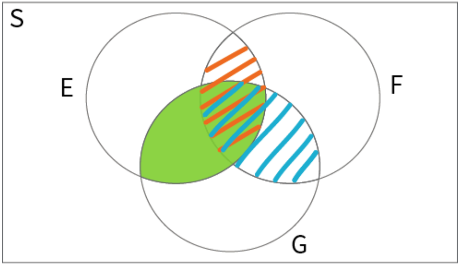
c. at least one event occurs: $E \cup F \cup G$



Problem 1

Let E, F, G be three events. Find expressions for the following events using notations for intersection, union, and complement:

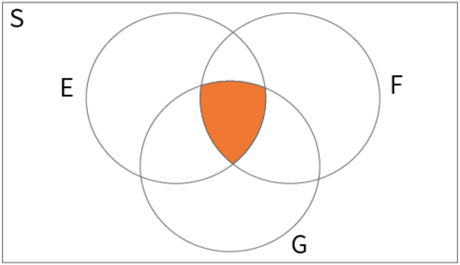
d. at least two events occur: $(E \cap F) \cup (E \cap G) \cup (F \cap G)$



Problem 1

Let E, F, G be three events. Find expressions for the following events using notations for intersection, union, and complement:

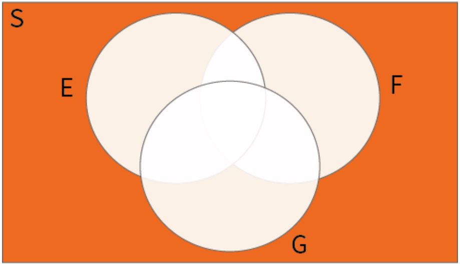
- (e) all three events occur: $E \cap F \cap G$



Problem 1

Let E, F, G be three events. Find expressions for the following events using notations for intersection, union, and complement:

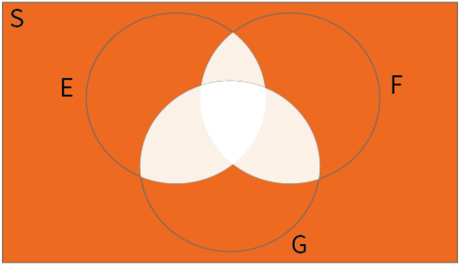
f. none occurs: $\overline{(E \cup F \cup G)}$



Problem 1

Let E, F, G be three events. Find expressions for the following events using notations for intersection, union, and complement:

g. at most one occurs: $\overline{(E \cap F)} \cap \overline{(E \cap G)} \cap \overline{(F \cap G)}$

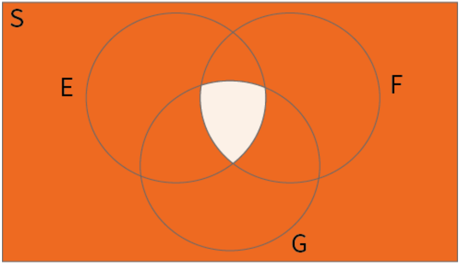


Remove those "regions" where two events or all three events happen.

Problem 1

Let E, F, G be three events. Find expressions for the following events using notations for intersection, union, and complement:

h. at most two occur: $\overline{E \cap F \cap G}$



Remove those "regions" where all three events happen.

Problem 2

If the occurrence of B makes A more likely, does the occurrence of A make B more likely?

Solution:

Given: $P(A|B) > P(A)$.
$\Rightarrow \frac{P(A \cap B)}{P(B)} > P(A)$ (conditional probability formula)
$\Rightarrow P(A \cap B) > P(B)P(A)$
$\Rightarrow \frac{P(A \cap B)}{P(A)} > P(B)$
$\Rightarrow P(B|A) > P(B)$ (conditional probability formula) .

Yes, the occurrence of A make B more likely. □

Problem 3

In a class there are four freshman boys, six freshman girls, and six sophomore boys. How many sophomore girls must be present if sex and class are to be independent when a student is selected at random?

Solution:
Given: 4 freshman boys, 6 freshman girls, 6 sophomore boys.

- Let F be the event that the student is a freshman.
- Let S be the event that the student is a sophomore.
- Let B be the event that the student is a boy.
- Let G be the event that the student is a girl.

Asked: The number of sophomore girls.
Strategy: Let x be the number sophomore girls.

$P(F) = \frac{\text{num of freshmen}}{\text{total num of students}} = \frac{\text{4 boys+6 girls}}{\text{4 freshman boys+6 freshman girls+6 sophomore boys}+x} = \frac{10}{16+x}$. (2 pts)

$P(B) = \frac{\text{num of boys}}{\text{total num of students}} = \frac{\text{4 freshman boys+6 sophomore boys}}{\text{4 freshman boys+6 freshman girls+6 sophomore boys}+x} = \frac{10}{16+x}$. (2 pts)

$P(F \cap B) = \frac{\text{num of freshman boys}}{\text{total num of students}} = \frac{\text{4 freshman boys}}{\text{4 freshman boys+6 freshman girls+6 sophomore boys}+x} = \frac{4}{16+x}$. (2 pts)

We need to solve for x so that $P(F \cap B) = P(F)P(B)$.

In a class there are four freshman boys, six freshman girls, and six sophomore boys. How many sophomore girls must be present if sex and class are to be independent when a student is selected at random?

Solution:

We need to solve for x so that $P(F \cap B) = P(F)P(B)$. (2 pts)

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

$\Rightarrow \frac{4}{16+x} = \frac{10}{16+x}\frac{10}{16+x}$.

$\Rightarrow 16 + x = \frac{100}{4}$.

$\Rightarrow x = 9$. (2 pts)

There must be 9 sophomore girls. □

Problem 4

Suppose we repeatedly roll two fair six-sided dice, considering the sum of the two values showing each time. What is the probability that the first time the sum is exactly 7 is on the third roll?

Solution:

The probability of getting 7 on any one roll is $\frac{6}{36} = \frac{1}{6}$.

Thus, the probability of not getting 7 on the first two rolls, and then getting it on the third roll, is equal to $\left(1 - \frac{1}{6}\right)\left(1 - \frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{25}{216}$. □

Problem 5

Suppose that we ask randomly selected people whether they share your birthday.

a. Give an expression for the probability that no one shares your birthday (ignore leap years).
b. How many people do we need to select so that the probability is at least 0.5 that at least one person shares your birthday?

Solution:

a. $\frac{(364)(364)\cdots(364)}{365^n} = \frac{364^n}{365^n}$.
b. We need to solve for n such that the following relationship holds:
$\Rightarrow 1 - \left(\frac{364}{365}\right)^n = 0.5$
$\Rightarrow 1 - 0.5 = \left(\frac{364}{365}\right)^n$
$\Rightarrow 0.5 = \left(\frac{364}{365}\right)^n$
$\Rightarrow \log(0.5) = \log\left\{\left(\frac{364}{365}\right)^n\right\}$
$\Rightarrow \log(0.5) = n \log\left(\frac{364}{365}\right)$
$\Rightarrow \frac{\log(0.5)}{\log\left(\frac{364}{365}\right)} = n$
$\Rightarrow \frac{\log(0.5)}{\log\left(\frac{364}{365}\right)} = n \Rightarrow n \approx 253.$

Previously...

Counting Rules

- Multiplication Principle: If one thing can be done in m ways and another thing can be done in n ways, the two things can be done in mn ways.
- Permutation: the number of ways to order r objects out of n total objects where each object is different is:

$$P_r^n = \frac{n!}{(n-r)!}.$$

- Permutation w/ repetition: the number of ways to order r objects out of n total objects where each object is different is:

$$P_r^n = \frac{n!}{q_1!q_2!q_3!\dots},$$

  where $q_1, q_2, q_3, \dots$ are the number of times each object is repeated.
- Combination: the number of ways to group r objects out of n total objects is:

$$C_r^n = \frac{n!}{(n-r)!r!}.$$

Conditional Probability

- As you obtain additional information, how should you update probabilities of events?
- Formula:

$$P(A|B) = \frac{\text{probability of events A and B both occurring}}{\text{probability of event B occurring}} = \frac{P(A \cap B)}{P(B)}, \quad \text{provided } P(B) > 0.$$

Independence

- Two events are said to be independent if knowing one occurs does not change the probability of the other occurring.
- Any one of the following should hold:

$$
\begin{aligned}
P(A|B) &= P(A) \\
P(B|A) &= P(B) \\
P(A \cap B) &= P(A)P(B)
\end{aligned}
$$

  Otherwise, the events are said to be *dependent*.
- Independent events are different from mutually exclusive events.

Bayes' Rule

Partitioning the Sample Space

Definition 2.11

For some positive integer k, let the sets $B_1, B_2, \ldots, B_k$ be such that

- $S = B_1 \cup B_2 \cup \ldots \cup B_k$.
- $B_i \cap B_j = \emptyset$, for $i \neq j$.

Then the collection of sets $\{B_1, B_2, \ldots, B_k\}$ is a *partition* of S.

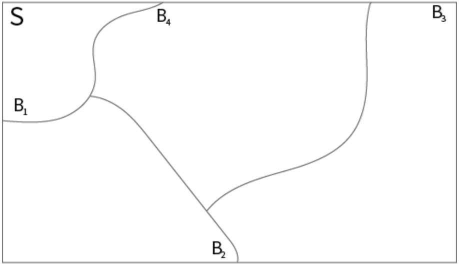


Figure: S is the union of 4 mutually exclusive sets: B_1, B_2, B_3, and B_4.

Decomposition of Any Subset of S

If A is any subset of S and $\{B_1, B_2, \ldots, B_k\}$ is a partition of S, A can be decomposed as follows:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_k).$$

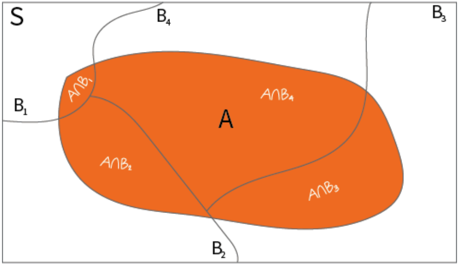


Figure: Decomposing A using B_1, B_2, B_3, and B_4.

Law of Total Probability

Theorem 2.8: Law of Total Probability

Assume that $\{B_1, B_2, \ldots, B_k\}$ is a partition of S such that $P(B_i) > 0$ for $i = 1, 2, \ldots, k$. Then for any event A,

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i).$$

Law of Total Probability

Proof of Theorem 2.8: Need to show $P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$

- If $\{B_1, B_2, \ldots, B_k\}$, then any subset A of S can be written as

$$\begin{aligned} A = A \cap S &= A \cap (B_1 \cup B_2 \cup \ldots \cup B_k) \\ &= (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_k). \quad \text{(Eq. 1)} \end{aligned}$$

- If $i \neq j$, $(A \cap B_i)$ and $(A \cap B_j)$ are mutually exclusive events.
  This is because $(A \cap B_i) \cap (A \cap B_j) = A \cap (B_i \cap B_j) = A \cap \emptyset = \emptyset$. Note $(B_i \cap B_j) = \emptyset$ since $\{B_1, B_2, \ldots, B_k\}$ is a partition of S, meaning any B_i and B_j are disjoint.
- We can write $P(A)$ as

$$\begin{aligned} P(A) &= P\{(A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_k)\} \quad \text{(Eq. 1)} \\ &= P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_k) \quad \text{(Additive Law)} \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_k)P(B_k) \quad \left(P(A|B) = \frac{P(A \cap B)}{P(B)}\right) \\ &= \sum_{i=1}^{k} P(A|B_i)P(B_i). \quad \square \end{aligned}$$

Example

I have three bags that each contain 100 marbles:

- Bag 1 has 75 red and 25 blue marbles;
- Bag 2 has 60 red and 40 blue marbles;
- Bag 3 has 45 red and 55 blue marbles.

I choose one of the bags at random and then pick a marble from the chosen bag, also at random. What is the probability that the chosen marble is red?

Solution:

Let R be the event that the chosen marble is red. Let B_i be the event that I choose Bag i.

- Given: $P(R|B_1) = 0.75$, $P(R|B_2) = 0.6$, and $P(R|B_3) = 0.45$.
- Question: What is $P(R)$?
- Strategy: Apply the law of total probability since B_1, B_2, and B_3 are disjoint and $B_1 \cup B_2 \cup B_3 =$ sample space.

$$\begin{aligned} P(R) &= P(R|B_1)P(B_1) + P(R|B_2)P(B_2) + P(R|B_3)P(B_3) \\ &= (0.75)\frac{1}{3} + (0.6)\frac{1}{3} + (0.45)\frac{1}{3} = 0.6. \end{aligned}$$

Theorem 2.9: Bayes' Rule

Assume that $\{B_1, B_2, \ldots, B_k\}$ is a partition of S such that $P(B_i) > 0$ for $i = 1, 2, \ldots, k$. Then

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}.$$

Proof of Theorem 2.9:

$$\begin{aligned} P(B_j|A) &= \frac{P(A \cap B_j)}{P(A)} && \text{(cond. prob. formula)} \\ &= \frac{P(A|B_j)P(B_j)}{P(A)} && \left(P(A|B) = \frac{P(A \cap B)}{P(B)}\right) \\ &= \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}. && \text{(law total prob.)} \end{aligned}$$

History of Bayes' Rule (Theorem)

Figure: Reverend Thomas Bayes

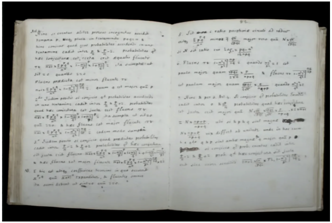


Figure: Bayes' Notebook

- We modify our opinions with objective information:
  Initial Beliefs + Recent Objective Data = A New and Improved Belief
- It is one of the most important theorem in data science.
- More than two centuries old but still heavily used in machine learning and AI

Sources:
1. McGrayne, Sharon Bertsch. The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, & Emerged Triumphant from Two Centuries of C. Yale University Press, 2011.
2. https://www.youtube.com/watch?v=8oD6eBkjF9o

Bayes' Rule (Theorem) Applications

- Insurance: Every time a cyclone or flood hits a region, insurance premiums skyrocket. Why?
- Alan Turing: deciphered Germany's naval Enigma code. Thus, knowing the positions of the German U-boats, trajectories for the supply ships...
- Search and Retrieval Operations: missing ships, planes, or bombs
- biologists: searching for a few plausible solutions for infectious disease

Sources:
1. McGrayne, Sharon Bertsch. The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, & Emerged Triumphant from Two Centuries of C. Yale University Press, 2011.
2. https://www.youtube.com/watch?v=8oD6eBkjF9o

Bayes Rule (Theorem) in ML and AI

- "Thanks to Bayes, we can filter spam, assess medical and other risks, search the Internet for the web pages we want, and learn what we might like to buy, based on what we've looked at in the past. The military uses it to sharpen the images produced when drones fly overhead, and doctors use it to clarify our MRI and Pet Scan images. It is used on Wall Street and in astronomy and physics, the machine translation of foreign languages, genetics, and bioinformatics. The list goes on and on."

Sources:
1. McGrayne, Sharon Bertsch. The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, & Emerged Triumphant from Two Centuries of C. Yale University Press, 2011.
2. https://www.youtube.com/watch?v=8oD6eBkjF9o

Bayes Rule (Theorem) in Sports

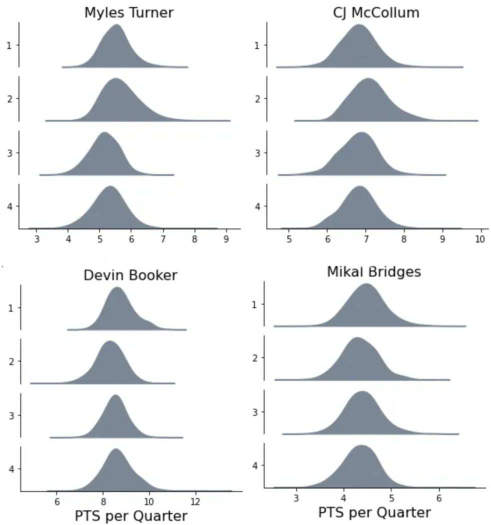


Figure: Example of players with consistent performance from 1st to 4th quarter.

Source: https://binomialbasketball.substack.com/p/modeling-player-decline-throughout

Bayes Rule (Theorem) in Sports

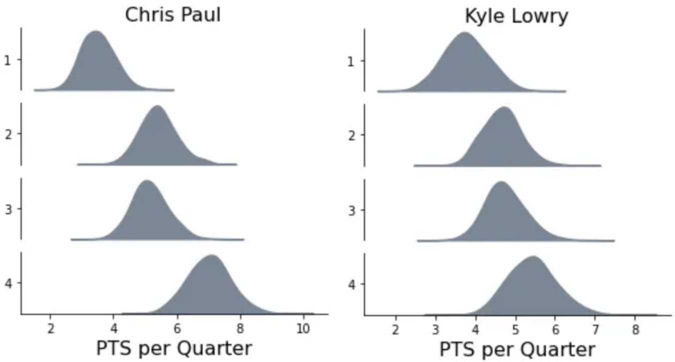


Figure: Example of players with improving performance from 1st to 4th quarter.

Source: https://binomialbasketball.substack.com/p/modeling-player-decline-throughout

Bayes Rule (Theorem) in Sports

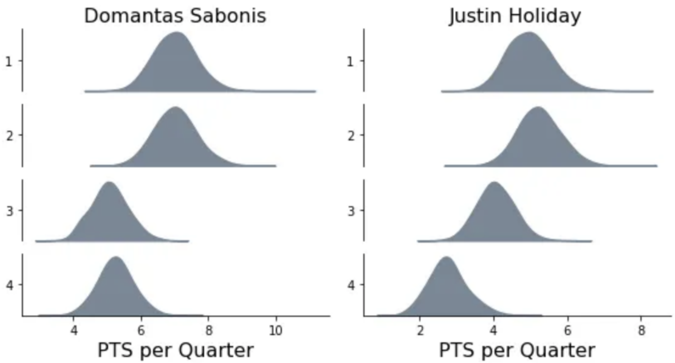


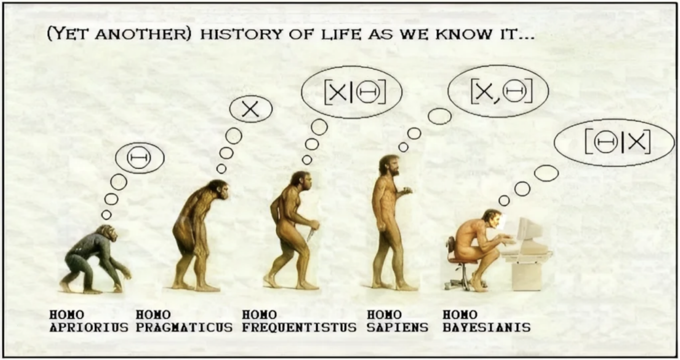
Figure: Example of players with declining performance from 1st to 4th quarter.

Source: https://binomialbasketball.substack.com/p/modeling-player-decline-throughout

Frequentist vs. Bayesian

Source: https://agostontorok.github.io/2017/03/26/bayes_vs_frequentist/

Frequentist vs. Bayesian



Source: https://www2.isye.gatech.edu/isyebayes/jokes.html

Example

Your neighbour is watching their favorite football (or soccer) team. You hear them cheering. Assume probability of goal is 2% and the probability of cheering given a goal is scored is 90%. The probability of cheering when the goal is not scored assumed to be 1%. What is the probability their team has scored?

Solution:

- Given: $P(\text{cheer}|\text{goal}) = 0.9$, $P(\text{goal}) = 0.02$, and $P(\text{cheer}|\text{no goal}) = 0.01$.
- Question: What is $P(\text{goal}|\text{cheer})$?
- Strategy: Apply the Bayes' rule

$$
\begin{aligned}
P(B_j|A) &= \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)} \\
P(\text{goal}|\text{cheer}) &= \frac{P(\text{cheer}|\text{goal})P(\text{goal})}{P(\text{cheer}|\text{goal})P(\text{goal}) + P(\text{cheer}|\text{no goal})P(\text{no goal})} \\
&= \frac{(0.9)(0.02)}{(0.9)(0.02) + (0.01)(1 - 0.02)} = 0.647.
\end{aligned}
$$

Questions?

Homework Exercises: 2.15, 2.17, 2.19, 2.21, 2.23, 2.75, 2.77

Solutions will be discussed this Friday by the TA.