MATH 7343 Applied Statistics

Prof. (Aidong) Adam Ding

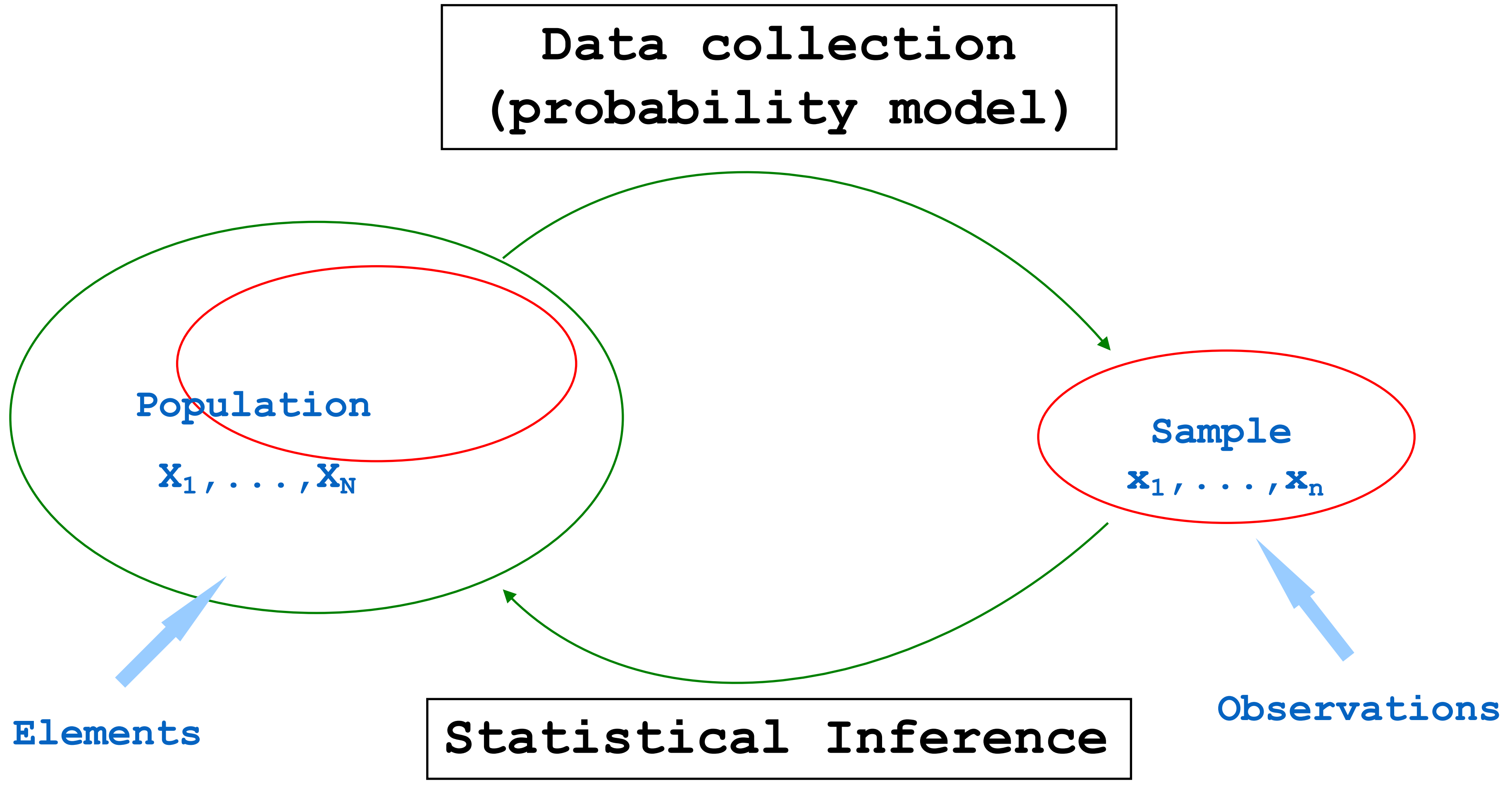
Northeastern University

Review

- Last time, finished Module 2. We learned how to do basic descriptive statistics and using R to do them. Pay attention to statistical concepts.

- Today we start the next topic on probability theory review.

Basic thinking process of statistical inference



Basic thinking process of statistical inference Example: Same sex marriage amendment poll

Population (All MA residents)	Sample (329 persons in the poll)
Parameter:	Statistics:
Proportion of those who support	173/329

If the parameter = 100% → Probability model → always observe 329 out of 329

Hence NOT 100% support ← Statistical inference ← In fact 173 out of 329.

Basic thinking process of statistical inference Example: Same sex marriage amendment poll

Parameter: p=Proportion of those who support the amendment

Statistics: 173/329

Is there majority support ($p \geq 0.5$) or minority support ($p \leq 0.5$)?

Probability model →

If majority support ($p \geq 0.5$) then more likely to observe 173 out of 329 than when $p \leq 0.5$.

← Statistical inference

Hence majority support ($p \geq 0.5$) is more likely.

But how much more likely that the support is majority? For quantification, we need probability theory.

Probability review (Chapter6 event probability)

Definitions:

- Outcome: The result from an experiment.
- Event: A collection (set) of outcomes.

Example: The outcomes of next day's weather can be {snow, rain, sunshine, …}

- Probability: A measure of the likelihood of an event.

Probability: A measure of the likelihood of an event.

There are two interpretations of the probability.

1. (Frequentist) The long-run relative frequency of an event. This is the one we use mainly.

$$\text{Relative frequency} = \frac{\#\ of\ times\ that\ an\ event\ occurs}{\#\ of\ experiments}$$

2. (Bayesian) A personal assessment of the likelihood based on knowledge or prior experience.

Probability definition

2. (Bayesian) A *personal* assessment of the likelihood based on knowledge or prior experience.

For Bayesian probability, the emphasis is on *logical* and *self-consistent* personal assessment.

For example, if you believe that the probability of any woman to be 7 feet tall is zero, and then you met a 7feet tall woman. Then you need to increase the probability (update through Bayes Theorem.)

Event probability

Example of event description:

- In a toxicity study, we have 5 mice: A,B,C,D,E. We inject 1 gram of Heroin to each mouse.
- Possible outcomes:

Only mice A and B died, or only mice B,D and E died, or …

- Event using set notation:

{AB}, {BDE} or {AB, BDE}, …

Example of event description: toxicity study

- In another way, we can instead let A denote that mouse A died (other mice may or may not died.)
- Then {Only mice A and B died} is denoted not as {AB} anymore, but is $A \cap B \cap C^c \cap D^c \cap E^c$ where E^c denotes the *complement* of E (i.e., event E does not occur), $\cap$ denotes intersect.
- Unimportant as to which notations we choose, but we must use set theory notations since events are sets!

Example of event description: toxicity study

- Let A denote that mouse A died then

the event {at least one mouse died, but A lived} is denoted as $(\mathrm{B} \cup C \cup D \cup E) \cap A^c$ where $\cup$ denotes union.

- Set theory notation "union" corresponds to "or" in the event, "intersect" corresponds to "and" in event.
- $\mathrm{B} \cup C$ means that B or C occurs, $\mathrm{B} \cap C$ means that B and C both occurs.

Event Probability Rules

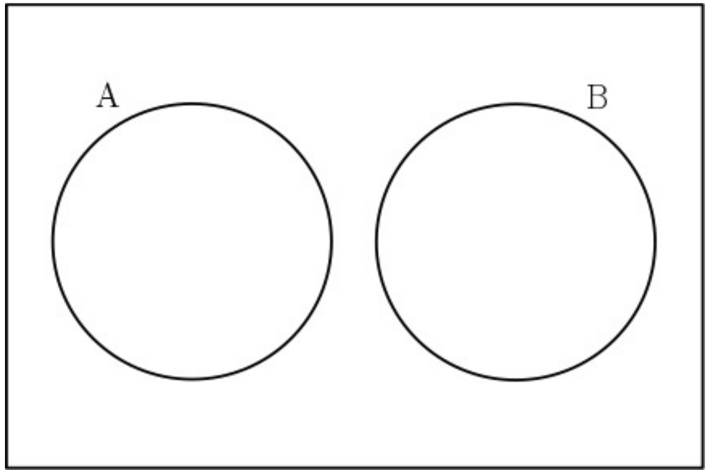
(1) Additive rule for the probabilities:

- If events A and B are mutually exclusive, then the probability that either A or B occurs is

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

- Mutually exclusive: If only one of the two events A and B can occur.

In Venn Diagram,



Event Probability Rules

(1) Additive rule for the probabilities:

- If events A and B are mutually exclusive, then the probability that either A or B occurs is

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

- Example: A={no mice died}, B={all mice died}. Then P(either none or all mice died)

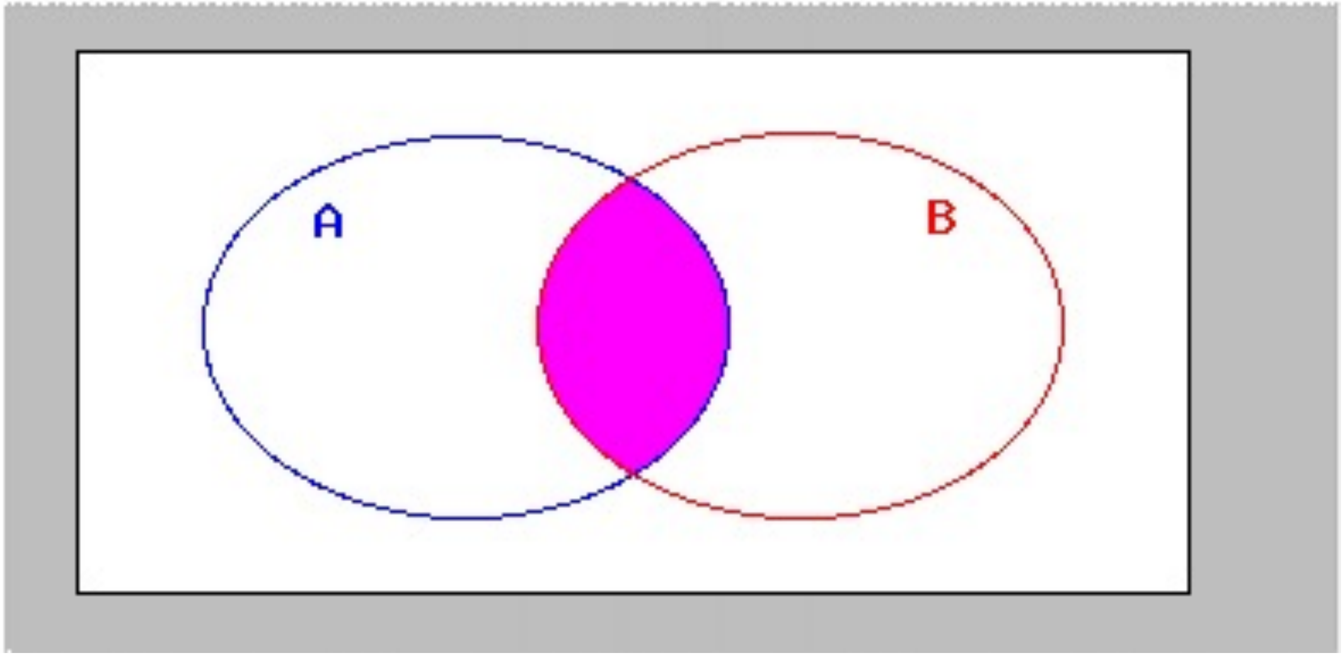
= P(none died) + P(all died)

Event Probability Rules

(1) Additive rule for the probabilities:

- For two general events A and B

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



Event Probability Rules

(1) Additive rule for the probabilities:

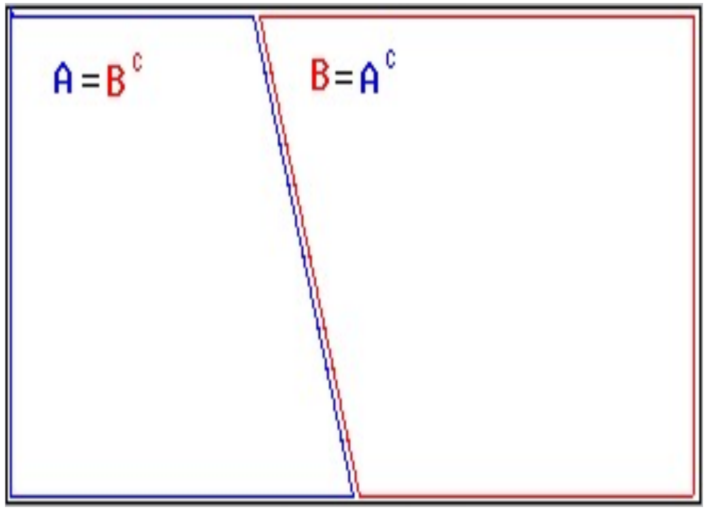
$$P(A \text{ or } B) = P(A) + P(B) - P(A \cap B).$$

- Particularly,

$$P(A \text{ or NOT } A) = P(A \cup A^C) = 1$$

$$= P(A) + P(\text{NOT } A) = P(A) + P(A^C).$$

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



- Example, P(Snow next May) = 0.3,

Then P(No snow next May) = 1 - 0.3 = 0.7

Event Probability Rules

(2) Multiplicative rule for the probabilities:

- If events A and B are independent, then the probability that both A and B occurs is

$$P(A \text{ and } B) = P(A \cap B) = P(A) \cdot P(B).$$

- Independent: If occurrence of one event in no way affects the occurrence of the other.
- Example: A={It snows in Boston next July},

  B={The student ___ gets an "A" in this course}

Event Probability Rules

- Independent: If occurrence of one event in no way affects the occurrence of the other.
- Example: A={It snows in Boston next July},
  B={The student __ gets an "A" in this course}
- Example:

Randomly draw a card from the 52 cards deck.

A={Its suite is Spade }, B={It is an Ace}

$P(A)=1/4$, $P(B)=1/13$, $P(A \cap B)=1/52$

Event Probability Rules(Additive/Multiplicative)

- Example: In the Mass Cash game, you select five numbers between 1-35. You win the jackpot if all five numbers match those drawn. If a player
- buys one ticket, P(wins the jackpot)= $1/\binom{35}{5}$
  =1/324,632
- buys two (different) tickets in the same game,

P(one of the tickets wins the jackpot)

= P(Ticket A or Ticket B wins the jackpot)

= P(A wins) + P(B wins) =2/324,632 (since mutually exclusive)

Event Probability Rules(Additive/Multiplicative)

- A player buys ticket A this week, and buys ticket B next week.

- P(Both ticket win the jackpot)= P(Ticket A and Ticket B wins)

= P(A wins) ·P(B wins) =$1/(324,632)^2$ (since independent)

- P(one of the tickets wins the jackpot)

= P(Ticket A or Ticket B wins the jackpot)

= P(A wins) + P(B wins) - P(A and B wins)

=$2/324,632 - 1/(324,632)^2$

Conditional Probability

- Probability of event A occurs given that B occurs:

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

- Notice that, if A and B are independent, according to above definition:

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

The occurrence of A in no way affects the occurrence of B.

Conditional Probability

- $P(A|B) = \dfrac{P(A \text{ and } B)}{P(B)}$
- Example 1: 20% of all USA residents are of age 40 and older with college degrees. Assume that the median age of USA residents is 40. What is the conditional probability that a USA resident completed college given that the person is 40 years and older?
- Solutions: $P(\text{college}|\geq 40) = \dfrac{P(\text{college and } \geq 40)}{P(\geq 40)} = \dfrac{0.2}{} =$

Conditional Probability

- $P(A|B) = \dfrac{P(A \text{ and } B)}{P(B)}$
- Example 2: If 18% of all USA residents are of age 65 and older. What is the conditional probability that a USA resident is of age 65 and older given that the person is at least 40 years old?
- Solutions: $P(\geq 65 | \geq 40) = \dfrac{P(\geq 65 \text{ and } \geq 40)}{P(\geq 40)} = \dfrac{}{P(\geq 40)}$

$$= \frac{}{0.5} =$$

Conditional Probability

- Probability of event A occurs given that B occurs:

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

- Hence $P(A\ and\ B)$ =P(A|B) P(B)

We can use both formulas.

Conditional Probability

- Example : ELISA test is used to screen donated blood for HIV. For HIV contaminated blood, ELISA is positive 98% of the times. If the blood is not contaminated, ELISA is positive with 7% probability. Suppose 1% of all blood donors have HIV. What is the probability that a donor is HIV positive given that the person is tested positive by ELISA?
- Solutions: P(HIV+ and ELISA+) = P(ELISA+ | HIV+)·P(HIV+)

  = ( )·( ) = 0.0098

Conditional Probability

- ELISA Example Solutions continued:

P(HIV- and ELISA+) = P(ELISA+ | HIV-)·P(HIV-)

= ( )·( ) = 0.0693

P(ELISA+) = P(HIV+ and ELISA+) + P(HIV- and ELISA+)

= 0.0098 + 0.0693 = 0.0791

$$P(\text{HIV+} \mid \text{ELISA+}) = \frac{P(\text{HIV+ and ELISA+})}{P(\text{ELISA+})} = \frac{0.0098}{0.0791} = 12.4\%$$

Anything wrong with this conclusion?

Some terms on test

- Sensitivity

= P(test detects that an event occurs | event occurs)

= P(test + | true +)

- Specificity = P(test - | true -)
- Prevalence = proportion of disease in a population
- In the Example above on ELISA tests for HIV,

Sensitivity = 98%, Specificity = 93%, Prevalence = 1%

Prior and Posterior Probabilities

- The probability before seeing data is called *prior probability*, and the probability after seeing data is called *posterior probability*.
- In the Example above on ELISA tests for HIV,

  Prior P(A donor is HIV+) = Prevalence = 1%

  If a donor is tested positive, then

  Posterior P(A donor is HIV+) = P(HIV+ | ELISA+)

  = 12.4%

Prior and Posterior Probabilities

- Posterior P(A donor is HIV+) = P(HIV+ | ELISA+)
- Recall how did we calculate this:

$$P(\text{HIV+} \mid \text{ELISA+})$$

$$= \frac{P(\text{HIV+ and ELISA+})}{P(\text{ELISA+})}$$

$$= \frac{P(\text{ELISA+} \mid \text{HIV+})P(\text{HIV+})}{P(\text{ELISA+} \mid \text{HIV+})P(\text{HIV+}) + P(\text{ELISA+} \mid \text{HIV}-)P(\text{HIV}-)}$$

Generally, this follows the Bayes Theorem.

Bayes Theorem

$$P(A \mid B)$$

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

$$= \frac{P(A \cap B)}{P(A \cap B) + P(A^C \cap B)}$$

$$= \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A^C)P(A^C)}$$

Summary

- We have finished Chapter 6
- Homework 1 is due at next Lecture time
- You should submit the homework as a PDF file.